

Comparative Analysis of Primordial Germ Cell Specification:

“What Is a Germline and How Do I Get One?”

By Stephany Foster

B.A., Brown University, 2014

Thesis

Submitted in partial fulfillment for the degree of Doctor of Philosophy in the Division of
Biology and Medicine at Brown University

Providence, Rhode Island

May, 2022

© Copyright 2022 by Stephany Foster

This dissertation by Stephany Foster is accepted in its present form
by the Department of Molecular Biology, Cell Biology, and Biochemistry as satisfying
the dissertation requirement for the degree of Doctor of Philosophy.

Date _____

Dr. Gary Wessel, Advisor

Recommended to the Graduate Council

Date _____

Dr. Kristi Wharton, Reader

Date _____

Dr. Nicola Neretti, Reader

Date _____

Dr. Richard Freiman, Reader

Date _____

Dr. Richard Behringer, Reader

Approved by the Graduate Council

Date _____

Dr. Andrew Campbell,
Dean of the Graduate School

Curriculum Vitae

Stephany Foster

stephany_foster@brown.edu

Education

- 2017-Present *PhD Candidate: Molecular Biology, Cell Biology and Biochemistry*
Brown University, Providence, RI
Gary Wessel Laboratory. Investigating the role of cell signaling in germ cell development in the sea star embryo. Single-cell RNA sequencing analysis of early echinoderm embryogenesis.
- 2014 *Bachelor of Arts, Biology*
Brown University, Providence, RI
Lab experience: George Church Lab, Harvard Medical School. Center for Causal Consequences of Variation (CCV) Summer Research Intern, 2012, 2013. Developing recombinant bacterial genomes for metabolic pathway scaffolding.

Fellowship

- 2020-2022 NIH Research Supplement to Promote Diversity in Health-Related Research

Work Experience

- 2021-2022 **Science Communicator**, Brown Technology Innovations, Brown University
- 2018-2020 **Student Writer**, Brown University Graduate School Newsletter, Brown University
- 2014-2017 **Associate Scientific Writer, Community Coordinator**
Recombine, a CooperSurgical Company, New York, NY

Publications

- Feb 2022 **Foster S.**, Oulhen, N. Fresques, T., Zaki, H., Wessel, G. Distinct mechanisms of germ cell factor regulation for an inductive germ cell fate. *Elife* (Submitted.)
- Feb 2022 Spurrel M., Oulhen N., **Foster S.**, Perillo M., Wessel G.M. Gene regulatory divergence among echinoderms underlies appearance of pigment cells in sea urchin development. *Sci Reports* (Submitted.)
- Jan 2022 Satoh N., Hisata K., **Foster S.**, Morita S., Nishitsuji K., Oulhen N., Tominaga H., Wessel G.M. A single-cell RNA-seq analysis of Brachyury-expressing cell clusters demonstrates a morphogenesis-associated signal center of oral ectoderm in sea urchin embryos. *Dev Biol*, 2022. 483: p. 128-142.
- Aug 2020 **Foster S.**, Oulhen N., Wessel G.M. A single cell RNA-seq resource for early sea urchin development. *Development* 2020.
- Aug 2020 Perillo M., Oulhen N., **Foster S.**, Spurrel M., Wessel G.M. Regulation of dynamic pigment cell states at single cell level. *Elife* 2020, **9**.

- Jan 2019 **Foster S.**, Teo Y.V., Neretti N., Oulhen N., Wessel G.M. Single cell RNA-seq in the sea urchin embryo shows marked cell-type specificity in the Delta/Notch pathway. *Mol Reprod and Dev.* 2019;1-4.
- Jan 2019 Oulhen N., **Foster S.**, Wray G., Wessel G.M. Identifying gene expression from single cells to single genes. *Methods Cell Biol* 2019, **151**:127-158.

Scientific Communications

- Aug 2021 Inductive germline formation in the sea star. Dr. Samuel M. Nabrit Conference for Early Career Scholars (virtual talk)
- Mar 2021 Inductive germline formation in the sea star. Developmental Biology of the Sea Urchin and other Marine Invertebrates Meeting (virtual talk)
- July 2020 **Foster S.**, Oulhen N., Wessel G.M. A single cell RNA-seq resource for early sea urchin development. Society for Developmental Biology 79th Annual Meeting Online.
- Apr 2019 **Foster S.**, Oulhen N., Wessel G.M. Mechanisms of germ cell induction during embryonic development. New England Science Symposium; Harvard Medical School, Boston, MA.
- Mar 2019 Foster S., Oulhen N., Wessel G.M. Mechanisms of germ cell induction during embryonic development. Scientista Symposium 2019; Boston, MA.
- Oct 2018 Foster S., Oulhen N., Wessel G.M. Inductive mechanisms for the germ line during development. Developmental Biology of the Sea Urchin XXV; Marine Biological Laboratories, Woods Hole, MA.
- Oct 2018 Reyes G., Vargas H., Foster S., Onorato T., Veras I.; Oulhen N.; Wessel G.M. Exploring the function and regulation of MYP in the developing embryo of the sea star, *Patiria miniata*. Developmental Biology of the Sea Urchin XXV; Marine Biological Laboratories, Woods Hole, MA.
- Oct 2016 Chen, S., Shenhert, S., Ash, W., Kumar, N., **Foster, S.**, Yarnall, S., Carlsson, I. The Doctor Will Tweet You Now: Expanding Access to Care Through Social Media Engagement. American Society for Reproductive Medicine 2016 Scientific Congress & Expo.
- Sept 2016 Canteenwalla, U., **Foster, S.**, Yarnall, S., Gamma, A., Paolino, N., Wieloch, S., Carlsson, I. Bias in Genetic Counseling: Where does it really live?. 35th Annual Education Conference of the National Society of Genetic Counselors.

Service and Leadership

- 2021-2022 MCB Graduate Program Admissions Committee Member
- 2020-2022 Co-Chair of MCB Graduate Program Seminar Series Committee

2021 Mentor, New York City College of Technology, Stem RA Summer Program

2019-2021 Financial Coordinator for Graduate Women in Science and Engineering (GWiSE), Brown University Chapter

Spring 2020 Graduate Student Teaching Assistant, Developmental Biology course

2018-2020 Educational outreach

- Molecular Biology – STEM II program, Wheeler School
- St Andrews School, eight grade class
- Francis J Varieur Elementary Afterschool Program

Acknowledgements

I am so grateful to the Wessel Lab for an experience that was full of laughter, learning, and support. Gary Wessel's thoughtful advising formed me into the scientist and thinker I am now. Thank you for creating a space where everyone feels welcome to contribute and to be their true selves. Thank you for leading by example. There are not enough words in the English or French language for me to thank Nathalie Oulhen. A superhero of science! You are brilliant and honest and I thank you for teaching me so much over the years. Cosmo Pieplow and Gerardo Reyes, I'm so happy that we shared time in this lab together, from conferences to concerts in the injection room.

I'd like to thank my thesis committee, Richard Freiman, Kristi Wharton, Nicola Neretti, and Richard Behringer, for engaging and fruitful discussions regarding my thesis work. Your guidance helped me think critically about my research and your dedication to students is admirable. Kristi Wharton's Developmental Biology lectures influenced my love of development and my decision to return to Brown. I'd like to thank the MCB Graduate Program Coordinators for their support and I want to thank my fellow graduate students for being such warm and welcoming friends.

I am blessed to have made such wonderful friends who have supported me throughout my graduate school journey. Thank you, Emily Gonzalez, for helping me make a home in Providence. Our comfy couch on Hope Street lives forever in my heart. Kimberly Madwhani and Kathryn Russo, you are such pillars of my everyday life! You kept me motivated, encouraged, and met me countless times at "the cry doors".

My sisters, Elizabeth and Vanessa, you are everything. I deeply admire you. Your drive, tenacity, and strength inspired me to pursue this dream of mine. I thank my mom, Maria Lucia, for growing alongside us all these years. My accomplishments and dedication are a reflection of your love and support. To my dad, George, thank you for teaching me I could achieve anything by working hard (and being on time). Finally, I want to thank Wesley Tejeda for being so steadfast. I don't believe you planned to be a graduate school "plus-one" but I am so grateful to have you. Thank you for believing in me and lifting me up.

Table of Contents

Title Page	i
Copyright Page.....	ii
Signature Page.....	iii
Curriculum Vitae.....	iv
Acknowledgements.....	viii
Table of Contents	viii
List of Figures.....	xi
Abstract.....	1
Chapter 1: Comparative analysis of primordial germ cell specification.....	3
Introduction	4
Inductive germline specification in the mouse	8
Inductive germline specification in non-mammalian species	13
Techniques used to study germline development	14
Studying the ancestral mode of germ cell specification in the sea star	20
Germline specification in the sea urchin.....	23
Single-cell RNA sequencing analysis of echinoderm putative PGCs	26
Summary.....	28
References	31
Chapter 2: Single cell RNA-seq in the sea urchin embryo shows marked cell-type specificity in the Delta/Notch pathway.....	37
Introduction	38
Materials and Methods	42
References	44
Figures	45
Chapter 3: A single cell RNA-seq resource for early sea urchin development.....	52
Highlights	53
Abstract	54
Introduction	55
Results	59
Discussion.....	70
Materials and methods	76
References	80

Figures	85
Chapter 4: Distinct mechanisms of germ cell factor regulation for an inductive germ cell fate	93
Highlights	94
Abstract	95
Introduction	96
Results	99
Discussion	109
Materials and Methods	115
References	119
Figures	122
Chapter 5: Impact and Future Directions	142
Research summary	143
Transcriptional Regulation of Nanos and Vasa	146
Post-Transcriptional Regulation of Nanos and Vasa	154
Do all cells have the capacity to become germ cells?	159
How does cell contact regulate Nanos gene expression?	166
The importance of mesoderm.....	167
Additional considerations for germ cell specification	170
Future experimental plans	174
Analysis of Nanos protein expression.....	174
Evolution of Nanos transcript localization methods	175
A reporter for Nanos protein activity	176
Quality control measures in scRNAseq datasets	177
References	179
Appendix A: Identifying gene expression from single cells to single genes.....	184
Summary.....	185
A. Single cell sequencing.....	186
B. Nascent RNA capture.....	215
C. Chromatin Immunoprecipitation.....	219
Conclusions	231
References	233
Appendix B: Regulation of dynamic pigment cell states at single-cell resolution	234
Abstract	235

Introduction	236
Results	241
Discussion.....	255
Materials and Methods	263
References	273
Figures	281
Appendix C: A single-cell RNA-seq analysis of Brachyury-expressing cell clusters suggests a morphogenesis-associated signal center of oral ectoderm in sea urchin embryos	302
Abstract	303
1. Introduction	305
2. Results	307
3. Discussion	320
4. Conclusions	327
5. Materials and methods	328
References	330
Figures	334

List of Figures

Chapter 1: Comparative analysis of primordial germ cell specification.....	3
Figure 1. Germ cell specification mechanisms.....	5
Figure 2. Cell signaling mechanisms and signaling perturbation using small molecules.	11
Figure 3. Phylogenetic relationship among leading models for germ cell specification studies.	19
Figure 4. Germline formation in echinoderms.	23
Figure 5. Comparison of inductive germ cell specification between mouse and sea star.	30
Chapter 2: Single cell RNA-seq in the sea urchin embryo shows marked cell-type specificity in the Delta/Notch pathway.....	37
Figure1:.....	45
Supplement Table 1:.....	46
SFigure 1:	46
SFigure 2:	47
Supplement Table 2:.....	48
SFigure 3:	49
SFigure 4:	50
SFigure 5:	51
Chapter 3: A single cell RNA-seq resource for early sea urchin development.....	52
Chapter 4: Distinct mechanisms of germ cell factor regulation for an inductive germ cell fate	93
Figure 1. Cell states identified across early sea star development.	122
Figure 2. Gene expression at blastula stage.....	123
Figure 3. Gene expression at mid-gastrula stage.....	124
Figure 4. Wnt signaling regulates Vasa expression at gastrula stage.	125
Figure 5. Delta/Notch regulation of germline genes.....	126
Figure 6. Analysis of cell interactions.	127
Figure 7. Inductive mechanism in mouse and sea star embryos.	128
Supplementary Figure S1. Identification of cell states across early sea star development.....	129
Supplementary Figure S2. Marker gene expression at blastula stage.	130
Supplementary Figure S3. Marker gene expression at mid-gastrula stage.	131
Supplementary Figure S4. Wnt signaling component expression.....	132

Supplementary Figure S5. MifL1-enriched cluster expression.	133
Supplementary Figure S6. Vasa, Nanos, and FoxY3 expression.	134
Supplementary Figure S7. Vasa and FoxY3 expression.	135
Supplementary Figure S8. FoxY morpholino knockdowns.	137
Supplementary Figure S9. Changes in transcript abundance with dissociation.	138
Supplementary Figure S10. BMP4 treatment in early sea star embryos.	139
Supplementary Figure S11. Nanos and FoxY gene expression.	140
Chapter 5: Impact and Future Directions	142
Figure 1. Embryonic context of germ cell specification project.	145
Figure 2. Schematic representation of experimental approach to testing Wnt regulation of Nanos transcription.....	148
Figure 3. Capacity for germ cell gene expression in early embryo.	160
Figure 4. Analysis of chromatin landscape in the sea star embryo.	162
Figure 5. Changes in transcript abundance seen with dissociation.....	165
Figure 6. Scheme for single-cell RNA sequencing of embryos and single cells treated with Nodal agonist, Activin.	165
Figure 7. BAC injection approach for testing mesoderm-driven expression of Nanos.	169
Appendix A: Identifying gene expression from single cells to single genes	184
Appendix B: Regulation of dynamic pigment cell states at single-cell resolution	234
Figure 1. Dissection of <i>S. purpuratus</i> ectodermal cell-types by scRNA-seq.....	281
Figure 2. gcm marks pigment cell mitotic and post-mitotic populations.....	282
Figure supplement 1.....	284
Figure supplement 2.....	285
Figure 3. Gene expression in the pigment cells revealed by single cell RNAseq.	286
Figure 4. A unique group of fmo genes is specifically expressed in pigment cells. ...	288
Figure supplement 1.....	290
Figure 5. Fmos from clade one are expressed in embryonic pigment cells and in adult coelomocytes.....	291
Figure supplement 1.....	293
Figure 6. Gcm controls pigment cell specification.....	294
Source data 1.	296
Figure supplement 2.....	298
Figure supplement 3.....	299
Figure supplement 4.....	300

Figure 7. Diagram summarizing the fate of pigment cells in wild type and gcm MO embryos.....	301
Appendix C: A single-cell RNA-seq analysis of Brachyury-expressing cell clusters suggests a morphogenesis-associated signal center of oral ectoderm in sea urchin embryos	302
Fig. 1.....	334
Fig. 2.....	336
Fig. 3.....	337
Fig. 4.....	338
Fig. 5.....	339
Fig. 6.....	342
Fig. 7.....	343
Fig. 8.....	344
Fig. 9.....	345
Fig. 10.....	346

Abstract

Specification of the germline, the cells which give rise to eggs and sperm, is essential for sexually reproducing organisms. The mechanism by which animals specify their germline falls into two categories: inherited or inductive. Echinoderms appear to use diverse mechanisms in germline formation. The sea urchin (*Strongylocentrotus purpuratus*) specifies its germline by inheritance of stored maternal molecules, representative of the inherited mechanism. We recently performed a scRNA-seq analysis of early sea urchin development and identified the transcript profile of the germ cell precursors. The inductive mechanism, used by mammals, relies on cell signaling interactions to direct a subset of embryonic cells to a germ cell fate. Implantation of the mammalian embryo precludes direct, detailed examination of this mechanism in mammals, but previous work suggested that sea star (*Patiria miniata*) embryos, which develop in simple media and are markedly transparent, also use inductive mechanisms to specify their germline. The germ cell factors Nanos and Vasa become restricted in a subset of cells within the posterior enterocoel (PE), the presumptive germline. Nodal signaling was observed to negatively regulate Vasa and Nanos mRNAs to restrict their expression to the PE. Despite the recurrence of the inductive germline mechanism across metazoa, the full network of signals that directs this process remains poorly understood. We learned that Wnt and Delta/Notch signaling enhances Nanos and Vasa expression, whereas a test of dependency in cell interactions reveals that Nanos and Vasa are regulated distinctly. Single-cell analysis of early sea star embryos reveals co-expression of germ cell genes with mesodermal markers suggesting a mesodermal precursor state similar to what has been seen in other species employing inductive

germ cell specification. This work provides insight into the ancestral mechanism of germ cell specification and its evolution across the tree of life. Additionally, the results from this work contrast the results of PGC specification in the sea urchin and enables deeper comparative mechanistic studies in tractable *in vivo* models.

Chapter 1: Comparative analysis of primordial germ cell specification

Introduction

For sexually reproducing organisms, specification of the primordial germ cells (PGCs), the cells which eventually give rise to eggs and sperm, is essential. The PGCs must be partitioned from the rest of the embryo and remain pluripotent until they differentiate into the gametes. Abnormal PGC specification or maintenance can result in animals that are infertile (Persani et al., 2010; Robertson et al., 1999; Tsuda et al., 2003). While the process by which organisms specify the germline is diverse, two major themes have been identified: the inductive and the inherited mechanism. The inherited mechanism, also referred to as preformation, is characterized by the presence of localized maternal factors in the egg. These factors are sequestered or inherited by a subset of cells early in embryonic development and these cells later give rise to the germ cells (Figure 1) (Extavour and Akam, 2003). This mechanism is evident in the fruit fly, *Drosophila melanogaster*, zebrafish, *Danio rerio*, frog, *Xenopus laevis*, and the sea urchin, *Strongylocentrotus purpuratus* (Extavour and Akam, 2003; Yajima and Wessel, 2012). The maternal factors required for germ cell specification, known as pole plasm in the fly and germ plasm in the frog, have been identified by transplantation experiments (Ikenishi et al., 1986; Illmensee and Mahowald, 1974; Tada et al., 2012). When the *Drosophila* pole plasm was transplanted from the posterior to the anterior end of the embryo, germ cells developed in the anterior end, indicating that germline specification happens through inheritance of maternally deposited, localized germline determinants (Illmensee and Mahowald, 1974).

The inductive mechanism, also referred to as epigenesis, differs from the inherited mechanism in that a combination of activating and inhibitory signals from surrounding cells directs a restricted cell population to take on a germ cell fate (Figure 1). This mechanism does not rely on maternal factors, and the process of germ cell specification happens later in development compared to the inherited mechanism (Extavour and Akam, 2003). Many animals use the inductive mechanism, including the cricket *Gryllus bimaculatus*, the mouse *Mus musculus*, and the sea star *Patiria miniata* (Fresques and Wessel, 2018; Nakamura and Extavour, 2016; Ohinata et al., 2009). Transplantation experiments in the mouse showed that signals from surrounding tissues to the proximal epiblast cells are required for PGC specification. Contrary to the *Drosophila* pole plasm, when mouse proximal epiblast cells are transplanted to another region of the embryo, they do not take on a germ cell fate, and distal epiblast cells take on a germ cell fate when transplanted to the proximal region (Tam and Zhou, 1996). Following the inductive mechanism, the cells that receive inductive signals will express germline determinants and take on a germ cell fate.

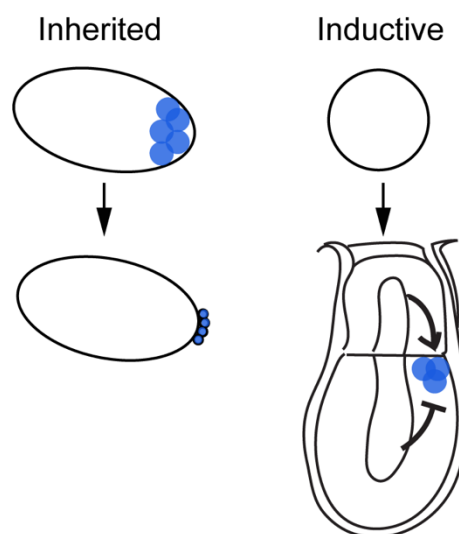


Figure 1. Germ cell specification mechanisms. (Left) Graphic representation of inherited mechanism as seen in *Drosophila*. Localized germ plasm containing

maternally deposited RNAs and proteins required for germ cell fate is present in the oocyte. Germ plasm is inherited asymmetrically by a small number of cells which later become the PGCs. (Right) Graphic representation of the inductive mechanism as seen in the mouse. Species following the inductive mechanism lack localized determinants in the oocyte and zygote. Later in development, activating and inhibitory cell signals converge to specify germ cell gene expression in select cells. Germ cell determinant expression is shown in blue.

Regardless of whether the mechanism of germline specification is inherited or inductive, the PGCs express the highly conserved germ cell factors: Nanos, Vasa, and Piwi. Nanos and Vasa are RNA binding proteins, and translational regulators. In *Drosophila*, Nanos works together with another RNA-binding protein, Pumilio to maintain the germline. Pumilio binds to Pumilio Response Elements (PRE, also known as Nanos Response Elements (NRE)) in the 3' untranslated region (UTR) of target RNAs and recruits Nanos to repress translation of Cyclin B, maintaining the pole cells, the PGC precursors, in cell cycle arrest. Pumilio and Nanos are also required for migration of the pole cells (Asaoka-Taguchi et al., 1999). Nanos expression is required for PGC maintenance in *C. elegans*, mouse, sea urchin, zebrafish, and *Drosophila* (Juliano et al., 2010; Kobayashi et al., 1996; Kopranner et al., 2001; Subramaniam and Seydoux, 1999; Tsuda et al., 2003). There are three Nanos homologues in the mouse, yet only one affects germ cell development, Nanos3. Nanos3 is expressed in the mouse PGCs as early as embryonic day 9.5 (E9.5) and is required for PGC development; Nanos3 knockout mice have smaller ovaries and testes and lack germ cells (Tsuda et al., 2003). Vasa is an RNA helicase, and in *Drosophila* it is required for pole plasm assembly (Hay et al., 1988; Lasko and Ashburner, 1988). In the mouse, Vasa is expressed in the PGCs after they have migrated to the embryonic gonad (Fujiwara et al., 1994). Loss of Vasa results in abnormal meiotic progression and PGC proliferation

in male mice. Male Vasa knockouts produce no sperm while female mice show normal development and fertility (Tanaka et al., 2000). Piwi and Piwi-interacting RNAs (piRNAs) function together to cleave target RNAs. Piwi has been linked to transposon silencing and maintenance of the germline in mouse, zebrafish, and *Drosophila*. Piwi mutants show DNA damage, retrotransposon, and repeat sequences in the germline cells, and the germline is eventually lost through apoptosis (Klattenhoff and Theurkauf, 2008). However, the mechanism of Piwi function in the germline is not fully understood. Well-studied model organisms like the fruit fly, frog, and zebrafish, have made it possible to extensively characterize the inherited mechanism (Hashimoto et al., 2004; Kloc et al., 2001; Raz, 2003; Santos and Lehmann, 2004). The inductive mechanism has been studied to some extent in several animals including axolotl, rabbit, and mouse (Chatfield et al., 2014; Irie et al., 2014; Johnson et al., 2003; Saitou and Yamaji, 2010). Most of what is known about the cell signaling mechanisms that regulate germline specification following the inductive method has been discovered using a mouse model. Characterizing the inductive mechanism requires identifying cell signaling within the context of the embryo. A significant limitation to *in vivo* studies in the mouse is access to and number of cells for experimentation. Knockout studies require developing genetic lines through multiple generations and surgical removal of embryos to observe phenotypes.

Echinoderms are a powerful model for the study of germline specification. The echinoderm phyla is ancestral to chordates providing an opportunity to study the inductive mechanism in a more ancestral model compared to mouse and other

mammals (Extavour and Akam, 2003). Additionally, the sea star and sea urchin spawn millions of gametes which fertilize and develop externally to the adults. The transparent embryos can be easily cultured, treated, dissociated, and analyzed in great quantities. Echinoderms also offer an opportunity to uncover evolutionary differences between two modes of germ cell specification through comparative studies of the sea star and sea urchin. To better define the specification of this important cell lineage, we must investigate the molecular mechanisms of inductive germline specification through diverse models. In this review, we highlight the cell signaling mechanisms identified in inductive germ cell specification across several species using different techniques. We compare recent findings on cell signaling regulation of sea star germ cell specification to what has been reported in mouse and other species. We also compare germ cell gene regulation between two echinoderm species that follow different modes of germ cell specification.

Inductive germline specification in the mouse

The mouse PGCs originate in the proximal epiblast cells of the early embryo. These cells are at the posterior end of the embryo, directly below the extraembryonic ectoderm. They can be seen via positive staining for the germ cell marker alkaline phosphatase at embryonic day 7 to 7.25 (E7-E7.25) (Ginsburg et al., 1990). The PGCs migrate from this region and colonize the embryonic gonad at E10.5 where they later differentiate into gametes. Because *Nanos* and *Vasa* are expressed later in PGC specification, the pluripotency genes, *Oct4*, *Nanog*, and *Sox2*, are commonly used to identify early PGCs (Saitou and Yamaji, 2010). Cell signaling regulates differentiation of

the PGCs and they express these pluripotency factors before expressing Nanos and Vasa. Wnt, Nodal, and BMP signaling have all been found to regulate germ cell specification in the mouse (Ohinata et al., 2009; Saitou and Yamaji, 2010).

Wnt, Nodal, and BMP signaling are an essential feature of development, regulating gene expression, cell proliferation, and pattern formation (MacDonald et al., 2009; Shen, 2007; Weiss and Attisano, 2013). Signaling in these three pathways occurs when a ligand expressed by one cell binds to a transmembrane receptor expressed by a target, or receiving, cell (Heldin et al., 2016). Receptor-ligand interactions can be between two neighboring cells or a ligand could travel to another region of the embryo altogether.

The family of secreted Wnt ligands is conserved across the metazoans. Mice, like humans, express 19 Wnt ligands and 10 Frizzled receptors (Wang et al., 2012). Canonical Wnt signaling refers to β -catenin dependent transcriptional activation. In the absence of Wnt signaling, the Axin complex, which includes Axin, glycogen synthase kinase-3 (GSK3), *adenomatous polyposis coli* gene product (APC), and casein kinase 1 (CK1), targets cytoplasmic β -catenin for degradation. Binding of a secreted Wnt ligand to its receptor, Frizzled, stabilizes cytoplasmic β -catenin in the receptor expressing cell, leading to nuclearization of β -catenin. Once in the nucleus, β -catenin binds to the transcription factor Tcf/Lef, and together they activate expression of Wnt target genes (Figure 2A,B) (Kikuchi et al., 2007; MacDonald et al., 2009). In the mouse, canonical Wnt signaling through Wnt3, is required for gastrulation and is involved in a positive feedback loop with Nodal and BMP4 signaling (Barrow et al., 2007; Wang et al., 2012).

BMP and Nodal are members of the TGF β signaling family. Signaling through both Nodal and BMP begins with a dimeric ligand binding a heterotetrameric receptor complex and ends with nuclear translocation of Smad proteins which regulate gene expression (Weiss and Attisano, 2013). BMP ligands bind two type I and type II receptors and there are three different type I receptors (ALK3, ALK6, ALK2) and three type II receptors (BMPR2, ACVR2A, ACVR2B). The downstream effector and transcriptional regulators for BMP are Smad1, Smad5, and Smad8 which translocate to the nucleus and regulate gene expression (Figure 2A,F) (Ehrlich, 2016; Weiss and Attisano, 2013). The Nodal ligand binds to two type I (ALK4) and type II receptors (ACVR2A, ACVR2B). Nodal binding to its receptors and coreceptor (EGF-CFC) leads to phosphorylation of cytoplasmic Smad2 or Smad3 which then associate with Smad4 and translocate to the nucleus to regulate gene expression (Figure 2A,D) (Shen, 2007).

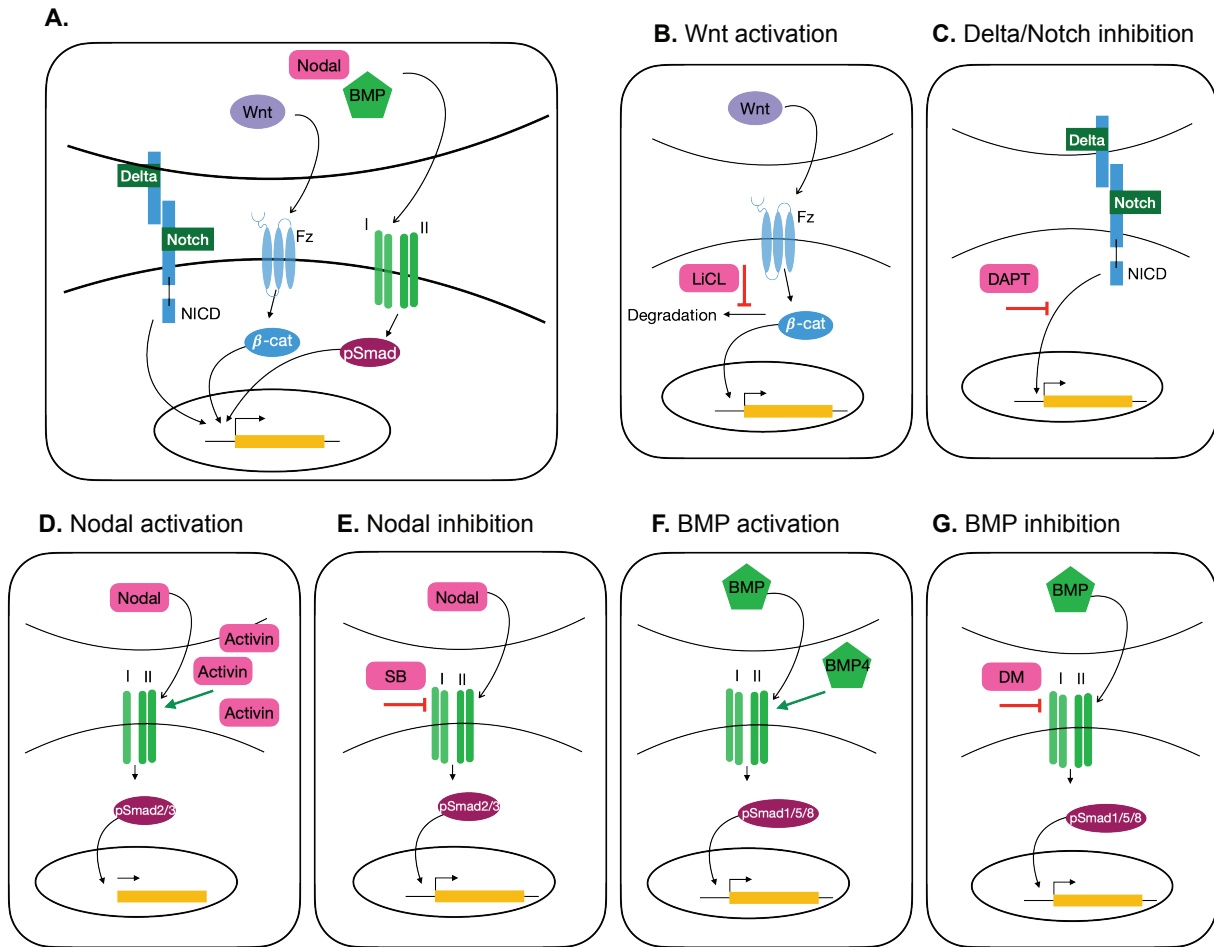


Figure 2. Cell signaling mechanisms and signaling perturbation using small molecules. (A) Summary of active Wnt, BMP, Nodal, and Delta/Notch cell signaling pathways all of which can regulate gene expression. (B) Lithium chloride (LiCl) activates Wnt signaling by inhibiting GSK3 activity leading to stabilization of β -catenin in the cytoplasm, β -catenin can then translocate to the nucleus (Zhang et al., 2003). (C) DAPT is a γ -secretase inhibitor, preventing cleavage of the Notch intracellular domain (NICD) and blocking Notch signaling (Geling et al., 2002). (D) Activin protein binds to the Nodal receptors and initiates downstream signaling without the need for coreceptor binding (Luo and Su, 2012; Shen, 2007). (E) SB-431542 selectively inhibits the Nodal type I receptors, ALK5, ALK4, and ALK7, preventing signaling (Inman et al., 2002). (F) Treatment with BMP ligands results in upregulated BMP signaling which is conveyed by the downstream effector phosphorylated Smad1/5/8 (Weiss and Attisano, 2013). (G) Dorsomorphin (DM) is a BMP antagonist which inhibits the BMP type I receptor preventing phosphorylation of BMP Smads and thus downstream signaling (Luo and Su, 2012; Yu et al., 2008).

Signaling by Wnt3 (E5.5) conveys competence to the proximal epiblast cells to respond to BMP4 signaling (Ohinata et al., 2009; Saitou and Yamaji, 2010). Nodal

signaling in the epiblast is required for the establishment of the anterior-posterior and proximal-distal axes. Through Smad2, Nodal maintains anterior end identity and restricts the PGC niche. However, Nodal also functions to maintain PGC specification through maintenance of the pluripotency gene Nanog and BMP4 expression (Saitou and Yamaji, 2010; Senft et al., 2019). PGC specification in the mouse requires BMP4 signaling from the extraembryonic ectoderm in a dose-dependent fashion (Ohinata et al., 2009). B-lymphocyte induced-maturation protein 1, (Blimp1, also known as PRDM1), a PR-domain containing transcriptional regulator and a marker of early mouse PGCs, is expressed in the proximal epiblast cells in response to BMP4 signaling. Blimp1 is required for PGC maintenance by repressing somatic mesodermal fate (Ohinata et al., 2009; Ohinata et al., 2005). As PGC specification proceeds, a regulatory event takes place that results in loss of response to BMP signaling (at E6.5, E7.5) (Senft et al., 2019). This is an important step in differentiating between the PGCs and somatic cells that will become mesoderm since BMP also regulates mesoderm development. The PGCs differentiate from their neighboring somatic cells by repressing several homeobox genes and the BMP signaling effectors Smad1 and Smad4 (Saitou et al., 2003). Single cell RNA sequencing analysis of the PGCs from early embryonic stages showed that the PGCs are derived from a population of mesodermal cells. That is, the early germ cells (E7.5) express mesodermal markers like Brachyury (T) before they are eventually repressed (Saitou et al., 2003).

Inductive germline specification in non-mammalian species

In the cricket, *Gryllus bimaculatus* of the arthropod phyla, BMP was shown to regulate expression of the germ cell factor, Piwi. Reducing transcript levels of the *Gryllus* ortholog of Blimp1, *Gb-Blimp1*, and the downstream BMP signaling effector, *Gb-Mad*, by RNA interference (RNAi) leads to a loss of Piwi expression and a reduction of PGCs (Nakamura and Extavour, 2016). This suggests a conserved role for BMP signaling and Blimp1 in germ cell specification. Whether Blimp1 transcriptional regulation of Piwi is direct or indirect is unknown. Moreover, the downstream regulator Mad was found to regulate Blimp1, however, whether that regulation is direct or indirect also remains unknown. BMP signaling is also required for germline gene expression in the hemichordate *Ptychodera flava*. Inhibiting BMP signaling with the small molecule DMH1 results in loss of Nanos, Vasa, and Piwi1 expression. Expression of phospho-Smad1/5/8, which is seen in Vasa-positive cells is also lost (Lin et al., 2021). Notably, BMP is not sufficient to induce ectopic PGC development supporting the multi-signal concept of inductive germline formation.

In the axolotl *Ambystoma mexicanum*, an amphibian, PGCs are derived from unspecified mesoderm in response to signaling from vegetal cells at blastula stage. Both Wnt and BMP seem to be involved in this process. While the cell signals coming from the vegetal end of the blastula have not been fully characterized, PGC specification was induced in isolated animal caps when axolotl embryos were co-injected with eFGF (embryonic fibroblast growth factor) and BMP4 RNA and not in the animal caps of control embryos injected with water (Johnson et al., 2003). The axolotl

homologue of Brachyury, a well conserved mesodermal marker, *Axbra*, is seen in all cells of the mesodermal region known to include PGC precursors, suggesting that the PGC precursors express Brachyury and are specified from early mesoderm. Similarly, the axolotl homolog of Wnt8, *Axwnt-8*, is expressed uniformly in the mesodermal region known to include PGC precursors (Bachvarova et al., 2001). Further studies are necessary to understand whether *Axwnt-8* plays a role in inducing mesoderm and/or PGC specification. Investigating the mechanism by which BMP4, whether independently or in combination with eFGF, induces PGC specification could reveal whether BMP signaling is a conserved element of inductive germline specification.

Techniques used to study germline development

Analysis of whole embryos contributed to our understanding of how the germ cells are specified in a particular site within the embryo and how they migrate to the gonads. We can observe changes in the expression patterns of the germ cell genes considering the context of the intact embryo as we are not disrupting immediate nor far-reaching cell interactions. Analysis of single cells makes a wealth of new information available to us because we can gain insight into whether there is diversity within a particular cell identity and document cell fate transitions in greater detail. *In vitro* approaches are also an important aspect of germline studies. A cell culture setting yields a controlled environment for cells of interest. In addition, *in vitro* cell culture facilitates certain experimental tests where access to a cell type is prohibitive.

Single-cell RNA sequencing

The first single-cell RNA sequencing (scRNA-seq) analysis of mouse PGCs was completed by dissecting the region of PGC development in the embryo, dissociating those cells, and manually choosing at random 100 cells for cDNA development and analysis. This study discovered two novel genes, *Stella* and *Fragilis*, expressed in the PGC precursors, as well as molecular differences between the PGCs and their somatic neighbors (Saitou et al., 2003). Since this study was published, the field has gained access to several robust, high-throughput methods to analyze RNA captured from single cells (Haque et al., 2017). Single cell transcriptomes for PGCs have been captured in zebrafish, *Xenopus*, and sea urchin (Briggs et al., 2018; Foster et al., 2020; Wagner et al., 2018). Transcriptomic changes of germline cells across several stages of development in mouse testes and ovaries have also been characterized using scRNAseq (Tan et al., 2020; Wang et al., 2020). Single-cell analysis of male and female human fetal gonads identified reciprocal expression of Nodal and BMP signaling ligands and their receptors between gonadal niche cells and the PGCs (Li et al., 2017). The relationship and signaling between germ cells and their neighboring cells are essential in the *Drosophila* male gonad. Knockdown of Delta expression in the germ cells and Notch expression in the neighboring cyst cells causes significant loss in germ cell number (Ng et al., 2019). Delta/Notch signaling takes place between two adjacent cells. One cell expressing a Delta ligand can signal to a neighboring cell expressing the Notch receptor. This binding results in cleavage of the Notch intracellular domain (NICD) by γ -secretase and the NCID then translocates to the nucleus where it regulates gene expression (Figure 2A) (Kopan, 2012). The signaling events involved in germ cell gene

regulation, and their targets, are not fully understood. If and how the PGCs signal to other cells is also an interesting question. It is possible that cell signaling between the PGCs and their niche plays a role in PGC maintenance in the mouse. Having detailed the gene expression landscape of the PGCs, we are now poised to pick apart regulatory events of the PGC developmental trajectory through scRNAseq analysis of mutants.

In vitro stem cell culture

Cell culture studies have made significant contributions to our understanding of PGC specification. This system grants greater freedom for controlling the environment through changes in media, and drug treatments, for example, as well as the timing of these manipulations. Cell culture studies revealed that all epiblast cells have the ability to express Blimp1 in response to BMP4 signaling, emphasizing the importance of restriction in PGC development. Further, BMP4 alone can induce Blimp1 expression and PGC specification by signaling through the receptors ALK3 and ALK5, and the transcriptional regulators Smad1 and Smad5 (Ohinata et al., 2009). More recently, mature oocytes were produced using induced pluripotent stem cells (iPS) derived from mouse fibroblasts, and a mouse embryonic stem cell line (ES). The iPSCs were differentiated into epiblast-like cells (EpiLC) through treatment with Activin and BMP4, and PGC-like cells (PGCLC) were aggregated with female mouse gonadal somatic cells. The mouse PGCLCs were differentiated, grown, and matured in an *in vitro* setting; healthy embryos and pups were derived from these oocytes through IVF (Hikabe et al., 2016).

Human PGC-like cells (hPGCLC) have also been induced from human iPS cells (hiPSC) when cultured with BMP4, epidermal growth factor (EGF), stem cell factor (SCF), and leukemia inhibitory factor (LIF), based on conditions known to induce EpiLC into mouse PGCLCs (Hikabe et al., 2016). Studies revealed that hiPSC treated with Activin A and a Wnt agonist (CHIR99021) produced epithelial cells expressing naive pluripotency genes and genes of a mesodermal cell state. The incipient mesoderm-like cells (iMELC) were later induced to PGC-like cells more quickly and at a greater rate compared to direct induction to PGCLC from hiPSC (Sasaki et al., 2015). A roadmap of human germline specification has been outlined, drawing from research in both *in vivo* and *in vitro* settings. It shows expression of pluripotency markers in human embryonic stem cells (hESC) and BMP and Wnt3 signaling leading to PGC specification, followed by Vasa expression in migrating PGCs (Hayashi et al., 2012). A branch in the pathway shows *in vitro* development of human embryonic germ cells (hEGC), this can now be updated with our knowledge of an iMELC state before the development of hEGC and will continue to be updated as more discoveries are made.

With various models existing for inductive germline studies, some may wonder how sea star studies supplement these efforts. The earliest functional evidence of the inductive mechanism in an invertebrate was reported in cricket (Donoughe et al., 2014). Crickets are a basally branching insect and germ cell studies in this animal further supported the hypothesis that the inductive mechanism is ancestral to insects (Extavour and Akam, 2003). If we consider germ cell studies in axolotl we gain an understanding of this cell type in chordates. The discovery of an inductive mechanism of germ cell

specification in the axolotl brought new knowledge that a signaling based method is an ancestral trait in vertebrates and that the inherited mechanism evolved in frogs such as *Xenopus* (Johnson et al., 2001). Currently, our knowledge of this mechanism in axolotl is limited beyond the reported roles for FGF and BMP signaling and derivation of the PGCs from mesodermal precursors (Chatfield et al., 2014). The mouse is a prominent model for studying inductive germ cell specification in a mammal and work in this animal model has detailed the involvement of signaling mechanisms in greater detail (Saitou and Yamaji, 2010). A relevant detail of both axolotl and mouse studies is that Nanos and Vasa expression is first detected late in development as the germ cells reach the developing gonads (Bachvarova et al., 2004; Fujiwara et al., 1994; Tsuda et al., 2003). Investigating how these genes are regulated requires extracting the PGCs from gonads or full embryos in the case of the mouse. All these species have their own advantages based on their unique evolutionary context.

Still, the sea star offers unique strengths and can certainly fill gaps in our understanding of the germ cells. The sea star produces an abundance of embryos and we can quickly measure changes in Nanos and Vasa expression and the progression of germ cell fate since these genes are seen early in embryogenesis. Importantly, this model is both ancestral and closely related to chordates (Figure 3) (Extavour and Akam, 2003; Wessel et al., 2014). Therefore, there is an opportunity to identify what regulatory events have been conserved or newly arose through evolution by comparing sea star and mouse studies. Moreover, we can compare findings in the sea star with another

echinoderm species the sea urchin which shares many similarities in embryonic development.

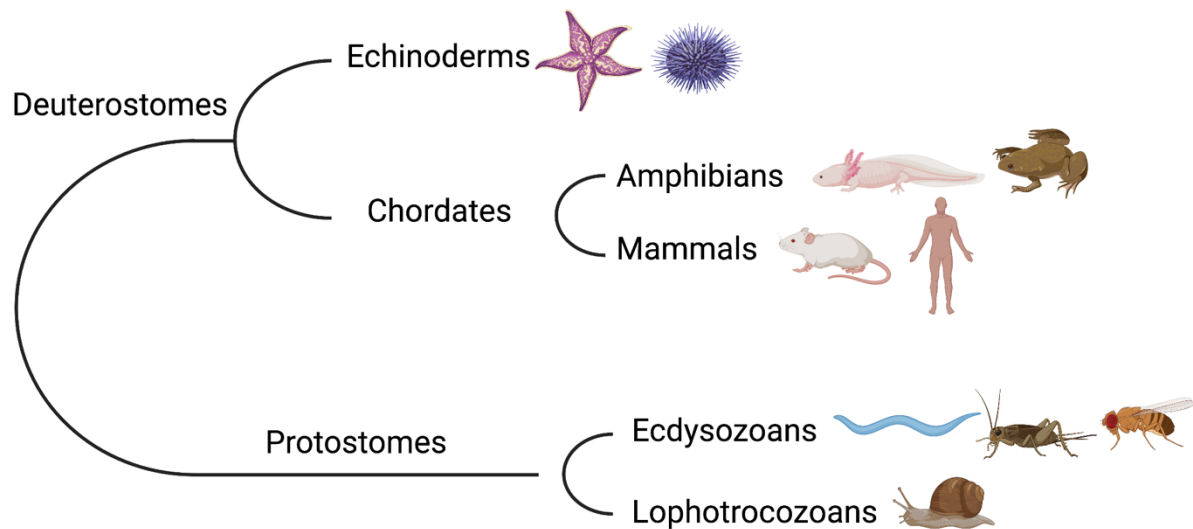


Figure 3. Phylogenetic relationship among leading models for germ cell specification studies. Modified from Wessel et al., 2014. Created with [BioRender.com](https://www.biorender.com).

Having identified the germ cell specification method used by various species and which method is ancestral to a lineage, we can begin to ask the question of what may be the evolutionary advantage of employing either method in a species specific basis. Several hypotheses have been proposed for both the evolution of and evolutionary advantage of the inductive and inherited mechanisms of germ cell specification (Whittle and Extavour, 2017). Species in vastly different environments employ the same mechanism for germ cell specification; the sea star, a satellite spawning marine animal, employs the inherited mechanisms as does the mouse, a land animal that fertilizes internally. Differences between species which range from environment to mating behaviors make it challenging to draw generalized conclusions about the advantage of one germ cell specification method over the other. For example, acquisition of the inherited mechanism has been linked to accelerated evolution in vertebrates (Evans et

al., 2014). The hypothesis related to this being that early specification of the germline alleviates selective pressure constraints on the embryo and results in rapid protein evolution. However, studies that consider protein evolution and diversification across both vertebrates and invertebrates report no link between inheritance and accelerated evolution (Whittle and Extavour, 2016). Indeed, the sea urchin (inherited) embryo forms micromeres and skeletogenic cells which are absent in the sea star (inductive) (Shashikant et al., 2018). In the case of these two echinoderm species, it is possible that an inherited mechanism led to evolution of new cell types. Further research of germ cell specification in echinoderms paired with gene regulatory network studies and mutation rate analyses could support or dispute what we understand about germ cell specification methods and their evolutionary impact in other animals.

Studying the ancestral mode of germ cell specification in the sea star

Previous work identified that the sea star, *Patiria miniata*, follows the inductive mode of germline formation (Fresques and Wessel, 2018). The conserved germline factors Nanos, Vasa, and Piwi do not show specific localization in the oocyte or the early embryo but rather are restricted later in development to a left-sided pouch, the posterior enterocoel (PE) (Fresques et al., 2016; Fresques et al., 2014). The PE is posited to be the source of the germline as it is the source of restricted and specific germline gene enrichment in the sea star larva. Moreover, removal of the PE results in significantly fewer germline cells in larvae (Inoue et al., 1992). This PE removal via micropipetting was performed without labeling any cells of the PE. Thus, it is possible that not all presumptive germline cells were removed. Additionally, both treated and

control larvae in this experiment failed to metamorphose. Labeling PE cells to detect complete removal of those cells remains an important future experiment, along with following larvae through metamorphosis to test whether germline cells are present in the juveniles and adults.

Vasa transcripts are present in the oocyte and the early embryo, while Nanos is first detected by in situ hybridization in mid-gastrula (Fresques et al., 2016; Fresques et al., 2014). Both Nanos and Vasa are expressed as a vegetal ring and at the tip of the archenteron in early gastrula. An opposing Nodal is responsible for restricting germ cell factor gene expression to the PE (Fresques and Wessel, 2018). Nodal is expressed first in the ventral side of the embryo, then in the ventral right side of the developing gut and ectoderm. As a result, Nanos and Vasa are restricted dorsally and to the left side of the developing gut, where the PE forms. Nodal signaling both inhibits Nanos and Vasa transcription and induces degradation of existing Nanos and Vasa transcripts where it is expressed. Nodal also inhibits tissue morphogenesis, preventing the formation of more than one pouch in the gut (Fresques and Wessel, 2018).

My work adds to current knowledge of inductive germline specification by identifying additional cell signaling mechanisms that regulate Nanos and Vasa mRNA expression in the sea star. Cultured sea star embryos can be treated with cell signaling perturbants by adding them directly to the sea water. This method allowed me to test a role for various signaling mechanisms, at different developmental stages. My research shows that Wnt signaling regulates Vasa mRNA expression, and Delta/Notch regulates

both Nanos and Vasa mRNA in gastrulae (Chapter 4) (Foster et al., 2022). I

hypothesize that Wnt, Delta/Notch, and Nodal signaling act at the same time to maintain Nanos and Vasa expression in the early gastrula and restrict it to the left side of the embryo.

To test whether cell signaling is required for germ cell gene expression, we dissociated embryos, removing cell signaling by suspending individual cells in culture and concomitantly greatly diluting cell signaling ligands. Dissociated cells, like intact embryos, can be treated with small-molecule cell signaling perturbants. Embryos were cultured until blastula stage, at which point Vasa expression is detected but Nanos is not yet detected. Embryos dissociated at blastula stage and cultured for up to 18 hours show an upregulation of Nanos mRNA abundance compared to intact gastrula. Vasa mRNA in dissociated cells is downregulated compared to intact gastrula. For Vasa, even in the absence of Nodal restriction, cell signaling is necessary to maintain its expression while Nanos does not show this same pattern. The increase in Nanos mRNA abundance is lost with incubation with a Nodal signaling agonist, suggesting a direct effect of Nodal on Nanos-positive cells and revealing an opposing trend between Nanos and Vasa (Chapter 4) (Foster et al., 2022). In summary, the overlapping pattern of Nanos and Vasa results from distinct regulation of each factor. Further studies will reveal the involvement of specific receptors and downstream effectors in the regulation of these germline genes.

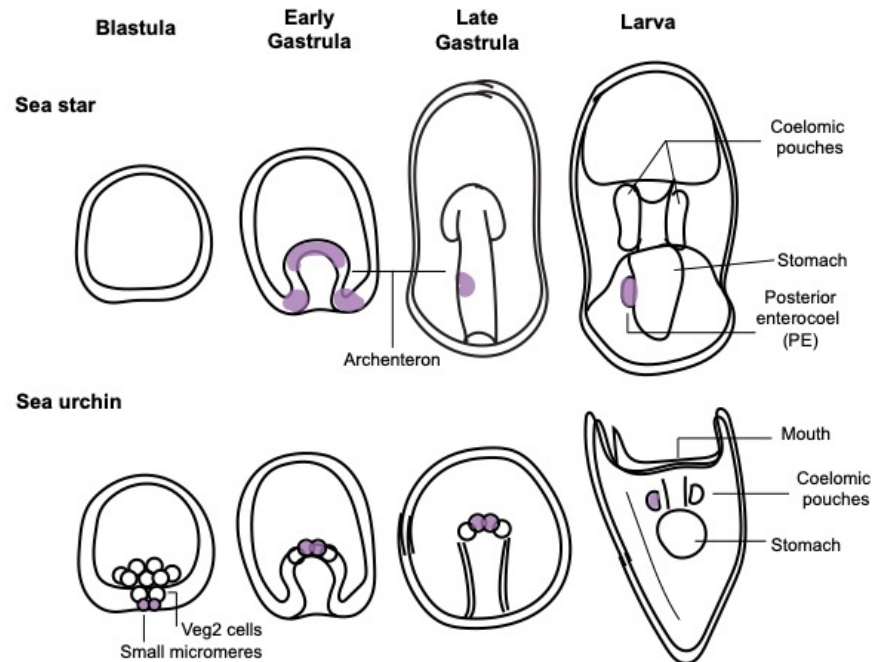


Figure 4. Germline formation in echinoderms. The sea star follows the inductive mechanism. Germ cell gene expression is sequentially restricted to the region of PE formation, site of the presumptive PGC precursors. The sea urchin follows the inherited mechanism. The presumptive PGCs which selectively express the germline genes are specified by the blastula stage. Germline precursor cells and germline gene expression region shown in purple.

Germline specification in the sea urchin

The sea urchin shares the characteristics that make the sea star an accessible model for the inductive mechanism such as rapid embryonic development, and early expression of the germ cell genes. The urchin, however, employs the inherited mechanism for germ cell specification and this allows us to compare what we learn in the sea star studies to what we know and continue to investigate in the sea urchin. Germ cell specification occurs early in sea urchin, relative to sea star, through the inherited mechanism. Two asymmetric cell divisions, the fourth and fifth cleavage divisions, result in the formation of the micromeres and then small micromeres, the

source of the PGCs, at the vegetal end of the embryo (Figure 4). While the fate of the small micromeres has not been followed via lineage tracing, data suggests their role as the source of the germline cells (Yajima and Wessel, 2011). When formed, the small micromeres accumulate germline determinants including Nanos, Vasa, and Seawi (Juliano et al., 2006). The site of small micromere formation is also where gastrulation takes place. When the vegetal plate invaginates, the small micromeres de-epithelialize into the blastocoel and are passively transported as the vegetal epithelium continues to invaginate, forming the gut. The small micromeres remain at the tip of the archenteron until they migrate to the left and right coelomic pouches (Yajima and Wessel, 2012). Derivatives of the small micromeres and Veg2 mesoderm lineage contribute to the coelomic pouches, and the left coelomic pouch is the site of adult rudiment formation (Cameron and Davidson, 1991; Cameron et al., 1987). The Veg2 mesoderm of the macromere lineage is formed at the fifth cell division, adjacent to the micromeres. The PGCs in the right coelomic pouch are lost to apoptosis (Luo and Su, 2012) and expression of the germ cell factors Nanos, Vasa, and Seawi remains in the left coelomic pouch (Juliano et al., 2006). Removal of the small micromeres results in adults which lack gametes but otherwise show normal development, including gonad development (Yajima and Wessel, 2011).

There are three Nanos homologs in the sea urchin. Nanos1 is present in the ovary, while Nanos2 is selectively enriched in the small micromeres when these cells form and is required for rudiment formation, and Nanos 3 is expressed in the small micromere lineage later in gastrulation (Juliano et al., 2010). Vasa mRNA and protein

are first expressed broadly in the early embryo before becoming enriched in the small micromeres (Voronina et al., 2008). Seawi is one of two Piwi homologs identified in the sea urchin; it expresses a common Piwi RNA binding motif (Rodriguez et al., 2005). Seawi RNA and protein, first expressed broadly in the early embryo, become restricted to the PGC precursor cells (Juliano et al., 2006). Seawi regulates Vasa protein expression in the early embryo; loss of Seawi results in loss of PGC descendants through apoptosis (Yajima et al., 2014).

Notably, removal of the parent cells of the small micromeres results in adults with developed gametes. Following micromere removal, the embryos show a broad upregulation of Vasa protein (Yajima and Wessel, 2011). We hypothesize that a compensatory mechanism involving Vasa upregulation is initiated following micromere removal. Moreover, secondary Nanos expression is detected outside of the small micromeres after gastrulation, in the adjacent somatic Veg2 mesoderm. The regulation of Nanos transcript expression varies between these two cell types. β -catenin, the downstream effector of canonical Wnt signaling, regulates Nanos transcription in the small micromeres. Delta/Notch signaling, through the forkhead transcription factor, FoxY, is essential for Nanos expression in the Veg2 cells. FoxY knockdown via morpholino results in loss of Nanos expression specifically in the Veg2 cells (Oulhen et al., 2019). Whether this secondary Nanos expression in the Veg2 cells plays a role in reconstituting the germline remains an interesting, open question. In another sea urchin species, *Hemicentrotus pulcherrimus*, Nanos expression is seen in other cell lineages following micromere deletion (Fujii et al., 2009). Further testing may reveal whether loss

of the micromeres activates a compensatory mechanism for germ cell specification that relies on cell signaling.

The forkhead (Fox) family of transcription factors has been identified in several metazoans including humans and obtain a highly conserved DNA-binding domain. Phylogenetic analysis of all forkhead domains identified in the sea urchin (*S. purpuratus*), *Drosophila*, and human genomes showed that the sea urchin FoxY did not belong to any vertebrate class (Tu et al., 2006). However, it was previously reported as FoxC-like and FoxQ-like in another urchin species (Hibino et al., 2006; Ransick et al., 2002). FoxY is likely not an echinoderm specific transcription factor but future genome assemblies for echinoderm species will grant a clearer understanding of related genes in other metazoans including other echinoderms.

Single-cell RNA sequencing analysis of echinoderm putative PGCs

We employed a scRNA-seq approach to perform transcriptomic analysis of early sea star and sea urchin development to identify the PGC precursor cells based on their expression of Nanos and Vasa. We identified and characterized expression for 22 cell states across 8 sea urchin developmental stages, from 8-cell stage to late gastrula. Changing clustering parameters reveals more cell types of interest such as different regions of the gut, pigment cells, and neuronal cell types (Foster et al., 2020). Coexpression of germ cell genes Nanos, Vasa, and Seawi was detected in two clusters, one of which we identify as the Veg2 based on expression of Sox4/C, a transcription factor known to be expressed in the Veg2 lineage. In the germline cluster, we detect a

decrease in Cyclin B, Cyclin A, and CNOT6 transcripts with development, consistent with the sea urchin PGCs becoming quiescent after their formation at the 32-cell stage (Chapter 3) (Foster et al., 2020; Oulhen et al., 2017; Oulhen and Wessel, 2017). This particular study was a catalyst for the analysis of other cell types of the sea urchin embryo as well. We discovered unique gene expression and distinct populations of the pigment cells, a cell type with a fundamental role in the sea urchin's innate immune system (Appendix C). Another collaboration aimed to further analyze the Brachyury-positive cells of the ectoderm and revealed co-expression of organizer-related genes, Nodal, BMP2/4, chordin, and ADMP, emphasizing a role for these cells in body plan formation (Appendix D).

We identified 21 cell states across 6 early sea star developmental stages via scRNAseq (Foster et al., 2022). We focused our analysis on the mid-gastrula stage, where Nanos and Vasa are both expressed and will later be restricted. Nanos and Vasa co-expression is detected in regions identified as the archenteron and hindgut, as seen previously by in situ hybridization (Fresques and Wessel, 2018). The forkhead box transcription factor FoxY3 stands out as being both enriched and specific to the Nanos, Vasa-positive cells (Foster et al., 2022). Additionally, these putative PGC precursor cells express mesodermal markers at mid-gastrula stage. Future studies will be essential for revealing FoxY3 gene targets and whether it plays a role in regulating PGC specification. scRNA-seq analysis of early larvae in which the PE is formed, and the germ cell factors are restricted to this region, will allow for comparison of gene

expression between early and later stage putative PGC precursor cells further revealing transitions in the development of this cell type.

Summary

Analysis of several species across the tree of life reveals that there is no evidence of the inherited mechanism in the most basal branching lineages to metazoans, such as jellyfish and hydra. Additionally, instances where entire phyla follow the inherited mechanism are rare; the inherited mechanism appears to have evolved independently in derived lineages within phyla (Extavour and Akam, 2003). These studies will help reveal mechanisms of germ cell specification conserved through evolution in sexually reproducing organisms. Our work also fuels essential comparative studies within echinoderms (sea urchin and sea star) to understand evolutionary divergence across species. Another motivation for investigating PGC specification is to better understand, diagnose, and potentially treat human diseases that arise from abnormal germ cell specification or expression, including infertility and cancer (Zhang et al., 2020). Reconstitution of PGC specification in an *in vitro* setting is a promising opportunity for treatment for those struggling to start or grow a family (Hayashi et al., 2012).

My research set out to answer the question, “What is a germline and how do I get one?” Analysis of the germ cell transcriptome is one approach to investigating what makes these cells unique. Following the inductive mechanism, you “get” a germline through cell signaling. Identifying what cell signaling mechanisms affect germline gene

expression brings us closer to identifying what downstream transcription factors are involved, and how this regulation evolved in other species. In the early sea star embryo, before the germ cell genes are restricted to the PE, the Nanos and Vasa-positive cells express mesodermal marker genes, like their neighbors. This is a similarity shared with PGC precursors in the mouse, axolotl, and cricket pointing to conservation of a mesodermal precursor state (Chatfield et al., 2014; Donoughe et al., 2014). Another similarity between sea star and mouse germ cell specification is present in signaling regulation. Wnt has a conserved role as an activating signal and Nodal has a conserved role in restricting the germline niche. Wnt signaling may be one of the earliest signals for inductive germline formation. BMP signaling is reported again and again as a regulator of germline gene expression and accordingly germ cell specification (Donoughe et al., 2014; Johnson et al., 2003; Lin et al., 2021; Ohinata et al., 2009). Preliminary data suggests that BMP signaling may regulate Vasa mRNA expression in the sea star (Foster et al., 2020). Testing a role for BMP in germline specification in the sea star is an important future direction and another point of comparison in inductive germ cell specification between mouse and sea star (Figure 2).

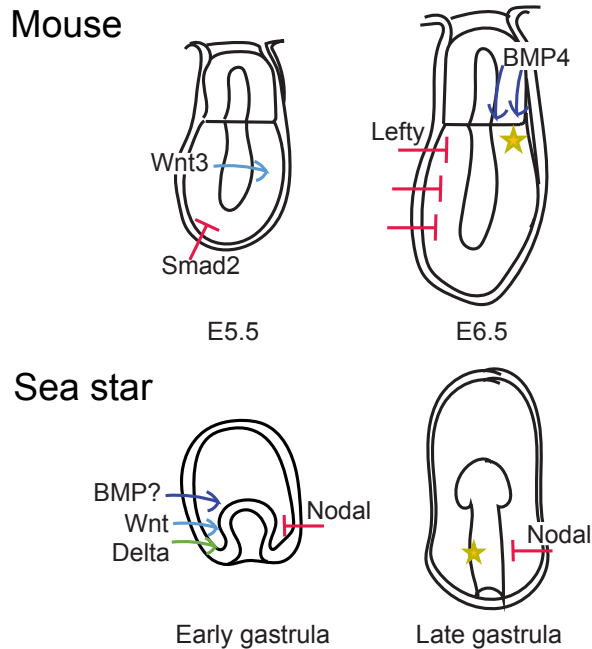


Figure 5. Comparison of inductive germ cell specification between mouse and sea star. Model depicting similarities and differences between mouse and sea star inductive germline specification.

Our findings on germ cell gene regulation in sea star contrast how *Nanos* and *Vasa* are enriched in the sea urchin small micromeres. The sea urchin, which follows the inherited mode of germ cell specification, shows enriched *Nanos* mRNA expression in the PGCs early in development compared to the sea star. This has been shown via RNA in situ hybridization and now scRNAseq analysis as well (Foster et al., 2022; Foster et al., 2020; Fresques et al., 2016; Juliano et al., 2006). In addition, we find many points of comparison in the regulation of the germ cell genes in both the sea star and sea urchin revealing both conserved nodes and regulatory evolution. Our studies of sea star PGC precursor cells enable us to perform deeper comparative studies of germ cell specification in echinoderms and investigate the evolution of germ cell specification across the tree of life, from hydra to mammals.

References

- Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K., Kobayashi, S., 1999. Maternal Pumilio acts together with Nanos in germline development in *Drosophila* embryos. *Nat Cell Biol* 1, 431-437.
- Bachvarova, R.F., Masi, T., Drum, M., Parker, N., Mason, K., Patient, R., Johnson, A.D., 2004. Gene expression in the axolotl germ line: *Axdazl*, *Axvh*, *Axoct-4*, and *Axkit*. *Dev Dyn* 231, 871-880.
- Bachvarova, R.F., Masi, T., Hall, L., Johnson, A.D., 2001. Expression of *Axwnt-8* and *Axszi* in the urodele, axolotl: comparison with *Xenopus*. *Dev Genes Evol* 211, 501-505.
- Barrow, J.R., Howell, W.D., Rule, M., Hayashi, S., Thomas, K.R., Capecchi, M.R., McMahon, A.P., 2007. Wnt3 signaling in the epiblast is required for proper orientation of the anteroposterior axis. *Dev Biol* 312, 312-320.
- Briggs, J.A., Weinreb, C., Wagner, D.E., Megason, S., Peshkin, L., Kirschner, M.W., Klein, A.M., 2018. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* 360.
- Cameron, R.A., Davidson, E.H., 1991. Cell type specification during sea urchin development. *Trends Genet* 7, 212-218.
- Cameron, R.A., Hough-Evans, B.R., Britten, R.J., Davidson, E.H., 1987. Lineage and fate of each blastomere of the eight-cell sea urchin embryo. *Genes Dev* 1, 75-85.
- Chatfield, J., O'Reilly, M.A., Bachvarova, R.F., Ferjentsik, Z., Redwood, C., Walmsley, M., Patient, R., Loose, M., Johnson, A.D., 2014. Stochastic specification of primordial germ cells from mesoderm precursors in axolotl embryos. *Development* 141, 2429-2440.
- Donoughe, S., Nakamura, T., Ewen-Campen, B., Green, D.A., 2nd, Henderson, L., Extavour, C.G., 2014. BMP signaling is required for the generation of primordial germ cells in an insect. *Proc Natl Acad Sci U S A* 111, 4133-4138.
- Ehrlich, M., 2016. Endocytosis and trafficking of BMP receptors: Regulatory mechanisms for fine-tuning the signaling response in different cellular contexts. *Cytokine Growth Factor Rev* 27, 35-42.
- Evans, T., Wade, C.M., Chapman, F.A., Johnson, A.D., Loose, M., 2014. Acquisition of germ plasm accelerates vertebrate evolution. *Science* 344, 200-203.
- Extavour, C.G., Akam, M., 2003. Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* 130, 5869-5884.
- Foster, S., Oulhen, N., Fresques, T., Zaki, H., Wessel, G., 2022. Distinct mechanisms of germ cell factor regulation for an inductive germ cell fate. *bioRxiv*, 2022.2002.2004.479164.
- Foster, S., Oulhen, N., Wessel, G., 2020. A single cell RNA sequencing resource for early sea urchin development. *Development* 147.
- Fresques, T., Swartz, S.Z., Juliano, C., Morino, Y., Kikuchi, M., Akasaka, K., Wada, H., Yajima, M., Wessel, G.M., 2016. The diversity of nanos expression in echinoderm embryos supports different mechanisms in germ cell specification. *Evol Dev* 18, 267-278.
- Fresques, T., Zazueta-Novoa, V., Reich, A., Wessel, G.M., 2014. Selective accumulation of germ-line associated gene products in early development of the

- sea star and distinct differences from germ-line development in the sea urchin. *Dev Dyn* 243, 568-587.
- Fresques, T.M., Wessel, G.M., 2018. Nodal induces sequential restriction of germ cell factors during primordial germ cell specification. *Development* 145.
- Fujii, T., Sakamoto, N., Ochiai, H., Fujita, K., Okamitsu, Y., Sumiyoshi, N., Minokawa, T., Yamamoto, T., 2009. Role of the nanos homolog during sea urchin development. *Dev Dyn* 238, 2511-2521.
- Fujiwara, Y., Komiya, T., Kawabata, H., Sato, M., Fujimoto, H., Furusawa, M., Noce, T., 1994. Isolation of a DEAD-family protein gene that encodes a murine homolog of *Drosophila vasa* and its specific expression in germ cell lineage. *Proc Natl Acad Sci U S A* 91, 12258-12262.
- Geling, A., Steiner, H., Willem, M., Bally-Cuif, L., Haass, C., 2002. A gamma-secretase inhibitor blocks Notch signaling in vivo and causes a severe neurogenic phenotype in zebrafish. *EMBO Rep* 3, 688-694.
- Ginsburg, M., Snow, M.H., McLaren, A., 1990. Primordial germ cells in the mouse embryo during gastrulation. *Development* 110, 521-528.
- Haque, A., Engel, J., Teichmann, S.A., Lonnberg, T., 2017. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* 9, 75.
- Hashimoto, Y., Maegawa, S., Nagai, T., Yamaha, E., Suzuki, H., Yasuda, K., Inoue, K., 2004. Localized maternal factors are required for zebrafish germ cell formation. *Dev Biol* 268, 152-161.
- Hay, B., Jan, L.Y., Jan, Y.N., 1988. A protein component of *Drosophila* polar granules is encoded by *vasa* and has extensive sequence similarity to ATP-dependent helicases. *Cell* 55, 577-587.
- Hayashi, Y., Saitou, M., Yamanaka, S., 2012. Germline development from human pluripotent stem cells toward disease modeling of infertility. *Fertil Steril* 97, 1250-1259.
- Heldin, C.H., Lu, B., Evans, R., Gutkind, J.S., 2016. Signals and Receptors. *Cold Spring Harb Perspect Biol* 8, a005900.
- Hibino, T., Ishii, Y., Levin, M., Nishino, A., 2006. Ion flow regulates left-right asymmetry in sea urchin development. *Dev Genes Evol* 216, 265-276.
- Hikabe, O., Hamazaki, N., Nagamatsu, G., Obata, Y., Hirao, Y., Hamada, N., Shimamoto, S., Imamura, T., Nakashima, K., Saitou, M., Hayashi, K., 2016. Reconstitution in vitro of the entire cycle of the mouse female germ line. *Nature* 539, 299-303.
- Ikenishi, K., Nakazato, S., Okuda, T., 1986. Direct Evidence for the Presence of Germ Cell Determinant in Vegetal Pole Cytoplasm of *Xenopus laevis* and in a Subcellular Fraction of It. *Development, Growth & Differentiation* 28, 563-568.
- Illmensee, K., Mahowald, A.P., 1974. Transplantation of posterior polar plasm in *Drosophila*. Induction of germ cells at the anterior pole of the egg. *Proc Natl Acad Sci U S A* 71, 1016-1020.
- Inman, G.J., Nicolas, F.J., Callahan, J.F., Harling, J.D., Gaster, L.M., Reith, A.D., Laping, N.J., Hill, C.S., 2002. SB-431542 is a potent and specific inhibitor of transforming growth factor-beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. *Mol Pharmacol* 62, 65-74.

- Inoue, C., Kiyomoto, M., Shirai, H., 1992. Germ Cell Differentiation in Starfish: The Posterior Enterocoel as the Origin of Germ Cells in *Asterina pectinifera*. *Development, Growth & Differentiation* 34, 413-418.
- Irie, N., Tang, W.W., Azim Surani, M., 2014. Germ cell specification and pluripotency in mammals: a perspective from early embryogenesis. *Reprod Med Biol* 13, 203-215.
- Johnson, A.D., Bachvarova, R.F., Drum, M., Masi, T., 2001. Expression of axolotl DAZL RNA, a marker of germ plasm: widespread maternal RNA and onset of expression in germ cells approaching the gonad. *Dev Biol* 234, 402-415.
- Johnson, A.D., Crother, B., White, M.E., Patient, R., Bachvarova, R.F., Drum, M., Masi, T., 2003. Regulative germ cell specification in axolotl embryos: a primitive trait conserved in the mammalian lineage. *Philos Trans R Soc Lond B Biol Sci* 358, 1371-1379.
- Juliano, C.E., Voronina, E., Stack, C., Aldrich, M., Cameron, A.R., Wessel, G.M., 2006. Germ line determinants are not localized early in sea urchin development, but do accumulate in the small micromere lineage. *Dev Biol* 300, 406-415.
- Juliano, C.E., Yajima, M., 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.
- Kikuchi, A., Yamamoto, H., Kishida, S., 2007. Multiplicity of the interactions of Wnt proteins and their receptors. *Cell Signal* 19, 659-671.
- Klattenhoff, C., Theurkauf, W., 2008. Biogenesis and germline functions of piRNAs. *Development* 135, 3-9.
- Kloc, M., Bilinski, S., Chan, A.P., Allen, L.H., Zearfoss, N.R., Etkin, L.D., 2001. RNA localization and germ cell determination in *Xenopus*. *Int Rev Cytol* 203, 63-91.
- Kobayashi, S., Yamada, M., Asaoka, M., Kitamura, T., 1996. Essential role of the posterior morphogen nanos for germline development in *Drosophila*. *Nature* 380, 708-711.
- Kopan, R., 2012. Notch signaling. *Cold Spring Harb Perspect Biol* 4.
- Koprunner, M., Thisse, C., Thisse, B., Raz, E., 2001. A zebrafish nanos-related gene is essential for the development of primordial germ cells. *Genes Dev* 15, 2877-2885.
- Lasko, P.F., Ashburner, M., 1988. The product of the *Drosophila* gene vasa is very similar to eukaryotic initiation factor-4A. *Nature* 335, 611-617.
- Li, L., Dong, J., Yan, L., Yong, J., Liu, X., Hu, Y., Fan, X., Wu, X., Guo, H., Wang, X., Zhu, X., Li, R., Yan, J., Wei, Y., Zhao, Y., Wang, W., Ren, Y., Yuan, P., Yan, Z., Hu, B., Guo, F., Wen, L., Tang, F., Qiao, J., 2017. Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell* 20, 858-873 e854.
- Lin, C.Y., Yu, J.K., Su, Y.H., 2021. Evidence for BMP-mediated specification of primordial germ cells in an indirect-developing hemichordate. *Evol Dev* 23, 28-45.
- Luo, Y.J., Su, Y.H., 2012. Opposing nodal and BMP signals regulate left-right asymmetry in the sea urchin larva. *PLoS Biol* 10, e1001402.
- MacDonald, B.T., Tamai, K., He, X., 2009. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell* 17, 9-26.

- Nakamura, T., Extavour, C.G., 2016. The transcriptional repressor Blimp-1 acts downstream of BMP signaling to generate primordial germ cells in the cricket *Gryllus bimaculatus*. *Development* 143, 255-263.
- Ng, C.L., Qian, Y., Schulz, C., 2019. Notch and Delta are required for survival of the germline stem cell lineage in testes of *Drosophila melanogaster*. *PLoS One* 14, e0222471.
- Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., Saitou, M., 2009. A signaling principle for the specification of the germ cell lineage in mice. *Cell* 137, 571-584.
- Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., Saitou, M., Surani, M.A., 2005. Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* 436, 207-213.
- Oulhen, N., Swartz, S.Z., Laird, J., Mascaro, A., Wessel, G.M., 2017. Transient translational quiescence in primordial germ cells. *Development* 144, 1201-1210.
- Oulhen, N., Swartz, S.Z., Wang, L., Wikramanayake, A., 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.
- Oulhen, N., Wessel, G., 2017. A quiet space during rush hour: Quiescence in primordial germ cells. *Stem Cell Res* 25, 296-299.
- Persani, L., Rossetti, R., Cacciatori, C., 2010. Genes involved in human premature ovarian failure. *J Mol Endocrinol* 45, 257-279.
- Ransick, A., Rast, J.P., Minokawa, T., Calestani, C., Davidson, E.H., 2002. New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. *Dev Biol* 246, 132-147.
- Raz, E., 2003. Primordial germ-cell development: the zebrafish perspective. *Nat Rev Genet* 4, 690-700.
- Robertson, S.E., Dockendorff, T.C., Leatherman, J.L., Faulkner, D.L., Jongens, T.A., 1999. germ cell-less is required only during the establishment of the germ cell lineage of *Drosophila* and has activities which are dependent and independent of its localization to the nuclear envelope. *Dev Biol* 215, 288-297.
- Rodriguez, A.J., Seipel, S.A., Hamill, D.R., Romancino, D.P., M, D.I.C., Suprenant, K.A., Bonder, E.M., 2005. Seawi--a sea urchin piwi/argonaute family member is a component of MT-RNP complexes. *RNA* 11, 646-656.
- Saitou, M., Payer, B., Lange, U.C., Erhardt, S., Barton, S.C., Surani, M.A., 2003. Specification of germ cell fate in mice. *Philos Trans R Soc Lond B Biol Sci* 358, 1363-1370.
- Saitou, M., Yamaji, M., 2010. Germ cell specification in mice: signaling, transcription regulation, and epigenetic consequences. *Reproduction* 139, 931-942.
- Santos, A.C., Lehmann, R., 2004. Germ cell specification and migration in *Drosophila* and beyond. *Curr Biol* 14, R578-589.
- Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., Ohta, H., Moritoki, Y., Iwatani, C., Tsuchiya, H., Nakamura, S., Sekiguchi, K., Sakuma, T., Yamamoto, T., Mori, T., Woltjen, K., Nakagawa, M., Yamamoto, T., Takahashi, K., Yamanaka, S., Saitou, M., 2015. Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell* 17, 178-194.

- Senft, A.D., Bikoff, E.K., Robertson, E.J., Costello, I., 2019. Genetic dissection of Nodal and Bmp signalling requirements during primordial germ cell development in mouse. *Nat Commun* 10, 1089.
- Shashikant, T., Khor, J.M., Ettensohn, C.A., 2018. From genome to anatomy: The architecture and evolution of the skeletogenic gene regulatory network of sea urchins and other echinoderms. *Genesis* 56, e23253.
- Shen, M.M., 2007. Nodal signaling: developmental roles and regulation. *Development* 134, 1023-1034.
- Subramaniam, K., 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.
- Tada, H., Mochii, M., Orii, H., Watanabe, K., 2012. Ectopic formation of primordial germ cells by transplantation of the germ plasm: direct evidence for germ cell determinant in *Xenopus*. *Dev Biol* 371, 86-93.
- Tam, P.P., Zhou, S.X., 1996. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev Biol* 178, 124-132.
- Tan, K., Song, H.W., Wilkinson, M.F., 2020. Single-cell RNAseq analysis of testicular germ and somatic cell development during the perinatal period. *Development* 147.
- Tanaka, S.S., Toyooka, Y., Akasu, R., Katoh-Fukui, Y., Nakahara, Y., Suzuki, R., Yokoyama, M., Noce, T., 2000. The mouse homolog of *Drosophila Vasa* is required for the development of male germ cells. *Genes Dev* 14, 841-853.
- Tsuda, M., Sasaoka, Y., Kiso, M., Abe, K., Haraguchi, S., Kobayashi, S., Saga, Y., 2003. Conserved role of nanos proteins in germ cell development. *Science* 301, 1239-1241.
- Tu, Q., Brown, C.T., Davidson, E.H., Oliveri, P., 2006. Sea urchin Forkhead gene family: phylogeny and embryonic expression. *Dev Biol* 300, 49-62.
- Voronina, E., Lopez, M., Juliano, C.E., Gustafson, E., Song, J.L., Extavour, C., George, S., Oliveri, P., McClay, D., 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.
- Wagner, D.E., Weinreb, C., Collins, Z.M., Briggs, J.A., Megason, S.G., Klein, A.M., 2018. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* 360, 981-987.
- Wang, J., Sinha, T., Wynshaw-Boris, A., 2012. Wnt signaling in mammalian development: lessons from mouse genetics. *Cold Spring Harb Perspect Biol* 4.
- Wang, J.J., Ge, W., Zhai, Q.Y., Liu, J.C., Sun, X.W., Liu, W.X., Li, L., Lei, C.Z., Dyce, P.W., De Felici, M., Shen, W., 2020. Single-cell transcriptome landscape of ovarian cells during primordial follicle assembly in mice. *PLoS Biol* 18, e3001025.
- Weiss, A., Attisano, L., 2013. The TGFbeta superfamily signaling pathway. *Wiley Interdiscip Rev Dev Biol* 2, 47-63.
- Wessel, G.M., Fresques, T., Kiyomoto, M., Yajima, M., Zazueta, V., 2014. Origin and development of the germ line in sea stars. *Genesis* 52, 367-377.
- Whittle, C.A., Extavour, C.G., 2016. Refuting the hypothesis that the acquisition of germ plasm accelerates animal evolution. *Nat Commun* 7, 12637.

- Whittle, C.A., Extavour, C.G., 2017. Causes and evolutionary consequences of primordial germ-cell specification mode in metazoans. *Proc Natl Acad Sci U S A* 114, 5784-5791.
- Yajima, M., Gustafson, E.A., Song, J.L., Wessel, G.M., 2014. Piwi regulates Vasa accumulation during embryogenesis in the sea urchin. *Dev Dyn* 243, 451-458.
- Yajima, M., Wessel, G.M., 2011. Small micromeres contribute to the germline in the sea urchin. *Development* 138, 237-243.
- Yajima, M., Wessel, G.M., 2012. Autonomy in specification of primordial germ cells and their passive translocation in the sea urchin. *Development* 139, 3786-3794.
- Yu, P.B., Hong, C.C., Sachidanandan, C., Babitt, J.L., Deng, D.Y., Hoyng, S.A., Lin, H.Y., Bloch, K.D., Peterson, R.T., 2008. Dorsomorphin inhibits BMP signals required for embryogenesis and iron metabolism. *Nat Chem Biol* 4, 33-41.
- Zhang, F., Liu, R., Liu, C., Zhang, H., Lu, Y., 2020. Nanos3, a cancer-germline gene, promotes cell proliferation, migration, chemoresistance, and invasion of human glioblastoma. *Cancer Cell International* 20, 197.
- Zhang, F., Phiel, C.J., Spece, L., Gurvich, N., Klein, P.S., 2003. Inhibitory phosphorylation of glycogen synthase kinase-3 (GSK-3) in response to lithium. Evidence for autoregulation of GSK-3. *J Biol Chem* 278, 33067-33077.

Chapter 2: Single cell RNA-seq in the sea urchin embryo shows marked cell-type specificity in the Delta/Notch pathway

Foster, S., Teo, Y.V., Neretti, N., Oulhen, N., Wessel, G.M., Single cell RNA-seq in the sea urchin embryo show marked cell-type specificity in the Delta/Notch pathway. *Mol Reprod Dev*, 2019. 86, 931-934.

This project was a collaboration between the Wessel and Neretti Labs. Yee Voan Teo completed the single-cell sequencing protocol. I contributed to the analysis and cell type identification and I wrote the manuscript.

Introduction

Sea urchin embryos are excellent for in vivo functional studies because of their transparency and tractability in manipulation. They are also favorites for pharmacological approaches since they develop in an aquatic environment and addition of test substances is straightforward. Precise interpretations from such experiments though are often not feasible because of potential pleiotropic effects of the reagents, and the uncertainty of which cells are impacted by the perturbant.

The Wnt and Delta/Notch pathways are intensively studied due to their influence and prevalence in animal development and homeostasis [1-4]. These pathways are particularly well studied in sea urchins. In the canonical Wnt pathway, the Wnt ligand binds to its Frizzled receptors as well as its LRP5/6 co-receptors to stabilize the β -catenin protein and induce its nuclear translocation to regulate the transcription of target genes. The Wnt pathway in the sea urchin is involved in the specification of the endomesoderm [5], the posterior ectoderm [6] and the neuroectoderm [7]. In contrast, the Delta/Notch pathway is required for mesoderm fate, in part by transcription of the transcription factor FoxY to specify the future coelomic pouch cell progenitors [8]. The Delta ligand is enriched in the sea urchin micromeres and activates the Notch receptor in adjacent cells. This interaction induces proteolytic cleavage of the cytoplasmic domain of Notch, which then translocates into the nucleus to regulate transcription [9].

Here we use single cell mRNA sequencing (scRNA-seq) for the first time as a metric of cell types in the sea urchin embryo and to determine the selectivity of two commonly used inhibitors, one each for the Wnt and the Delta/Notch pathways, on these nascent

cell types. Wnt signaling was inhibited using the small, cell permeable molecule c59. C59 inhibits Wnt signaling *in vivo* by inhibiting Porcupine (PORCN), and thereby blocking progression of Wnt in the secretory pathway [10]. DAPT (**N-[N-(3,5-Difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester**) is a γ -secretase inhibitor and blocks cleavage of the Notch intracellular domain following interaction with Delta [11].

Cells for scRNA-seq were collected from *Strongylocentrotus purpuratus* embryos at early and late gastrula stages and processed via DropSeq [12]. On a microfluidic chip, a single cell suspension from dissociated embryos, barcoded hydrogel beads, and oil meet to form individual droplets containing one cell and one barcoded bead. Beads are coated with primers that contain a barcode unique to each bead, to mark individual cells, and each primer also has a unique molecular identifier (UMI) to mark individual transcripts for quantitation purposes. When a cell is lysed within the droplet, polyadenylated mRNAs are captured and barcoded by the primers. A total of 15,798 single cells were profiled from eight samples on the 10x Chromium Controller. Sequencing of these mRNAs provided 67.2K to 142.2K reads per cell and on average 80% of the transcripts mapped confidently to the *S. purpuratus* genome. The scRNA-seq datasets from control and drug treated embryos were aligned by canonical correlation analysis (CCA) using the R package, Seurat. CCA analysis links individual cells across datasets based on similar gene expression profiles. Visualizing the scRNA-seq data by T-Distributed stochastic neighbor embedding (t-SNE) plots, we identified various cell types by distinct mRNA profiling in gastrulae, a period of diversification of cell fates. In the t-SNE plot, the distances between clusters correlate with the

differences in the mRNA profiles. Both conserved and differentially expressed markers were extracted for each cluster across a control-drug treatment pair. Quality control for the scRNA-seq datasets included setting lower and upper thresholds for gene number per cell. A lower threshold of 400 genes was used to omit empty droplets, and an upper threshold of 2500 genes was set to exclude doublets, cases where two cells were encapsulated in one droplet, from further analysis.

Several predictions were borne out by this analysis. First, we identified 11 distinct cell types based on mRNA profiling. This number is compatible with results from gene regulatory analysis (GRN) and in situ hybridizations (echinobase.org) but uses a distinct metric for identification. In addition to morphology, phenotype, and functionality, we anticipate RNA-profiling to enhance the definition of cell-types. Second, the cell lineages affected by Wnt and Delta/Notch inhibition were distinct from each other. For example, several marker transcripts expressed in the future pigment cell lineage (cluster 5) were specifically downregulated selectively after DAPT treatment ([13]), concurrently we observed an increase in proportion of oral NSM cells (cluster 9). We speculate that the mRNA profiles for the pigment cell lineage have shifted to more align with the lineage of oral NSM cells. Each of these cell types comes from a recent common lineage (Veg2 tier of cells) and suggests plasticity in this lineage under control of the Delta/Notch pathway. In contrast, Foxq2, among other transcripts, was upregulated in many of the cell types after C59 treatment, consistent with prior work indicating its repression by Wnt signaling ([14] [15]). In clusters where a change in Foxq2 mRNA level was not detected, a vast majority of the cells in the cluster do not express Foxq2. Finally, cluster analysis showed nearly complete separation of pigment cells in DAPT versus vehicle control

(cluster 5), and skeletogenic cells in Wnt inhibitor versus vehicle control (cluster 8).

These data support the specificity and distinct effects of the Wnt and Delta/Notch pathways in differential cell lineages and illuminate how these conserved pathways selectively regulate cell fates at a single cell level.

Overall, we conclude that single cell RNA-seq analysis in this embryo is revealing of the cell types present during development, of the changes in the gene regulatory network resulting from inhibition of various signaling pathways, and of the selectivity of these pathways in influencing developmental trajectories. Sea urchin embryos appear well suited to this technology.

Materials and Methods

Embryo Processing

Adult *Strongylocentrotus purpuratus* were obtained from South Coast Bio-Marine info@scbiomarine.com and were spawned using 0.5M KCl injections intracoelomically.

DAPT (Tocris) was added at fertilization to 8 μ M, a concentration previously determined to effectively perturb Delta/Notch signaling without affecting overall development [16].

C59 (Cellagen) was added at 0.5 μ M at fertilization as previously described [15].

Embryos were cultured and, at appropriate developmental stages, were dissociated into single cells as described [17]. We started with 50mL cultures containing 0.25% of embryos (125 μ L of embryos in 50mL). The full volume was dissociated into single cells.

We obtained 12 million single cells from which 5,000 were used for the DropSeq.

Briefly, embryos were washed 2x in calcium-free seawater (CaFSW; 4°C) by gentle centrifugation (500xg), and then resuspended in hyaline extraction media (HEM; 4°C) containing 1mM EGTA. Note: do not use EDTA since some of the early steps in the 10x protocol use enzymes that rely on magnesium for function. The embryos in HEM were inverted every few minutes while remaining at 4°C and monitored for indication of loss of adhesion (the cells transition from a cuboidal or squamous epithelium to a rounded morphology). At that point, the embryos are washed 1x with CaFSW (4°C) and replaced with 0.5M NaCl and dissociated by several passes through a loose fitting Dounce homogenizer. Gentle shear force is sufficient to achieve near complete dissociation with no detectable lysis. The cell suspension is then passed through a 40 micron Nitex filter to remove undissociated embryos, counted on a haemocytometer (see Supplemental

Figure 5), and then diluted into an appropriate concentration of 0.5M NaCl (4°C) for the 10x protocol.

Methods for Single cell RNA sequencing

The single-cell RNA-seq protocol was performed using the ChromiumTM Single Cell 3' reagent kit v2 chemistry and cells were loaded on a GemCode Single Cell Instrument (10x Genomics, Pleasanton, CA). Approximately 5000 single cells were targeted from each sample. Libraries were sequenced on Illumina HiSeq 2500 with the custom configuration of read 1 (26bp) and read 2 (98bp), i7 index (8bp) and i5 index (0bp).

The Cell Ranger Single Cell Software Suite 2.1.0 was used to perform single cells demultiplexing and UMI counting (<https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome>). The custom transcriptome reference was generated from assembly Spur_4.2 using CellRanger. Cells that have at least 400 or at most 2500 genes detected were included in the downstream analysis using Seurat 2.3.0 [18]. The union of the top 1000 highest dispersion genes from the compared datasets were selected for the alignment. T-SNE projection and clustering analysis (dims=1:20 and resolution=0.5) were performed using Seurat. Markers were found using FindConservedMarkers and FindMarkers functions.

Acknowledgements

The authors declare no conflicts of interest in this work, and are grateful for support from the National Institutes of Health 9RO1GM125071.

References

1. Duchartre, Y., Y.M. Kim, and M. Kahn, *The Wnt signaling pathway in cancer*. Crit Rev Oncol Hematol, 2016. **99**: p. 141-9.
2. Nusse, R. and H. Clevers, *Wnt/beta-Catenin Signaling, Disease, and Emerging Therapeutic Modalities*. Cell, 2017. **169**(6): p. 985-999.
3. Zhang, R., A. Engler, and V. Taylor, *Notch: an interactive player in neurogenesis and disease*. Cell Tissue Res, 2018. **371**(1): p. 73-89.
4. Li, L., et al., *Notch signaling pathway networks in cancer metastasis: a new target for cancer therapy*. Med Oncol, 2017. **34**(10): p. 180.
5. Croce, J., et al., *Wnt6 activates endoderm in the sea urchin gene regulatory network*. Development, 2011. **138**(15): p. 3297-306.
6. McIntyre, D.C., et al., *Short-range Wnt5 signaling initiates specification of sea urchin posterior ectoderm*. Development, 2013. **140**(24): p. 4881-9.
7. Range, R.C., R.C. Angerer, and L.M. Angerer, *Integration of canonical and noncanonical Wnt signaling pathways patterns the neuroectoderm along the anterior-posterior axis of sea urchin embryos*. PLoS Biol, 2013. **11**(1): p. e1001467.
8. Materna, S.C., S.Z. Swartz, and J. Smith, *Notch and Nodal control forkhead factor expression in the specification of multipotent progenitors in sea urchin*. Development, 2013. **140**(8): p. 1796-806.
9. Kopan, R. and M.X. Ilagan, *The canonical Notch signaling pathway: unfolding the activation mechanism*. Cell, 2009. **137**(2): p. 216-33.
10. Proffitt, K.D., et al., *Pharmacological inhibition of the Wnt acyltransferase PORCN prevents growth of WNT-driven mammary cancer*. Cancer Res, 2013. **73**(2): p. 502-7.
11. Geling, A., et al., *A gamma-secretase inhibitor blocks Notch signaling in vivo and causes a severe neurogenic phenotype in zebrafish*. EMBO Rep, 2002. **3**(7): p. 688-94.
12. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets*. Cell, 2015. **161**(5): p. 1202-1214.
13. Ohguro, Y., H. Takata, and T. Kominami, *Involvement of Delta and Nodal signals in the specification process of five types of secondary mesenchyme cells in embryo of the sea urchin, Hemicentrotus pulcherrimus*. Dev Growth Differ, 2011. **53**(1): p. 110-23.
14. Yaguchi, S., et al., *A Wnt-FoxQ2-nodal pathway links primary and secondary axis specification in sea urchin embryos*. Dev Cell, 2008. **14**(1): p. 97-107.
15. Cui, M., et al., *Specific functions of the Wnt signaling system in gene regulatory networks throughout the early sea urchin embryo*. Proc Natl Acad Sci U S A, 2014. **111**(47): p. E5029-38.
16. Materna, S.C. and E.H. Davidson, *A comprehensive analysis of Delta signaling in pre-gastrular sea urchin embryos*. Dev Biol, 2012. **364**(1): p. 77-87.
17. McClay, D.R., *Methods for embryo dissociation and analysis of cell adhesion*. Methods Cell Biol, 2004. **74**: p. 311-29.
18. Butler, A., et al., *Integrating single-cell transcriptomic data across different conditions, technologies, and species*. Nat Biotechnol, 2018. **36**(5): p. 411-420.

Figures

(A & B)

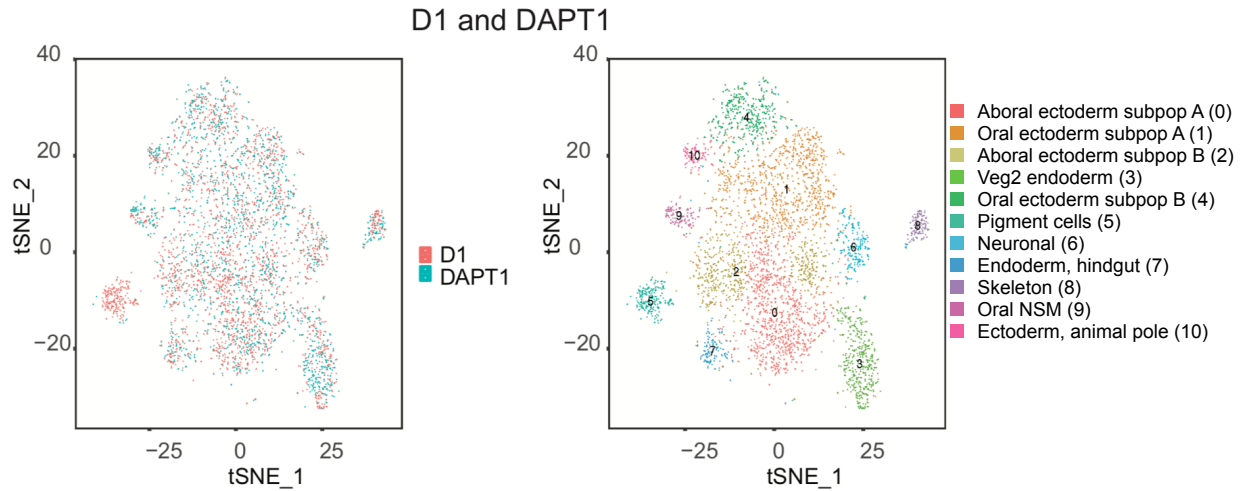
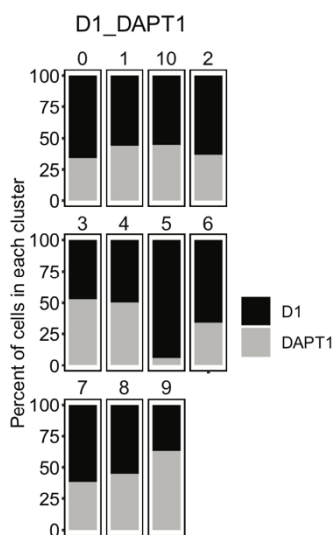


Figure1: Single cell analysis of late gastrula treated with DMSO only (D1) or with the Delta/Notch inhibitor (DAPT1). Results are presented as t-SNE plots, in which each dot represents a transcriptome from a single cell. (A) Comparison of the transcriptomes obtained between the control (red) and the DAPT treated embryos (blue). (B) Representation of the total number of clusters obtained for D1 (DMSO control) and DAPT1 grouped together. Note the distinct change in composition of cells in cluster 5 (blue) following DAPT treatment.

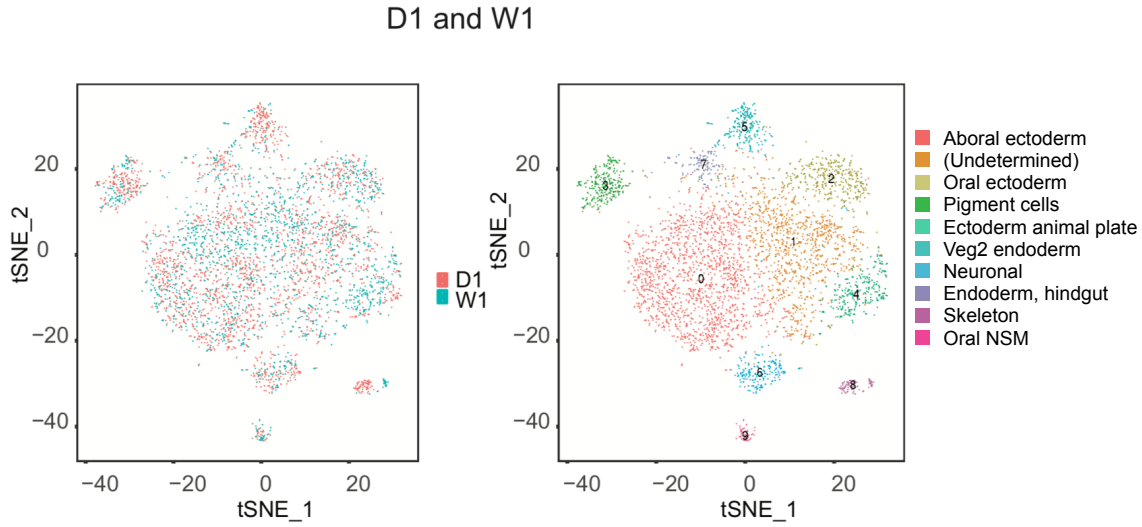
Cluster Number	Cell Type	Representative Marker Genes
0	Aboral ectoderm subpopulation A	spec1, spec2d
1	Oral ectoderm subpopulation A	foxG, hnf6
2	Aboral ectoderm subpopulation B	ARS
3	Veg2 endoderm	Blimp, Endo16, FoxA, GataE
4	Oral ectoderm subpopulation B	Lefty, Frizzled, Foxq2
5	Pigment cells	gcm
6	Neuronal	neural cadherin, AHNAK
7	Endoderm, hindgut	CBF-beta
8	Skeleton	Alx1, Sm50, Sm30a
9	Oral NSM	delta, erg
10	Ectoderm, animal pole	Foxq2

Supplement Table 1: Clusters seen in Figure 1 are identified with specific markers.



SFigure 1: Bar plots depicting percent of cells corresponding to each sample present in each cluster. Note distinct loss of cluster 5 cells in DAPT treated embryos seen also in t-SNE plot.

(A & B)



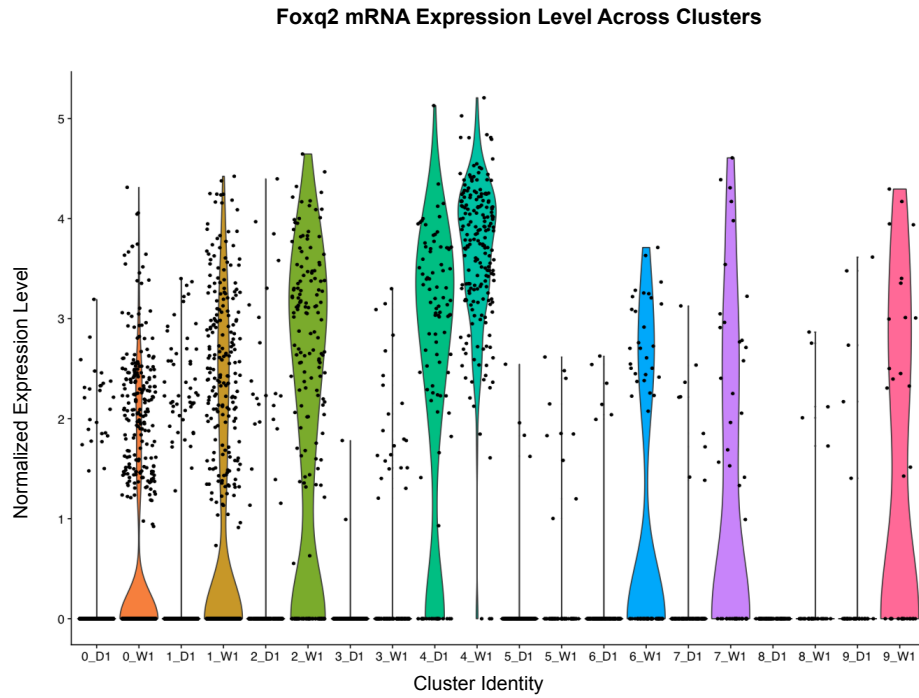
(C)

Cluster Number	Cell Type	Marker Gene
0	Aboral ectoderm	spec2d, ARS, hox1
1	Undetermined	otx
2	Oral ectoderm	Lefty
3	Pigment cells	gcm
4	Ectoderm animal plate	foxq2
5	Veg2 endoderm	Blimp, Endo16, FoxA
6	Neuronal	Neural cadherin
7	Endoderm, hindgut	CBF-beta
8	Skeleton	Alx1, Sm50, Sm30a
9	Oral NSM	delta, erg

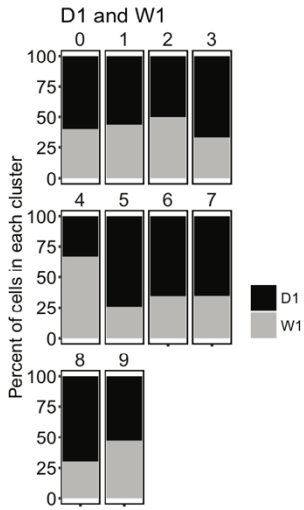
SFigure 2: Single cell analysis of late gastrulae treated with DMSO only (D1) or with the Wnt inhibitor, c59 (W1). Results are presented as t-SNE plots, each dot represents a transcriptome from a single cell. (A) Comparison of the transcriptomes obtained between the control (red) and the c59 treated embryos (blue). (B) Representation of the total number of clusters obtained for D1 and W1 grouped together. (C) Clusters seen in SFigure 2 are identified with specific markers. Note the difference in control clusters depicted in the t-SNE plots and marker genes. This variation results from aligning the same control dataset (D1) to two independent treatment datasets.

	D1-W1	
Cluster Number	LogFC	Adjusted p-value
0	- 1.188229076	2.31E-59
1	- 1.478037802	6.97E-39
2	- 1.745843531	6.97E-25
3	- 0.992697757	4.33E-06
4	- 0.706903072	2.97E-10
5	X	X
6	- 1.674281575	3.55E-07
7	- 1.958703206	0.001256764
8	X	X
9	X	X

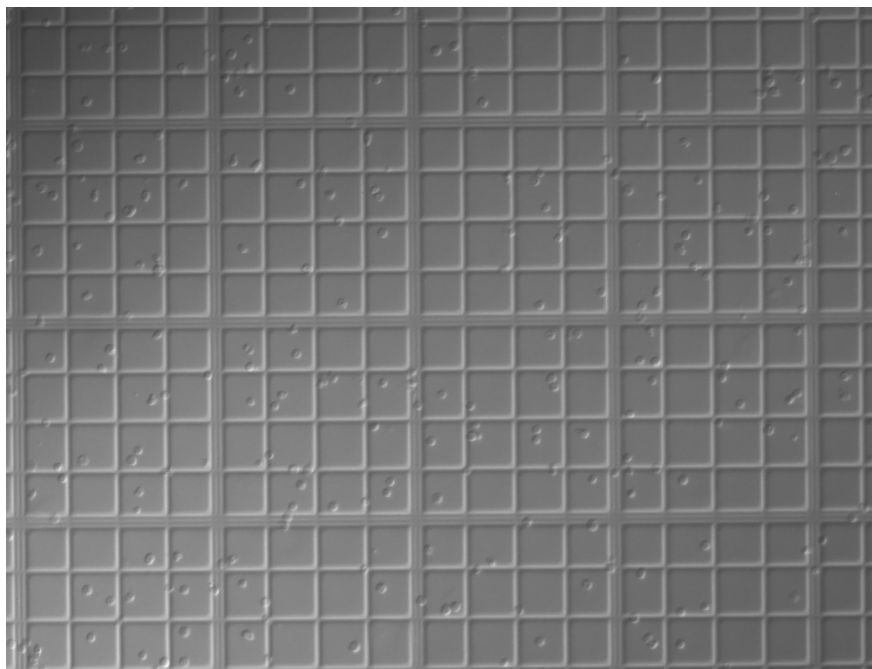
Supplement Table 2: Log fold change in Foxq2 mRNA expression level in control (D1) versus Wnt inhibitor-treated embryos (W1). Negative values indicate upregulation in treated embryos compared to control. Foxq2 mRNA level was significantly upregulated in a majority of the clusters. Blank= not detected as significantly differentially expressed.



SFigure 3: Normalized Foxq2 expression levels across all clusters in control (D1) versus Wnt inhibitor-treated embryos (W1). Each dot on the violin plot represents one cell and distribution of expression data is highlighted in color. Foxq2 expression was upregulated in most cell types after Wnt inhibition, consistent with extracted differential expression data summarized in previous figure. In clusters 5, 8, and 9, most cells do not exhibit Foxq2 expression. The x-axis nomenclature shows cluster number followed by sample identification.



SFigure 4: Bar plots depicting percent of cells corresponding to each sample present in each cluster. Note the distinct loss of cells from cluster 8 and 5 in Wnt inhibitor-treated embryos and increase in cluster 4 cells.



SFigure 5: Representative image of dissociated cells on a hemacytometer prior to loading on the 10x instrument. Note the individuality of the cells and the lack of cell clumps. Each cell is approximately 8 microns in diameter.

Chapter 3: A single cell RNA-seq resource for early sea urchin development

Foster S, Oulhen N, Wessel G: A single cell RNA sequencing resource for early sea urchin development. *Development* 2020, 147(17).

The embryo cultures used for this study were prepared by Nathalie Oulhen. The single cell RNA sequencing protocol was performed by myself and Nathalie Oulhen. I analyzed the datasets, performed cell type identification, and created the figures. Nathalie created the embryo graphic in Figure 1B. I wrote majority of the manuscript with support from Gary Wessel and Nathalie Oulhen.

Highlights

1. We discovered 22 major cell states in late gastrulae of *Strongylocentrotus purpuratus* by single cell mRNA-seq.
2. Sub-clustering of the major embryonic domains reveals over 50 cell states with distinct transcript profiles.
3. The ectodermal, endodermal, and mesodermal germ layers each have multiple distinct cell states.
4. We found two cell states in the embryo expressing germline markers.
5. Many mRNAs thought to be exclusive to various cells types, are instead more broadly expressed.

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 by single cell RNA-seq. 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 with focus on key transition stages for cell type specification of each germ layer. Sub-clustering of these major embryonic domains revealed over 50 cell states with distinct transcript profiles. Further, we identified the transcript profile of two cell states expressing germ cell factors, one we conclude are the primordial germ cells, and the other transiently present during gastrulation. We hypothesize that these cells of the veg2 tier of the early embryo may 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.

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 to reveal conserved principles in development, and nodes in which evolutionary change has occurred. While the functional end product of gene expression is the driving force of change, whether it is a protein, metabolites, or other biochemistry, it is usually very difficult to see such specific activities with current technologies. Instead, investigators focus on DNA and RNA polymers since they can be analyzed specifically through their sequence, and be amplified for remarkable sensitivity.

Single cell RNA-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 rapidly constructs 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/less similarity 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 though 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 might identify distinctions in genes expressed in each cell state that enables 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 scmRNA-seq (Foster et al., 2019). Two sibling populations of dissociated cells were analyzed at one time point (early gastrula), the only difference being presence or absence of an inhibitor (DAPT) of the Notch/Delta 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 Notch/Delta 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

[\[http://www.echinobase.org/endomes/\]](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 2 sequential, asymmetric divisions. These presumed PGCs promptly differ significantly from their sibling somatic cells. For example, while the somatic cells continue to divide rapidly, the 4 PGCs become mitotically quiescent shortly after they are formed and they will divide only once until after gastrulation, resulting in an embryo of 8 PGCs versus over 1000 somatic cells. In addition to the down regulation of their cell cycle, the PGCs also reduce their 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 of cell states from the 8-cell stage to late gastrulae using single cell mRNA-seq technology in the embryo of *S. 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 single cell RNA sequencing analysis via DropSeq using the 10x Genomics platform. Our analysis included eight time points spanning early development, from three hours post-fertilization to late gastrula stage (Table S1), selected based on 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. Here we identify cell states of the sea urchin arising from the three germ layers: ectoderm, endoderm, and mesoderm. Upon integration of the datasets and visualization with tSNE dimensionality reduction, we identify 22 major cell states, including the germline (Fig. 1A), based on known marker gene expression (Table S2). These 22 major clusters can be further segregated to reveal regions within the digestive system, distinct ectoderm cell states, and diverse blastocoelar populations (Fig. 1B and S8). The identification of 22 clusters is a direct result of the resolution parameter used in the clustering step of the analysis. The resolution used of 0.5 is a conservative setting whereas increased resolution resulted in 58 separable cell states (Fig. 1B). Here we characterize each of the major 22 clusters, which are well annotated cell states of the sea urchin embryo, and since no resolution is optimal for each cell type, we show results from sub-clustering of each of the major cell populations.

We first tested consistency in the single cell sequencing approach for this animal. While 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 variation in numbers of cells, in depth of sequencing, as well as differences in developmental timing between these wildtype embryos.

In analyzing the eight developmental datasets, differences between early and late time points were apparent. We detected a range of 2,000 to 7,000 genes per cell of embryos collected between 8- to 64-cell stage. In contrast, we detected on average fewer than 2,000 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 progresses (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) that cells have reduced potency during development and

acquire 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 this embryo sometimes until 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 is also consistent with other reports using a variety of cell types (Macosko et al., 2015; Siebert et al., 2019). At each stage of sea urchin embryogenesis (*S. purpuratus*), 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 would necessarily express less than this overall maximum so that the yield of mRNA information from each cell state could range as high as 60%. Thus, this scmRNA-seq approach does not yield a complete mRNA profile, but one sufficient for distinguishing diverse cells states not otherwise achievable.

Identification of 22 major cell populations through development to late gastrulae

Cluster visualization by tSNE plots show 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 then directed to further analysis in the supplement 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 is 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 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 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 is 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 blastula stage (Yaguchi et al., 2010). We defined clusters 9 and 17 as neural cell states based on co-expression 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 gastrula stage (Oliveri et al., 2006). Endo16, an endoderm enriched gene, is first detected via in situ hybridization in the vegetal pole of mesenchyme blastula embryos. At 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 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 (Fig. S6, Fig. S7). In gastrula stages, FoxA expression was more dynamic, present in cell states that do not express Endo16 (Fig. 3). FoxA expression is known to extend to the foregut while 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 7 cell populations. After subclustering, we can distinguish the hind and midgut, the foregut, the secondary mesenchyme cells, the pigment cells, the skeleton cells, and neuronal cells (Fig. S8).

Mesoderm representatives: The primary mesenchyme cells (PMC) are a mesodermally-derived cell type which gives rise to the larval skeleton. The PMC lineage starts with an 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 will 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 PMC 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 which is 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 scmRNA-seq at the 64-cell stage in cluster 21. At morula stage, *Alx1* expression is enriched in cluster 16 at 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 these clusters represent the PMC cells (16,19) (Fig. 4, Fig. S9, Fig. S10, Fig.S11). At late gastrula, the percentage of cells expressing the PMC markers was greatly reduced. For example, at mesenchyme blastula, clusters 16 and 19 represent respectively 2.6 and 1.9% of total cells whereas at 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 percent of cells seen by scmRNA-seq is 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 the asymmetric cell division of the micromeres at the 16-32 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 rarity of these 2 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 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 detected is expected since, 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 is first detected in cluster 20 at 64-cell stage and its expression is maintained in this cluster up to 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 is transiently detected in an additional cluster (16). These Nanos2 positive cells in both cluster 16 and 20 co-express two additional germline markers, vasa and seawi (Fig. 6B). Nanos2 expression is known to be very restricted in the embryo while vasa and seawi are more widespread, and that pattern is

reflected in the plot (Fig. 6A). Later in development, cells of cluster 16 start expressing genes such as *Alx1* and *SM50* 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). This transient *Nanos2* expression in cluster 16 (PMCs) suggests that the micromeres themselves might be expressing *Nanos2* before giving rise to the germline. *Nanos2* is 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. While both clusters are quite similar in terms of gene expression, we see high levels of *SoxC* expression in cluster 16 compared to 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 appear (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., 2019; 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 are 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 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) (Table S3, S4, 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 is already highly expressed in the earlier stages of development and becomes specifically retained in the PGCs (20), suggesting the maternal origin of this transcript. Vasa and Seawi are also more enriched in the germline compared to the Veg2 lineage (Fig. S14, Fig. 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 are differentially expressed between these Nanos2 positive cell states, making them distinct. These analyses revealed genes that are always co-expressed 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 (Kruppel like), Odz, Staufen and Maelstrom were also found in this analysis (Fig. S17-20). Krüppel-like factors are a family of zinc-finger transcription factors, essential for maintaining pluripotency (Bialkowska et al., 2017). Odz is a pair rule gene required for somatic gonad formation in *C.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; St Johnston et al., 1991). In mice, staufen is expressed in the germ cells in both males and females (Saunders et al., 2000). Maelstrom 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, 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), CNOT6 (RNA degradation)

(Oulhen et al., 2017; Swartz et al., 2014). Quiescence is a common phenotype of stem cells and this single cell RNA seq data set is essential to better understand 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 are highly enriched in the early stage PGCs compared to 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 to 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 find 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 have 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 instead have 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 based on 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 to test the result. We have explored whether this result could be artifactual. The broad distribution is most apparent in the violin plots in which each cell with a gene marker may be visualized, versus the feature plots, a more common figure format in the literature of single cell mRNA-seq in which the eye is directed to clusters of the greatest abundance [satijalab.org/seurat/v3.0/]. Further, we find clusters lacking the breadth of expression, suggesting that the mRNA was not a broad contaminant of the experiment. In previous results (Figure 3 within (Zazueta-Novoa and Wessel, 2014) (Fig. S21)), the selectivity of a 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. Since the scmRNA-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 since 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). Whereas by single cell RNA-seq here, we find Gus transcripts enriched in cells of the vegetal pole, but broadly detectable in all cells of the embryo (Fig. S21). If, however, the in situ hybridization signal reaction was extended, Gus mRNA was then seen present broadly throughout the embryo ((Zazueta-Novoa and Wessel, 2014) and Fig. S21) 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 though of this phenomenon from widely disparate cells and organisms ((Briggs et al., 2018; Siebert et al., 2019); 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 shown herein. We note that some transcripts are 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 though 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. Since 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 if 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 the expression of these common genes is not specific to the germline though, 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 RNAseq of the germ line cells (calcein-positive cells) against all the other cells of the embryo (calcein-negative cells) while 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 single cell mRNA-seq would show this same transcript broadly expressed in clusters throughout the vegetal hemisphere of the embryo.

Nanos is a 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 (e.g. *C. elegans* (Kraemer et al., 1999), *Xenopus* (Lai et al., 2011) planarians (Wang et al., 2007)). We recently discovered that Nanos2 in the sea urchin is not only expressed in the germline but is present also 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 cell, results in dramatic upregulation of Vasa protein throughout the embryo. Later, Vasa becomes restricted to sub-populations 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 will give rise to multiple cell types such as the pigment cells, the blastocoelar cells, the 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 the PGCs requires the maternal Wnt pathway for expression whereas the somatic Nanos2 expression instead requires Delta/Notch signaling through the forkhead family member FoxY (Oulhen et al., 2019). 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 now 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 supports this conclusion. This includes partial quiescence of protein synthesis and other metabolic features (Oulhen et al., 2017; Oulhen et al., 2019) that reflects a lower level of Nanos2 translational repression. Future experimentation will attempt to analyze scmRNA data in this embryo following micromere depletion to

determine how the Veg2 lineage may change in such instances in loss of PGC-precursors. Since some other animal embryos also appear to somehow compensate 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 scmRNA-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 0.5M KCl into the adult coelomic cavity. Fertilization was accomplished in sea water containing 10mM p-aminobenzoic acid to reduce cross-linking of the fertilization envelope, and which was washed out after 30 minutes. Embryos were cultured in filtered (0.2micron) 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 hyalin-extraction media (HEM) for 10-15 minutes, 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 40micron Nitex mesh, counted on a hemocytometer, and diluted to reach the appropriate concentration for the scmRNA-seq protocol. Equal numbers of embryos were used in each time point and at no time were cells or embryos pelleted in a centrifuge (Oulhen et al., 2019).

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 (2x150 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 10X 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. 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 the parameter `dimensions=15` and `resolution=0.5`. The use of 15 dimensions, in other words, the 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 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 with 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 dimensions:10, resolution: 1.0. Cluster markers were found using FindConservedMarkers and FindMarkers functions. The comparison of the Foster et al., 2019 dataset and the corresponding timepoint from this dataset was performed following the same pipeline described above and visualized with parameters dims=20, res=0.5. AverageExpression measured as count data normalized to library size and log transformed (Stuart et al., 2019).

Data availability: The sequencing files and gene expression matrices for the single-cell RNA-seq analysis presented here have been deposited in the GEO database and may be accessed with the accession number GSE149221.

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. We are grateful for support from the NIH (9RO1GM125071, 1R01GM132222, 1P20GM119943).

References

- Andrikou, C., Iovene, E., Rizzo, F., Oliveri, P., Arnone, M.I., 2013. Myogenesis in the sea urchin embryo: the molecular fingerprint of the myoblast precursors. *Evodevo* 4, 33.
- Anishchenko, E., Arnone, M.I., D'Aniello, S., 2018. SoxB2 in sea urchin development: implications in neurogenesis, ciliogenesis and skeletal patterning. *Evodevo* 9, 5.
- Annunziata, R., Perillo, M., Andrikou, C., Cole, A.G., Martinez, P., Arnone, M.I., 2014. Pattern and process during sea urchin gut morphogenesis: the regulatory landscape. *Genesis* 52, 251-268.
- Bialkowska, A.B., Yang, V.W., Mallipattu, S.K., 2017. Kruppel-like factors in mammalian stem cells and development. *Development* 144, 737-754.
- Briggs, J.A., Weinreb, C., Wagner, D.E., Megason, S., Peshkin, L., Kirschner, M.W., Klein, A.M., 2018. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* 360.
- Bruskin, A.M., Tyner, A.L., Wells, D.E., Showman, R.M., Klein, W.H., 1981. Accumulation in embryogenesis of five mRNAs enriched in the ectoderm of the sea urchin pluteus. *Dev Biol* 87, 308-318.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., Satija, R., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 36, 411-420.
- Cameron, R.A., Fraser, S.E., Britten, R.J., 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., Chen, D., 2009. E2f-mediated degradation of Cyclin A is essential for the maintenance of germline stem cells in *Drosophila*. *Development* 136, 4133-4142.
- Cho, P.F., Gamberi, C., Cho-Park, Y.A., Cho-Park, I.B., Lasko, P., Sonenberg, N., 2006. Cap-dependent translational inhibition establishes two opposing morphogen gradients in *Drosophila* embryos. *Curr Biol* 16, 2035-2041.
- Clegg, N.J., Findley, S.D., Mahowald, A.P., Ruohola-Baker, H., 2001. Maelstrom is required to position the MTOC in stage 2-6 *Drosophila* oocytes. *Dev Genes Evol* 211, 44-48.
- Davidson, E.H., Cameron, R.A., 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., 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.
- Ettensohn, C.A., Illies, M.R., Oliveri, P., 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.
- Findley, S.D., Tamanaha, M., Clegg, N.J., 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.

- Fink, R.D., McClay, D.R., 1985. Three cell recognition changes accompany the ingression of sea urchin primary mesenchyme cells. *Dev Biol* 107, 66-74.
- Foster, S., Teo, Y.V., Neretti, N., Oulhen, N., 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.
- Fujii, T., Sakamoto, N., Ochiai, H., Fujita, K., Okamitsu, Y., Sumiyoshi, N., Minokawa, T., Yamamoto, T., 2009. Role of the nanos homolog during sea urchin development. *Dev Dyn* 238, 2511-2521.
- Garner, S., Zysk, I., Byrne, G., Kramer, M., Moller, D., Taylor, V., Burke, R.D., 2016. Neurogenesis in sea urchin embryos and the diversity of deuterostome neurogenic mechanisms. *Development* 143, 286-297.
- George, N.C., Killian, C.E., 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.
- Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan, F., Kuo, A.H., Hsieh, R.W., Cai, S., Zabala, M., Scheeren, F.A., Lobo, N.A., Qian, D., Yu, F.B., Dirbas, F.M., Clarke, M.F., Newman, A.M., 2020. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* 367, 405-411.
- Gustafson, E.A., Wessel, G.M., 2010. Exogenous RNA is selectively retained in the small micromeres during sea urchin embryogenesis. *Mol Reprod Dev* 77, 836.
- Gustafson, E.A., Yajima, M., Juliano, C.E., 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.
- Irish, V., Lehmann, R., Akam, M., 1989. The Drosophila posterior-group gene nanos functions by repressing hunchback activity. *Nature* 338, 646-648.
- Juliano, C.E., Yajima, M., 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.
- Klein, A.M., 2017. What is your conceptual definition of "Cell type" in the context of a mature organism? *Cell Syst*, 256.
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., Kirschner, M.W., 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187-1201.
- Kraemer, B., Crittenden, S., Gallegos, M., Moulder, G., Barstead, R., Kimble, J., Wickens, M., 1999. NANOS-3 and FBF proteins physically interact to control the sperm-oocyte switch in *Caenorhabditis elegans*. *Curr Biol* 9, 1009-1018.
- Lai, F., King, M.L., 2013. Repressive translational control in germ cells. *Mol Reprod Dev* 80, 665-676.
- Lai, F., Zhou, Y., Luo, X., Fox, J., King, M.L., 2011. Nanos1 functions as a translational repressor in the *Xenopus* germline. *Mech Dev* 128, 153-163.
- Luo, X., Nerlick, S., An, W., King, M.L., 2011. *Xenopus* germline nanos1 is translationally repressed by a novel structure-based mechanism. *Development* 138, 589-598.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., Trombetta, J.J., Weitz, D.A., Sanes, J.R., Shalek, A.K., Regev, A., McCarroll, S.A., 2015. Highly Parallel

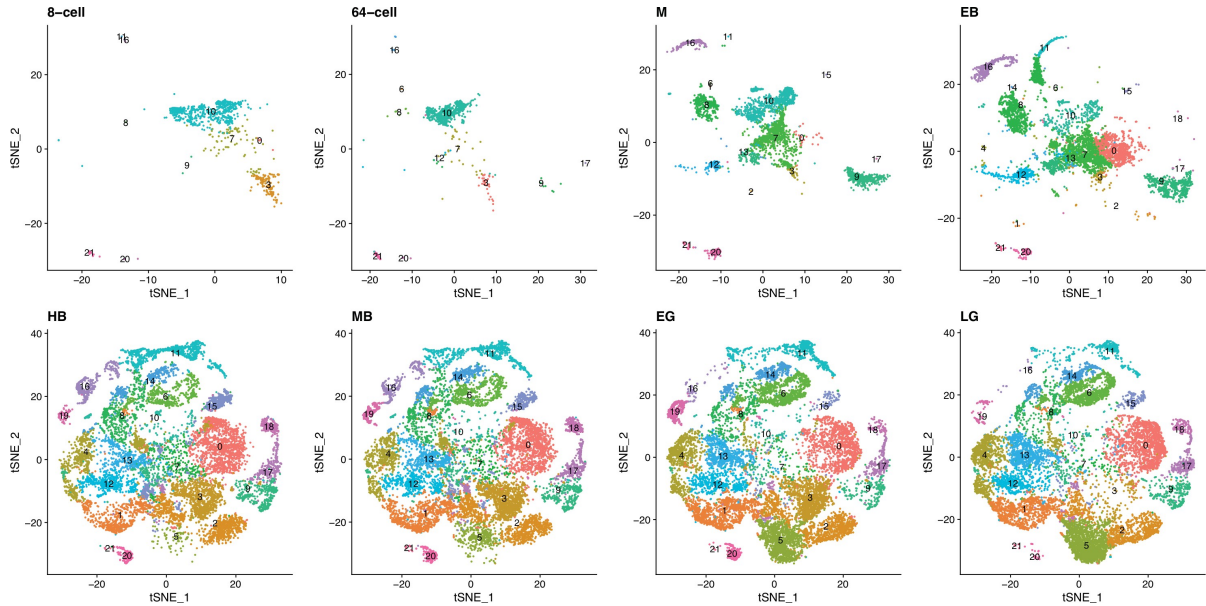
- Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202-1214.
- Materna, S.C., Swartz, S.Z., Smith, J., 2013. Notch and Nodal control forkhead factor expression in the specification of multipotent progenitors in sea urchin. *Development* 140, 1796-1806.
- McClay, D.R., Chambers, A.F., 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.
- McClay, D.R., 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.
- Oliveri, P., Walton, K.D., Davidson, E.H., McClay, D.R., 2006. Repression of mesodermal fate by *foxa*, a key endoderm regulator of the sea urchin embryo. *Development* 133, 4173-4181.
- Oulhen, N., Foster, S., Wray, G., Wessel, G., 2019a. Identifying gene expression from single cells to single genes. *Methods Cell Biol* 151, 127-158.
- Oulhen, N., Swartz, S.Z., Laird, J., Mascaro, A., Wessel, G.M., 2017. Transient translational quiescence in primordial germ cells. *Development* 144, 1201-1210.
- Oulhen, N., Swartz, S.Z., Wang, L., Wikramanayake, A., 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.
- Oulhen, N., 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.
- Patmanidi, A.L., Champeris Tsaniras, S., Karamitros, D., Kyrousi, C., Lygerou, Z., 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.
- Peter, I.S., Davidson, E.H., 2010. The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. *Dev Biol* 340, 188-199.
- Poustka, A.J., Kuhn, A., Groth, D., Weise, V., Yaguchi, S., Burke, R.D., Herwig, R., Lehrach, H., 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.
- Ransick, A., Ernst, S., Britten, R.J., 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.
- Roegiers, F., Jan, Y.N., 2000. Staufen: a common component of mRNA transport in oocytes and neurons? *Trends Cell Biol* 10, 220-224.
- Saunders, P.T., Pathirana, S., Maguire, S.M., Doyle, M., Wood, T., Bownes, M., 2000. Mouse *staufen* genes are expressed in germ cells during oogenesis and spermatogenesis. *Mol Hum Reprod* 6, 983-991.
- Siebert, S., Farrell, J.A., Cazet, J.F., Abeykoon, Y., Primack, A.S., Schnitzler, C.E., Juliano, C.E., 2019. Stem cell differentiation trajectories in *Hydra* resolved at single-cell resolution. *Science* 365.

- Song, J.L., Wessel, G.M., 2012. The forkhead transcription factor FoxY regulates Nanos. *Mol Reprod Dev* 79, 680-688.
- Soper, S.F., van der Heijden, G.W., Hardiman, T.C., Goodheart, M., Martin, S.L., de Boer, P., Bortvin, A., 2008. Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Dev Cell* 15, 285-297.
- St Johnston, D., Beuchle, D., Nusslein-Volhard, C., 1991. Stauf, a gene required to localize maternal RNAs in the Drosophila egg. *Cell* 66, 51-63.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., Satija, R., 2019. Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888-1902 e1821.
- Swartz, S.Z., Reich, A.M., Oulhen, N., Raz, T., Milos, P.M., Campanale, J.P., Hamdoun, A., Wessel, G.M., 2014. Deadenylation protects inherited mRNAs in primordial germ cells. *Development* 141, 3134-3142.
- Takacs, C.M., Amore, G., Oliveri, P., Poustka, A.J., Wang, D., Burke, R.D., 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.
- Tu, Q., Brown, C.T., Davidson, E.H., Oliveri, P., 2006. Sea urchin Forkhead gene family: phylogeny and embryonic expression. *Dev Biol* 300, 49-62.
- Tu, Q., Cameron, R.A., Davidson, E.H., 2014. Quantitative developmental transcriptomes of the sea urchin *Strongylocentrotus purpuratus*. *Dev Biol* 385, 160-167.
- Urry, L.A., Hamilton, P.C., Killian, C.E., 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.
- Voronina, E., Lopez, M., Juliano, C.E., Gustafson, E., Song, J.L., Extavour, C., George, S., Oliveri, P., McClay, D., 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.
- Wang, Y., Zayas, R.M., Guo, T., Newmark, P.A., 2007. nanos function is essential for development and regeneration of planarian germ cells. *Proc Natl Acad Sci U S A* 104, 5901-5906.
- Wei, Z., Yaguchi, J., Yaguchi, S., Angerer, R.C., Angerer, L.M., 2009. The sea urchin animal pole domain is a Six3-dependent neurogenic patterning center. *Development* 136, 1179-1189.
- Wessel, G.M., Morita, S., Oulhen, N., 2020. Somatic cell conversion to a germ cell lineage: A violation or a revelation? *J Exp Zool B Mol Dev Evol*.
- Wohlschlegel, J.A., Kutok, J.L., Weng, A.P., Dutta, A., 2002. Expression of geminin as a marker of cell proliferation in normal tissues and malignancies. *Am J Pathol* 161, 267-273.
- Wu, X., Ruvkun, G., 2010. Cancer. Germ cell genes and cancer. *Science* 330, 1761-1762.
- Yaguchi, S., Yaguchi, J., Angerer, R.C., Angerer, L.M., 2008. A Wnt-FoxQ2-nodal pathway links primary and secondary axis specification in sea urchin embryos. *Dev Cell* 14, 97-107.

- Yaguchi, S., Yaguchi, J., Wei, Z., Shiba, K., Angerer, L.M., Inaba, K., 2010. ankAT-1 is a novel gene mediating the apical tuft formation in the sea urchin embryo. *Dev Biol* 348, 67-75.
- Zazueta-Novoa, V., Wessel, G.M., 2014. Protein degradation machinery is present broadly during early development in the sea urchin. *Gene Expr Patterns* 15, 135-141.

Figures

A.



B.

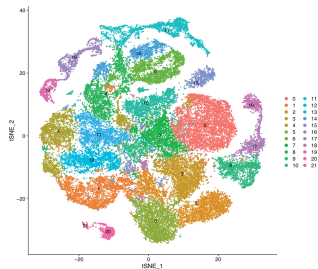


Fig 1. (A) tSNE plots of major developmental stages: 8-cell stage, 64-cell, morula, early blastula, hatched blastula, mesenchyme blastula, early gastrula, and late gastrula after alignment. 60,399 total single cells are colored by cluster identity. Twenty-two cell populations are detected across the 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 resolution 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 (left).

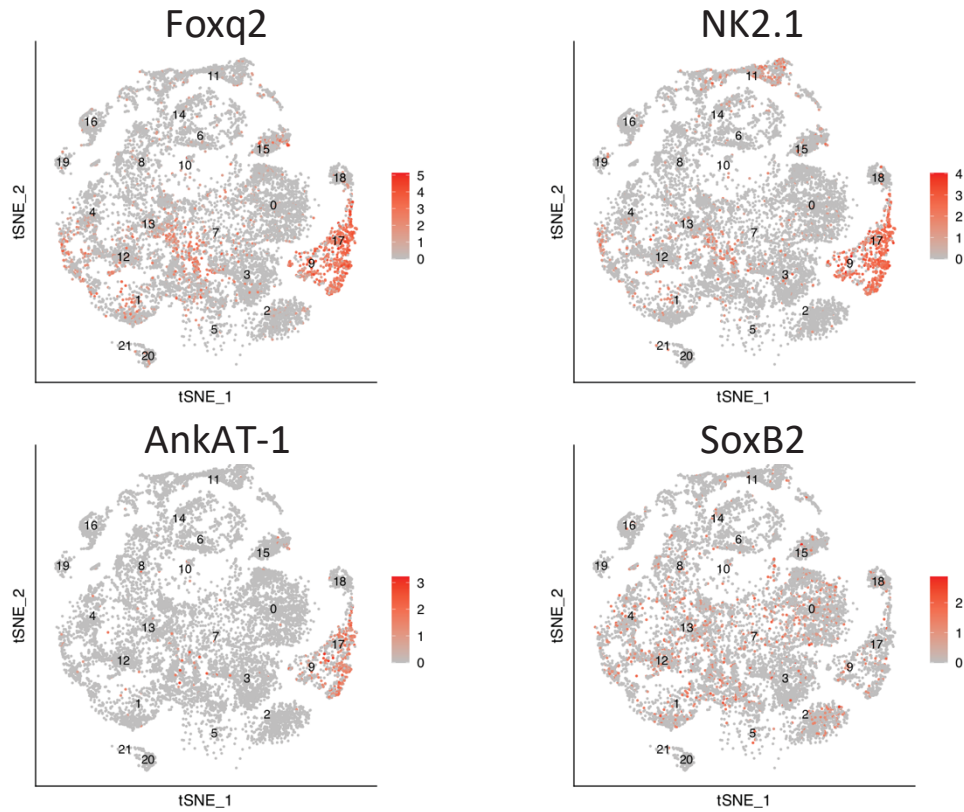


Fig 2. Feature plots showing coexpression of neural precursor genes at hatched blastula stage. Apical ectoderm markers, NK2.1, Foxq2, and AnkAT-1 expression enriched in clusters 9, 17. Neural precursor gene, SoxB2, known to be expressed throughout the ectoderm seen more broadly expressed. Cells are colored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

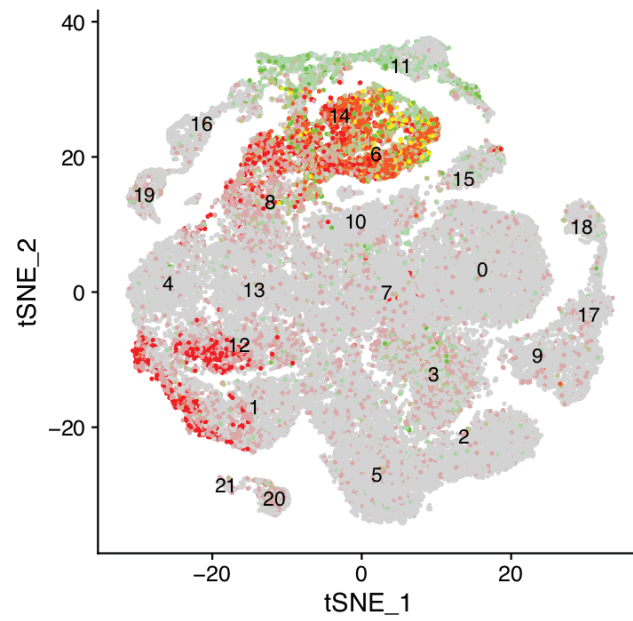


Fig. 3. Feature plot showing expression of FoxA (red) and Endo16 (green) across the dataset. Coexpression (yellow) of FoxA and Endo16 in clusters 6 and 14.

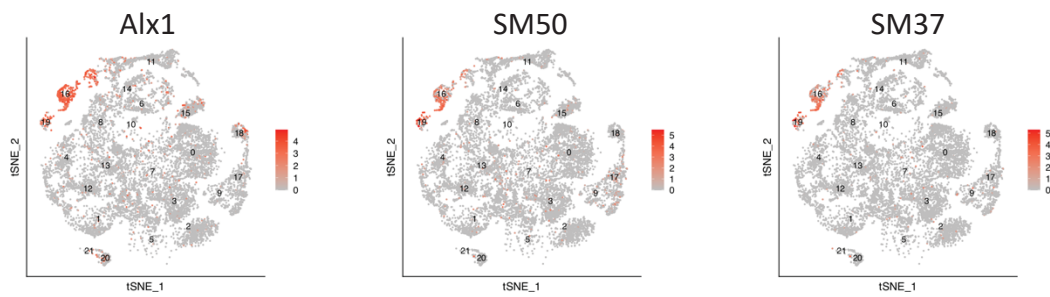


Fig. 4. Clusters 16 and 19 marked by expression of PMC marker genes, Alx1, SM50, SM37, at hatched blastula stage. Cells are colored by relative expression level.

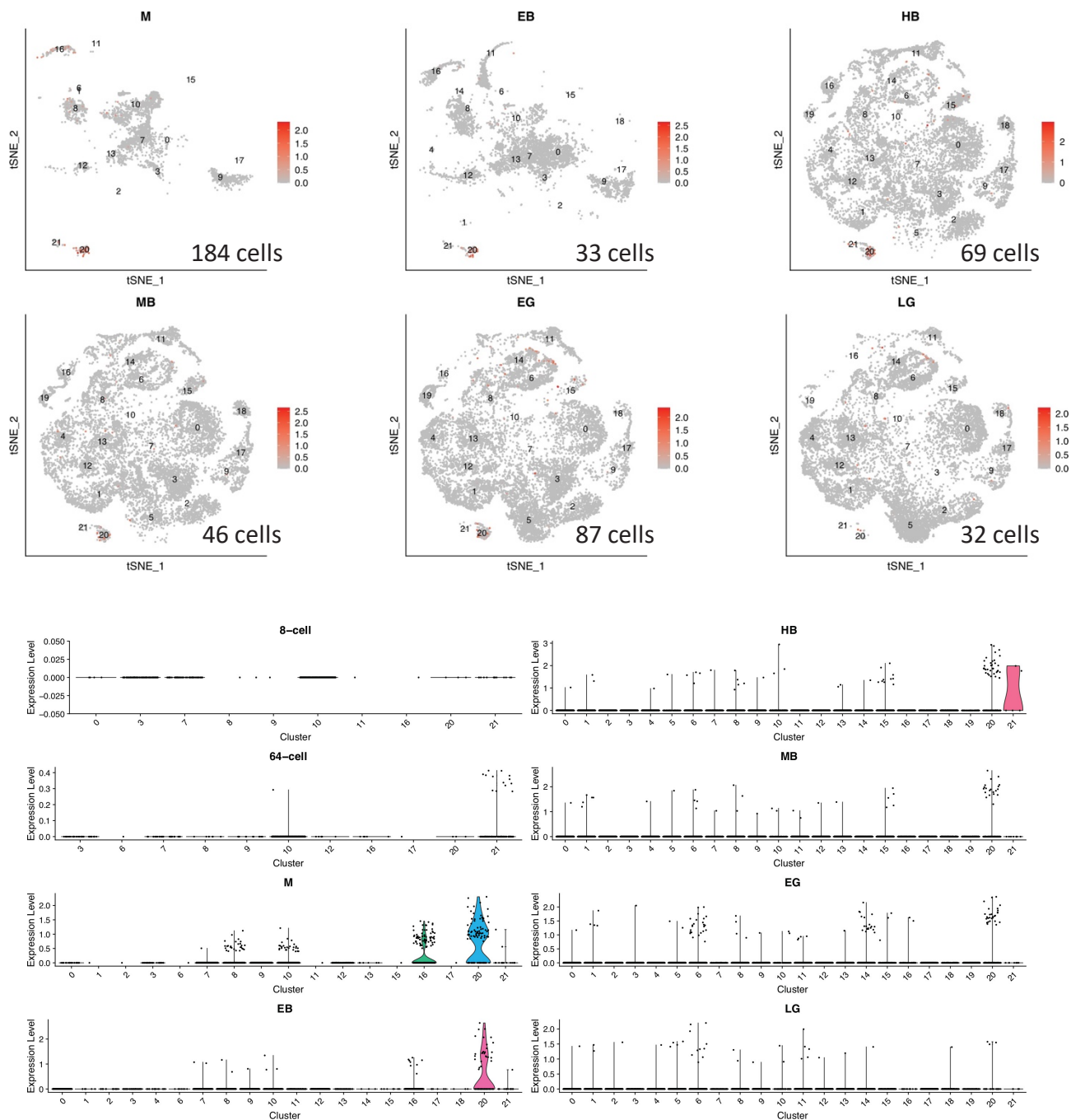
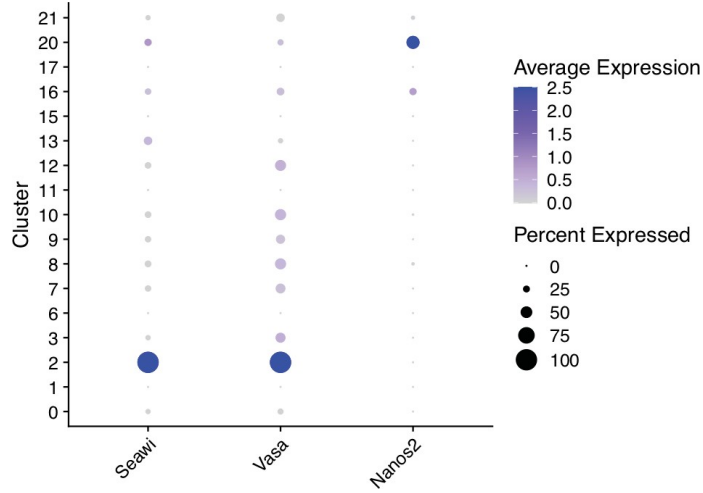
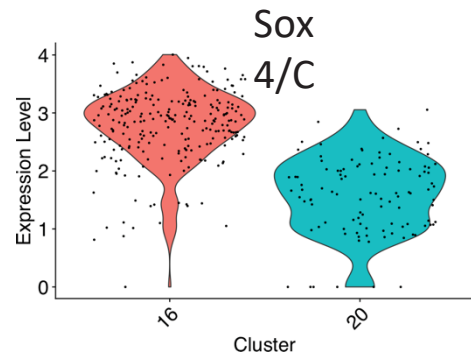


Fig. 5. Feature and violin plots of relative Nanos2 expression per cluster (cell state) at different developmental time points. Number of Nanos2 positive cells detected per time point included in feature plots. Cells are included based on normalized expression level with a lower threshold set to 1, except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

A.



C.



B.

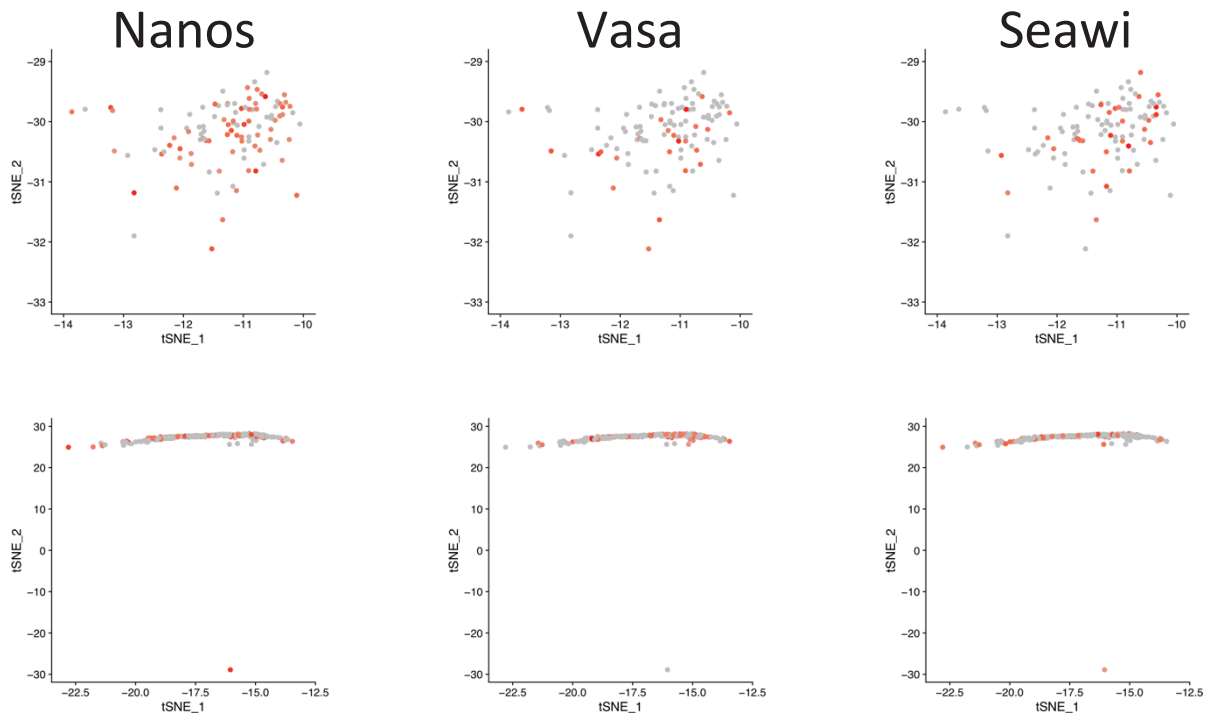


Fig. 6. Coexpression of germ line genes in clusters 16 and 20 at morula stage. (A)

Dotplot showing expression of germ line markers *nanos2*, *vasa*, and *seawi*. Dot size

relates percent of cells expressing 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.

Chapter 4: Distinct mechanisms of germ cell factor regulation for an inductive germ cell fate

Foster, S., Oulhen, N., Fresques, T., Zaki, H., Wessel, G., Distinct mechanisms of germ cell factor regulation for an inductive germ cell fate.

I performed the cell signaling perturbation, morpholino, and repressor construct experiments in this study. I performed the single cell sequencing protocol with support from Nathalie Oulhen. I performed the data analysis independently, and wrote the manuscript. Tara Fresques performed the experiment for and created the figures for Supplementary Figure S7A and Supplementary Figure S11. Hossam Zaki designed the morpholinos for the sea star FoxYs.

Highlights

- 1) Single-cell RNA-sequencing of sea star embryos demonstrates temporal differences in cell fate commitment among echinoderms.
- 2) Sea urchin and sea star embryos appear to ascribe their germ line by two extreme different mechanisms but share similar pathways in regulation of the germline genes.
- 3) Expression of the germline factors, Vasa and Nanos, is regulated by distinct mechanisms in the sea star.
- 4) Germline induction in the sea star uses similar signaling mechanisms as mammals.

Key words:

Inductive germline, germ cell, Vasa, Nanos, sea star, sea urchin, cell fates, scRNA-seq

Abstract

Specification of primordial germ cells (PGCs), the lineage which gives rise to eggs and sperm, is essential for sexually reproducing organisms. The mechanism by which animals specify their PGCs generally falls into two categories: inherited or inductive. The inductive mechanism, used by mammals, relies on cell signaling interactions to direct a subset of embryonic cells to a germ cell fate. Previous work suggested that sea star embryos, which develop in simple culture and are markedly transparent, also use inductive mechanisms to specify their germline. The germ cell factors Nanos and Vasa become restricted during early development into a localized region of cells within the posterior enterocoel (PE), the presumptive germline. Nodal signaling was observed to negatively regulate Vasa and Nanos mRNAs outside of the PE and restrict the germline to the PE. Here we employed single cell RNA sequencing to identify the transcriptional program of germ cells and their changes during development. We never see Nodal pathway members within Nanos/Vasa positive cells in the region known to give rise to the PE, and instead see members of the Wnt-signaling pathway and the FoxY family of transcription factors. We learned that Wnt and Delta/Notch signaling enhances expression of both Nanos and Vasa, whereas a test of dependency in cell interactions reveals that Nanos and Vasa are regulated distinctly. The results from this work contrasts the results of PGC specification in the sea urchin and enables deeper comparative mechanistic studies in tractable *in vivo* models.

Introduction

Sexually reproducing organisms have a distinct germline, the cell lineage that will give rise to either eggs or sperm in the adult, enabling a continuation of species. Two major mechanistic distinctions have been identified for specification of the primordial germ cells, the cells committed early in development to the germline. These are referred to as the inductive and the inherited mechanisms also referred to as epigenesis and preformation, respectively [1]. In the inherited mechanism, used by species including the insect *Drosophila melanogaster*, the frog *Xenopus laevis*, and zebrafish *Danio rerio*, localized maternal RNA transcripts and proteins in the egg are inherited by a particular set of cells in the embryo, which then give rise to gametes in the adult [1, 2]. In contrast, in the inductive mechanism, the germline is formed later in development through zones of cell-cell signaling. Current evidence suggests that a combination of activating and inhibitory signals converges to induce germline specification in a restricted region of the embryo. This mechanism is used by species including mammals, the cricket *Gryllus bimaculatus*, and the amphibian *Ambystoma mexicanum* (axolotl). Studies in organisms across many different phyla place the inductive mechanism as the ancestral mechanism of germline formation [1]. However, the full network of cell signaling pathways and their targets which direct the ancestral mode of germline formation remains poorly understood. A role for BMP signaling in germline specification is thought to be conserved. Loss of BMP signaling results in loss of primordial germ cells in mouse and cricket embryos [3, 4]. Wnt signaling was identified to induce PGC specification in both mouse and axolotl, although the specific Wnt ligand is not conserved [3, 5, 6].

The sea star, *Patiria miniata*, follows an inductive mechanism of germline formation [7].

This is in contrast to how its close relative, the sea urchin, forms its germline which instead relies on maternal factors. This divergence provides an important comparison between organisms within a taxon. Further, the sea star provides a transparent, tractable embryo, closely related to chordates, for *in vivo* analysis. Expression of the highly conserved germline marker genes Nanos and Vasa, among others, has been characterized [8]. Nanos and Vasa expression is seen in a vegetal ring at mid-gastrula stage. Later, Nanos and Vasa expression is lost in the ventral gut and then the right side of the developing gut at late gastrula stage. Finally, at the larval stage, Nanos and Vasa expression is restricted to the dorsal left side of the gut and those cells form the posterior enterocoel (PE). We refer here to the PE as the presumptive germline as previous work has shown that removal of the PE results in significantly fewer germ cells in larva compared to controls [9]. The PE also accumulates Nanos and Vasa transcripts selectively in the larva supporting its germline contention. The lineage of cells in the PE though, have not yet been traced through development to adulthood. Initially, Nanos and Vasa RNA are expressed broadly in the forming gut, and Nodal signaling inhibits expression of the germline genes, restricting germline gene expression eventually to the left side. A Nodal signaling gradient on the ventral side at blastula stage, and on the right side in late gastrula stage are responsible for restriction of Nanos and Vasa to the dorsal left side, in the PE [7]. While Nodal has been identified to play a role in restricting the germline by inhibiting expression of the germline genes, the signal or signals that activate germline gene expression is unknown.

Single-cell RNA-sequencing (scRNA-seq) is a powerful tool used to identify individual cells that make up an entire embryo. Single-cell RNA sequencing has identified

germline cells of *Xenopus*, zebrafish, and sea urchin embryos, and has been employed to identify transcriptomic changes in germline cells with development in male and female mice [10-14]. We performed scRNA-seq in the early sea star to identify the cell states expressing germline factors in the embryo, to test the cell signaling pathway components responsible for their fate, and to leverage analysis of the distinct mechanism used for germ cell formation in the sister taxon of sea urchins.

Results

Identification of cell states across early sea star development

Sea star embryos were cultured to six different developmental stages, 8 hpf, 10 hpf, 14 hpf, blastula, early gastrula, and mid-gastrula stage, at which point they were dissociated and processed for scRNAseq via Drop-seq [15]. All six datasets were analyzed using Seurat and integrated using Harmony, a single-cell RNA-sequencing data integration tool, to identify conserved cell types across all datasets [16, 17]. This analysis of 25,703 total cells yields 20 clusters, or cell states, across all timepoints identified by known marker gene expression (Figures 1A and 1B; Table S6). Here we use cell state to refer to a select accumulation of transcripts present in a cell as it reaches a developmental fate. In analyzing the six datasets together, it is evident that early (8-14h) and late (B-MG) cell types are transcriptionally distinct. Cell clusters of early stages (Clusters 3 and 7) decrease with development; these cells presumably are “lost” to differentiation (Figure S1). Clusters 3, 7, and 9, which make up the bulk of the earlier stage datasets, showed enrichment of CyclinA and lack of marker gene expression (Figure 1B).

Blastula

We assessed the blastula and gastrula stage datasets in greater detail as gene expression in these developmental stages have been characterized previously and germline marker gene expression is present. In blastulae, we detect 10 cell states which are identified by expression of such known marker genes (Figures 2A and 2B; Tables S1 and S3). We identified ectodermal cells by expression of SoxB1 [18]. Foxq2 and DKK3 are expressed in the animal pole domain where the apical organ of the nervous system

will form [19]. Cluster 0 expresses high levels of SoxB1, FoxQ2, and Dkk3. Cluster 1 expresses FoxQ2 and Dkk3 which we identify as the animal pole domain (Figure S2). While these apical ectodermal markers are widespread in the dataset, the vegetal markers are restricted to a subset of cell states. These vegetal markers include both mesodermal and endodermal progenitors. Ets1/2, a marker of the presumptive mesoderm, is expressed in the vegetal plate [20]. GataE, FoxA, brachyury, and Blimp1, all genes involved in endodermal development, will mark different regions of the embryonic gut. However, in blastulae, they are all expressed in the vegetal ectoderm, surrounding the blastopore [8, 21]. Cdxl, another endodermal marker, is also expressed as a ring, vegetally in blastulae; it is later restricted to the blastopore region in mid-gastrulae [22]. GataE, FoxA, and Blimp1 are expressed in the same cluster as Ets1/2 (cluster 2) while brachyury and Cdxl are markers of cluster 3 (Figures 2B and S2). At blastula stage, the transcription factor glial cells missing (gcm), which marks the pigment cells in a related species, the purple sea urchin shows no localized expression; gcm-positive cells are present throughout the ectoderm [23, 24]. In our dataset, cells with highest gcm expression clustered into a unique population in cluster 8 with some gcm-positive cells present within the ectodermal clusters (Figures 2B and S2).

Expression of the Wnt signaling ligands and receptors has previously been characterized in blastula and gastrula stages [25]. As signaling plays an important role in development and inductive germline specification, we sought to characterize Wnt ligand expression in our data. Our dataset showed coexpression of ligands Wnt3, Wnt16, and the Wnt receptor, Frizzled1 (Frizz), in cluster 2 (Figure S4). These three are known to be expressed in the vegetal pole at blastula stage. The ligand Wnt8 is also expressed in the

vegetal pole, not overlapping with Frizz expression. Additionally, the Wnt8 expression pattern showed overlap with Wnt3 (cluster 3) and non-overlap with Wnt3 (cluster 7) mirroring what was previously reported (Figure S4) [25]. Vasa mRNA is expressed in the vegetal plate of blastulae and is particularly enriched in cluster 2, like the Wnt3 and Wnt16 ligands, and the receptor Frizz (Figure S6H) [8]. Nanos accumulation is seen by in situ hybridization later, in mid-gastrulae, but is not detected in blastulae [26]. Indeed, Nanos expression was detected at low level at blastula stage in our dataset (Figure S6A).

Mid-gastrula

Analysis of the mid-gastrula stage dataset reveals 11 cell states (Figure 3A; Tables S1 and S4). At this stage, many characteristics of the animal ectoderm remain the same. Ectodermal cell types (cluster 0,1,2,4) are still distinguished by expression of SoxB1, FoxQ2, and Dkk3 (Figure S3). Now, expression of several vegetal markers detected at blastula stage, changes from overlapping to marking either vegetal ectoderm or different regions of the gut. In blastula stage, both Ets1/2 and GataE were expressed in the same cluster (cluster 2), at gastrula stage, Ets1/2 is particularly enriched in cluster 3, and GataE marks a different cluster (cluster 6) (Figure 2B). Ets1/2 at gastrula stage is present at the tip of the archenteron where it marks presumptive mesoderm and ingressing mesenchyme cells [20, 27]. There is some low level GataE expression in the archenteron, however, GataE is enriched in the mid- to hindgut region in mid-gastrula [21]. FoxA, which is expressed in the mid/hindgut region, is enriched in the same cluster as GataE [28]. In mid-gastrulae, brachyury expression surrounds the blastopore but is not within the blastopore [28]. Cdxl expression is restricted to the region around the

blastopore [22]. Brachyury and Cdxl are both enriched in cluster 8 (Figure S3).

Lack of unique marker gene expression makes cluster 5 difficult to identify. Cluster 5 showed enrichment of ECM protein expression, including fibrinogen (PMI_003448 Fic_9) and fibrillin (PMI_026729 Fbn3) (Table S4). Similar to blastula stage, one cluster is marked by high gcm expression, cluster 9. A small number of cells cluster on their own (cluster 10) and these cells are marked by Mifl1, macrophage inhibitory factor like 1 (Figure 3B). MIF plays a role in immune function in vertebrates [29, 30]. Coexpression of gcm and another MIF, mif5, was previously reported in sea urchin larvae [31]. Cluster 10 also expresses Mt1-4, a metalloprotease, and Fos, a transcription factor, both enriched in the sea urchin primary mesenchyme cells (PMCs) (Figure S5) [32, 33]. Further studies are necessary to reveal what role these genes play in sea star development as sea star embryos lack PMCs and pigment cells.

The sea star embryo is known to have left/right morphological asymmetry by gastrula stage. A Nodal signaling gradient on the right side of the embryo is responsible for restricting PE formation to the left side [7]. Nodal expression, along with Lefty, is enriched in one cluster (cluster 7) which we identify as the right-side ectoderm (Figure 2B). In contrast to Nodal signaling, Wnt pathway component expression is present in the same region as Nanos and Vasa RNA expression in the early embryo and in the forming gut during the gastrula stage. In situ hybridization shows enrichment of Nanos and Vasa transcripts at the top of the archenteron and in a vegetal ring in the hindgut region at mid-gastrula stage [7]. Six Wnt ligand proteins and three proteins for Frizzled (Fz), the Wnt receptor, are present in the *P. miniata* transcriptome [25]. Among the six

Wnt ligands, RNA expression patterns of Wnt3, Wnt8, and WntA/4 show overlap with the region of Nanos and Vasa RNA expression in the gastrula stage embryo [7, 25]. Among the three Wnt receptors, Fz1/2/7 RNA expression is concentrated in the forming gut of the early gastrula [25]. The three signaling ligands, Wnt3, Wnt8, and WntA/4, show appropriate expression patterns to be candidates for initiating germline specification by interacting with the Fz1/2/7 receptor, due to their tight colocalization with the Nanos and Vasa RNA expression pattern in the early embryo. At mid-gastrula, the Wnt signaling ligands have been described as having a nested gene expression pattern. Wnt16 shows most vegetal expression followed by Wnt3 and then Wnt8 along the ectoderm. In our dataset, Wnt16 enrichment is seen in cluster 6 with Wnt3 expression. Wnt3 is also coexpressed with Wnt8 in cluster 8. Frizz is present in cluster 3, which we identify as the tip of the archenteron (Figure S4) [25].

Derivation of the germline precursor cells

We focused our analysis on the mid-gastrula stage, where Nanos and Vasa are both expressed and will be restricted to the region of PE formation. Nanos expression is detected in 267 total cells, co-expression of Nanos and Vasa is seen in 212 cells. Nanos expression is much less abundant compared to Vasa; Vasa expression is seen in all clusters at mid-gastrula stage but expression in the hindgut cluster (cluster 8) was of greatest interest to us (Figure S6H). We took a closer look at the Vasa-positive cells in the hindgut cluster and compared gene expression in the hindgut cells that express Vasa to those that do not. In extracting a list of differentially expressed genes in Vasa-positive cells of the hindgut cluster, we find the transcription factor FoxY3 (Table S5). In situ hybridization analysis shows FoxY3 is

expressed in the same regions as both Nanos and Vasa at mid-gastrula stage (Figure S7A). FoxY3 expression is enriched in clusters 3, 6, and 8, which correspond to the archenteron and gut region, similar to Vasa expression (Figures S6H and S6I). A role for FoxY was also previously identified in regulating Nanos expression downstream of Delta/Notch signaling in the sea urchin [34]. ScRNA-seq analysis reveals that Vasa and FoxY3 are co-expressed (Figure S7B). Wnt8 and Wnt3 are expressed in the cells neighboring the Vasa, FoxY3-positive cells; Wnt3 is also expressed in the Vasa, FoxY3-positive cells, along with Wnt16 and Wnt1 (Figure S7C; Table S5). Moreover, at this stage, Vasa-positive cells express hindgut identity genes such as GataE and FoxA, and the homeobox gene Hox11/13b (Table S3; Figure S7C).

Wnt signaling regulates Vasa expression

The coexpression of Wnt signaling ligands and a downstream target of Delta/Notch signaling, FoxY3, in Vasa-positive cells led us to test whether Wnt and Delta/Notch signaling may regulate germline factor expression in the sea star. To test whether ectopic activation of canonical Wnt signaling alters *P. miniata* germline factor gene expression, we treated embryos with the Wnt signaling activator LiCl over a developmental time course and measured effects on levels and distribution of the germline-specific transcripts Nanos and Vasa [35]. Embryos were cultured in fresh sea water and treated with 5mM LiCl for 24 hours, at three time points: one-day post-fertilization (Day 1), two-days post-fertilization (Day 2), or three-days post-fertilization (Day 3) (Figure 4A). At 24 hours after each treatment time point, Nanos and Vasa transcript levels were measured by whole-embryo quantitative RT-PCR (qPCR).

Untreated embryos cultured and collected under the same conditions and timing were used as controls.

When Wnt signaling is upregulated, Vasa mRNA abundance is upregulated compared to controls; Nanos expression is unaffected (Figure 4B). Consistent with the qPCR results, Vasa mRNA expression expands with LiCl treatment at day 2 (Figure 4C). At this time point, the control embryos display Vasa staining as a rounded zone on the left edge of the gut. However, the LiCl-treated embryos display Vasa staining as a band from the left edge to the right edge of the gut. It is important to note that Vasa expression was not upregulated throughout the embryo but rather in a band along the gut. This result suggests that Wnt signaling regulates Vasa expression in the embryo, at gastrula stage. To further test whether Vasa is a direct target of the canonical Wnt signaling pathway, we made use of a β -catenin/Engrailed mRNA construct coding for a protein in which the transactivation domain on the β -catenin C-terminus is replaced with the repression domain of the *Drosophila* Engrailed gene [34, 36]. Embryos injected with the transcriptional repressor construct showed a significant downregulation of Vasa transcripts at gastrula stage, while Nanos mRNA level was unaffected (Figure 4D). This result suggests that in gastrulae, Vasa may be a direct transcriptional target of canonical Wnt signaling. We have identified four putative HMG-box binding motifs recognized by β -catenin and the transcription factor Tcf/Lef upstream of the Vasa promoter region that we will test in the future to determine whether Wnt regulation of Vasa is direct (5'-(T/A)ACAAAG-3') [34, 37].

Delta/Notch signaling regulates Nanos and Vasa expression

Delta/Notch signaling was found to regulate Nanos expression through the FoxY transcription factor, in somatic cells adjacent to the PGCs in the purple sea urchin [34]. In the sea star, the delta ligand is expressed in the same region of Nanos and Vasa expression, first in the invaginating gut of early gastrula, and then in the midgut region [27]. To test whether Delta/Notch may regulate germ cell factor gene expression in the sea star, we treated two-day old embryos with the Delta/Notch signaling inhibitor, DAPT [38]. Expression of both Nanos and Vasa is significantly decreased with DAPT treatment compared to controls. Expression of FoxY2 and FoxY3 is also decreased with DAPT treatment (Figure 5). *P. miniata* has three FoxY transcription factors, expression of all three is present at mid-gastrula, in the same region as Nanos and Vasa expression. By late gastrula stage, expression of FoxY1 and FoxY2 is lost (Figure S11). FoxY3 is expressed in the same domain as Nanos and Vasa at mid- and late gastrula stage (Figure S7A). Furthermore, we detect FoxY3 and Vasa coexpression at mid- gastrula by scRNA-seq (Figure S7B). To test whether FoxY3, downstream of Delta/Notch signaling, functions to regulate germline gene expression in the sea star, we used a morpholino approach to block translation of the protein. We do not see a significant change in mRNA transcript abundance for Nanos or Vasa with FoxY3 knockdown or knockdown of the other two FoxY transcription factors. We do note a significant upregulation in FoxY1, FoxY2, and FoxY3 transcripts with morpholino treatment suggesting that the three FoxY transcription factors behave as transcriptional repressors and regulate themselves through a negative feedback loop or that morpholino binding protects the transcripts from degradation (Figure S8). This data suggests that although FoxY3 is expressed in the same domain as Nanos and Vasa, it does not directly regulate their expression. The downstream Notch signaling effector, suppressor of hairless (SuH) is also

downregulated with Delta/Notch inhibition (Figure 5). A morpholino or CRISPR/Cas9 approach targeting SuH would be a powerful test for whether Nanos and Vasa are direct transcriptional targets of Delta/Notch through SuH activity.

Analysis of cell interactions

The small molecule perturbation experiments and morpholino approaches help reveal the transcriptional mechanisms involved in germline factor expression in the sea star embryo. However, these experiments do not address the question of what or whether cell interactions are required for germline gene expression. We took advantage of the ability to dissociate embryos into single cells to test this premise. In having dissociated cells suspended in sea water we have inhibited signaling through cell contact and we have greatly diluted signaling ligands. Sea star embryos were cultured until hatched blastula stage at which point Nanos mRNA abundance is low [26]. Embryos were then dissociated and cultured for four hours, to early gastrula stage (Figure 6A).

Dissociating cells at this stage in development allows us to ask the question of whether cell interactions are required to initiate Nanos expression and/or to maintain Vasa expression. Remarkably, Nanos transcript levels were significantly upregulated in dissociated cells compared to intact blastula and gastrula stage embryos whereas Vasa mRNA levels did not reach similar levels as in the intact gastrula. Nodal mRNA level was significantly decreased in dissociated cells compared to gastrula control, as might be expected when considering a positive feedback loop for Nodal in intact embryos [39,40]. Dissociated cells showed upregulated Wnt8 expression and downregulated FoxY3 expression, consistent with a loss of cell contact and thus loss of Delta/Notch signaling. The level of Blimp1 mRNA, expressed in the same region as

Nanos and Vasa, does not change with dissociation (Figure 6B). However, expression of endomesodermal genes known to correlate with nuclear β -catenin activity in the sea star embryo are downregulated in dissociated cells (Figure S9C) [41]. Additionally, Nanos mRNA abundance continues to increase in dissociated cells cultured for up to 18 hours (Figure S9B).

A potential explanation for the upregulation in Nanos transcripts is the drop in Nodal. Since Nodal represses Nanos mRNA accumulation, separation of cells from this interaction may minimize such a Nodal phenotype. To test the effect of Nodal signaling on dissociated cells, we treated dissociated cells with the Nodal signaling activator, recombinant human activin AB (rhActivinAB). In dissociated cells treated with rhActivinAB, Nodal expression is upregulated, the spike in Nanos mRNA seen with dissociation is lost, and Vasa expression is now upregulated (Figure 6C). We previously reported an inhibitory effect of Nodal on Nanos expression [7]. A loss of Nanos expression with rhActivinAB treatment suggests that Nodal acts directly on Nanos expressing cells to regulate Nanos expression. At this time, it is unclear whether the upregulation of Nanos transcripts in dissociated cells compared to controls is due to a greater number of cells expressing Nanos or a small population of cells greatly upregulating Nanos expression. The former would indicate that Nanos expression is cell autonomous and Nodal plays an essential role in restricting Nanos expression throughout the embryo.

Discussion

Inductive signals in sea star germ cell specification

Our current understanding of the inductive mechanism for germ cell specification comes from studies performed using mouse, cricket, and axolotl embryos. Identifying activating and inhibitory cell signaling pathways and their targets is essential to understand germline induction across different species, how these processes may be disrupted to effect fertility, and how they can be utilized for gamete development studies in vitro [42, 43]. The sea star possesses many strengths as a model organism for study of the inductive mechanism including access to greater numbers of oocytes, and manipulation and testing of early embryos at different points in the specification process. Here we present a role for Wnt and Delta/Notch signaling in inductive germline specification of sea star embryos through regulation of the germline genes, Nanos and Vasa. We propose that both Wnt and Delta/Notch signaling regulate Nanos and Vasa expression at gastrula stage. These pathways could be active concurrently with Nodal restriction which originates from a gradient on the right side of the embryo. While Delta/Notch signaling affects both Nanos and Vasa expression, Vasa alone seems to be under Wnt/ β -catenin regulation. This suggests that although Nanos and Vasa are expressed in the same regions, and similarly repressed by Nodal, how that expression pattern comes to be is distinct. Further investigation is necessary to uncover whether both genes are direct transcriptional targets of Delta/Notch and Wnt. We also report a difference in Nanos and Vasa regulation through dissociation experiments. In this setting, with cell signaling removed, Nanos transcript abundance increases and continues to do so with time, while Vasa transcript abundance does not. This experiment reveals that activating signals are essential for initiating and maintaining Vasa expression in the embryo. Nanos does not

seem to face the same regulation. Future studies will reveal whether Nanos expression is cell autonomous and if cells throughout the early embryo are capable of expressing Nanos in the absence of a repressor. It is possible that particularly for Nanos, Nodal repression is crucial in promoting proper development of the germline and the embryo as a whole.

At the sea star mid-gastrula stage, Vasa transcript is expressed more broadly than in the future germline, this contrasts its expression in the sea urchin where it is restricted to the small micromeres in gastrulae and the mouse where it is expressed only after the PGCs have colonized the embryonic gonad [44, 45]. This is a distinct strategy of cell-specific germ cell factor accumulation in which expression is spread broadly then cleared from cells to yield highly localized expression. In the sea star, Vasa protein is also under strict control through selective post-translational degradation by Gustavus, an E3 ligase [46]. In the future, investigating how Nanos protein localization is regulated will inform us whether these transcripts are also under distinct regulation at the protein level.

During gastrulation, Vasa and Nanos mRNA expression is present in the archenteron. And in larvae, Nanos and Vasa expression is present in the PE as well as the anterior pouches [7]. We detected Nanos and Vasa transcripts as well as FoxY3 at the tip of the archenteron (mid-gastrula cluster 3) (Figures S6G-I and S7). What role these transcripts, particularly Nanos and Vasa, have in this region is unknown. A notable difference in germline development in the sea urchin compared to sea star is that a PE structure is absent. Instead, the germline determinants, including Nanos and Vasa, accumulate in the small micromeres and become incorporated into both coelomic

pouches before becoming restricted to the left pouch [44]. Future studies are essential to determine whether these cells expressing Nanos and Vasa at the tip of the archenteron have a function in germline development conserved in the sea urchin embryo.

Distinct mechanisms of germ line specification in two echinoderm species

In comparable sea urchin developmental stages, more cell states and regions of the embryo are partitioned in distinct clusters than in the sea star. For example, at early blastula stage in the sea urchin, skeleton, pigment, neural cell states, and distinct oral ectoderm cell states are present. By early gastrula, scRNA-seq identified six ectodermal cell states and three neural cell states [12]. Of note, the skeletogenic and pigment cell lineages are mesodermal cell types of the sea urchin, absent in the sea star embryo.

The sea star (*P. miniata*) and sea urchin (*P. lividus*) have been compared to assess heterochronies in marker gene expression, revealing a later onset of maternal to zygotic transition in the sea star compared to this sea urchin [47]. It is possible that this temporal difference plays a role in the diversity of cell states present at the early developmental stages when comparing the two species. An alternative explanation is that cell fates in the sea star rely more on cell signaling for fate decisions than in the sea urchin, and the sea star embryonic cells are less committed than in the comparable stage of a sea urchin.

Echinoderms offer an opportunity to investigate two modes of germ cell development in closely related species and uncover evolutionary differences. Nanos and Vasa eventually localize to putative PGC precursors in both species (the small micromeres in

sea urchins and in the PE in the sea star), yet their embryonic expression patterns are vastly different. Early in sea urchin embryogenesis, Nanos and Vasa expression is restricted to the small micromeres [44]. In the sea urchin, Nanos transcription is regulated by direct action of β -catenin in the small micromeres or the Delta/Notch transcription factor, FoxY, in the adjacent somatic Veg2 mesoderm cells [34]. In the sea star, Vasa mRNA expression is seen earlier than Nanos, present in the oocyte, in the vegetal pole of blastulae, and later with Nanos in the hindgut and archenteron regions [7, 8, 26]. Similar to the sea urchin, Delta/Notch regulation of Nanos is also present in the sea star (Figure 5). However, we propose that the FoxY transcription factors present in the sea star are not transcriptional regulators of Nanos. This could point to an evolutionary shift between the two species where FoxY plays a role in germ line specification but its transcriptional targets have changed. A difference in Wnt signaling effect is also evident. Nanos expression in the small micromeres is regulated by β -catenin in the sea urchin. Although we do not report an effect on Nanos expression through Wnt signaling activation or repression in the sea star, analysis of different stages of development as well as the Nanos regulatory region are necessary to reach this conclusion. Notably, removal of the micromeres, the parent cells of the small micromeres, results in adults with developed gametes [48]. Whether Nanos expression, downstream of FoxY, in the Veg2 cells plays a role in reconstituting the germline remains an interesting question and could point to an auxiliary, inductive mechanism present in the sea urchin. Testing effects of cell signaling perturbation in micromere- deleted embryos will help decipher this process.

Evolution of germ line specification

Evidence argues the inductive mode (epigenesis) in metazoans such as sponges, jellyfish, and hydra is the ancestral mechanism of germ cell specification [1]. Discovery of an inductive mode for germline specification in the sea star, an echinoderm species, presents an opportunity for comparison with the mouse, a mammalian species for which we have the most comprehensive description of the inductive mechanism (Figure 7). Mouse PGCs acquire their fate in response to signaling. They arise from mesodermally differentiating cells. Single-cell analysis revealed that early germ cells express homeobox genes, including *Hoxb1* and *Hoxa1*, and the mesodermal marker *brachyury*, at the same level as their somatic cell neighbors. The early germ cells then initiate transcriptional repression of homeobox genes which remain highly expressed in the neighboring somatic cells as the embryo develops [49]. Here we see a similarity in the sea star where we see coexpression of *Hox11/13b* and *brachyury* with *Nanos*-positive cells (Figure S7D). In the mouse, expression of germline determinants with mesodermal markers is resolved by repression of the mesodermal markers in the PGC precursor cells; *Stella*-positive cells repress homeobox genes [49]. In the sea star, apart from restricting *Nanos* and *Vasa* transcript expression, repression of mesodermal markers in *Nanos*- and *Vasa*-positive cells remains an important step in germline specification which has yet to be identified.

In the mouse embryo, *Wnt3* and *Nodal* signaling serve as early regulators of PGC specification. These early signals give the epiblast cells competence to respond to BMP signaling, which regulates expression of germline determinants like *Blimp1* [3, 6]. This report of *Wnt* regulation of *Vasa* expression suggests that *Wnt* signaling may have an ancestral, conserved role in inductive germline formation. *Nodal* also has a role in

restricting the germline in both mouse and sea star. Further studies are necessary to identify which Wnt ligand is responsible for regulating germ cell development in the sea star. The early vegetal expression of Wnt3, Wnt16, Wnt8, and Frizzled1 receptor makes them potential candidates for testing their role in germ cell specification [25]. Following the inductive mechanism, PGC specification is a sequential process. This is evident in the sea star through sequential restriction of germline gene expression first to the dorsal side and later to the left side [7]. In the mouse, Wnt and Nodal signaling effects are required before a response to BMP signals is seen (Stuart et al.). An important future direction is the examination of BMP signaling in the sea star. BMP signaling has a conserved role in germline development by regulating expression of germline genes in mouse, cricket, axolotl (*Ambystoma mexicanum*) and the acorn worm (*Ptychodera flava*) [3, 4, 50, 51]. Preliminary data suggests that BMP may have a role in activating Vasa mRNA expression in the early gastrula (Figure S10). Further investigation of sea star embryos will uncover the sequence of signaling events required for proper development of the germ cells and whether these signaling events have been conserved through evolution.

Materials and Methods

Embryo culture and dissociation. Adult *Patiria miniata* animals were collected by either Peter Halmay (PeterHalmay@gmail.com) or Josh Ross (info@scbiomarine.com) off the Californian coast. Embryos were cultured essentially as described previously (Fresques et al., 2016). Embryos were cultured in filtered (0.2micron) 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 hyalin-extraction media (HEM) for 10-15 minutes, 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 40micron 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 (Oulhen et al., 2019).

In situ hybridization DIG-labeled RNA probes were made with a Roche DIG probe synthesis kit as described previously (Fresques et al., 2016). Probe-hybridized embryos were developed with NBT+BCIP for purple. Embryos were incubated with probe for 1 week and were developed essentially as described previously (Fresques et al., 2016).

Pharmacological perturbation Embryo cultures were incubated with 5 mM LiCl, 10uM

DAPT, 106ng/uL recombinant human activin AB, or recombinant human BMP4 at 50 or 100ng/mL, at the developmental time points indicated.

Morpholino injection Between 100 to 200 oocytes were injected with 1.0 mM of morpholino (at stock concentration, Gene-Tools) in injection solution: 10% glycerol and 1 mM Texas Red Dextran or 0.83 mM FITC Dextran (Molecular Probes). After incubation for 1 day at 16°C, injected fluorescent oocytes were selected and matured with 1-methyl adenine (3.0 μ M Acros Organics). Eggs were fertilized with diluted sperm, washed and incubated in filtered sea water at 16°C until they developed to the desired embryonic stage.

	Morpholino sequence	
Pm-FoxY1	5'-CCATCTCCCCGTCAGGATTCCAC-3'	
Pm-FoxY2	5'-GATATCATCTTGGCTCCACTTTAAA-3'	
Pm-FoxY3	5'-TGCGATTAGAATCAAAACGGAGTGA-3'	

Repressor construct injection Oocytes were injected with 1.0mM of the β -catenin/engrailed mRNA construct, cultured and collected at two-days post-fertilization for qPCR analysis.

qPCR analysis RNA was extracted from 50-100 embryos in each desired stage by using the RNeasy Micro kit (Qiagen). A reverse transcription reaction was performed using M-MLV Reverse Transcriptase (Promega) or M-MuLV Reverse Transcriptase (Thermoscientific, Maxima first-strand cDNA synthesis kit). Sybr green was used for qPCR analysis (ThermoFisher). Experiments were carried out in triplicate, and ubiquitin or 18 s was used to normalize RNA levels between samples. All data are represented as fold-change relative to the standards. A two-tailed t-test was used to calculate

significance. Error bars indicate ± 1 s.d.

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 ChromiumSingle Cell 3' Reagent kit v3 Chemistry. Libraries were sequenced by Genewiz on the Illumina Hiseq (2x150 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 10X Genomics. The custom transcriptome reference was generated from *P. miniata* assembly V2.0 (echinobase.org) using CellRanger mkref. The annotation for PmNanos was manually edited to account for a longer 3'UTR. 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 3.2.2 (Stuart and Butler et al., 2019). Cells of 8hpf to 14hpf stages with at least 1000 and at most 6000 genes (features), and cells with at least 1000 and at most 2500 genes in blastula to mid-gastrula stages 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 using the Seurat toolkit Harmony to identify conserved cell populations across the datasets. The clustering parameters used were dimensions: 20, resolution: 1.0. Cluster markers were found using FindConservedMarkers and FindMarkers functions. The reader is directed to the datasets deposited in GEO with the accession number GSE 196654. The code used is included with the supplementary

material, Pm Analysis.R.

References

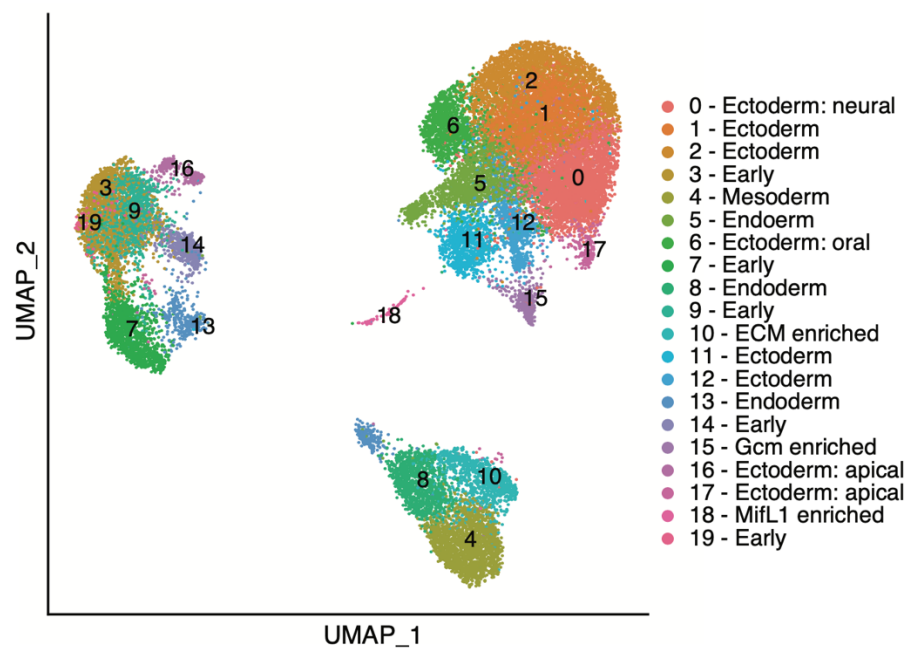
1. Extavour, C.G. and M. Akam, *Mechanisms of germ cell specification across the metazoans: epigenesis and preformation*. Development, 2003. **130**(24): p. 5869-84.
2. Illmensee, K. and A.P. Mahowald, *Transplantation of posterior polar plasm in Drosophila. Induction of germ cells at the anterior pole of the egg*. Proc Natl Acad Sci U S A, 1974. **71**(4): p. 1016-20.
3. Ohinata, Y., et al., *A signaling principle for the specification of the germ cell lineage in mice*. Cell, 2009. **137**(3): p. 571-84.
4. Nakamura, T. and C.G. Extavour, *The transcriptional repressor Blimp-1 acts downstream of BMP signaling to generate primordial germ cells in the cricket Gryllus bimaculatus*. Development, 2016. **143**(2): p. 255-63.
5. Bachvarova, R.F., et al., *Expression of Axwnt-8 and Axsxl in the urodele, axolotl: comparison with Xenopus*. Dev Genes Evol, 2001. **211**(10): p. 501-5.
6. Saitou, M. and M. Yamaji, *Germ cell specification in mice: signaling, transcription regulation, and epigenetic consequences*. Reproduction, 2010. **139**(6): p. 931-42.
7. Fresques, T.M. and G.M. Wessel, *Nodal induces sequential restriction of germ cell factors during primordial germ cell specification*. Development, 2018. **145**(2).
8. Fresques, T., et al., *Selective accumulation of germ-line associated gene products in early development of the sea star and distinct differences from germ-line development in the sea urchin*. Dev Dyn, 2014. **243**(4): p. 568-87.
9. Inoue, C., M. Kiyomoto, and H. Shirai, *Germ Cell Differentiation in Starfish: The Posterior Enterocoel as the Origin of Germ Cells in Asterina pectinifera*. Development, Growth & Differentiation, 1992. **34**(4): p. 413-418.
10. Wagner, D.E., et al., *Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo*. Science, 2018. **360**(6392): p. 981-987.
11. Briggs, J.A., et al., *The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution*. Science, 2018. **360**(6392).
12. Foster, S., N. Oulhen, and G. Wessel, *A single cell RNA sequencing resource for early sea urchin development*. Development, 2020. **147**(17).
13. Tan, K., H.W. Song, and M.F. Wilkinson, *Single-cell RNAseq analysis of testicular germ and somatic cell development during the perinatal period*. Development, 2020. **147**(3).
14. Wang, J.J., et al., *Single-cell transcriptome landscape of ovarian cells during primordial follicle assembly in mice*. PLoS Biol, 2020. **18**(12): p. e3001025.
15. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets*. Cell, 2015. **161**(5): p. 1202-1214.
16. Stuart, T., et al., *Comprehensive Integration of Single-Cell Data*. Cell, 2019. **177**(7): p. 1888-1902 e21.
17. Korsunsky, I., et al., *Fast, sensitive and accurate integration of single-cell data with Harmony*. Nat Methods, 2019. **16**(12): p. 1289-1296.
18. Yankura, K.A., et al., *Gene regulatory network for neurogenesis in a sea star embryo connects broad neural specification and localized patterning*. Proc Natl Acad Sci U S A, 2013. **110**(21): p. 8591-6.
19. Cheattle Jarvela, A.M., K.A. Yankura, and V.F. Hinman, *A gene regulatory*

- network for apical organ neurogenesis and its spatial control in sea star embryos.* Development, 2016. **143**(22): p. 4214-4223.
20. Cary, G.A., et al., *Systematic comparison of sea urchin and sea star developmental gene regulatory networks explains how novelty is incorporated in early development.* Nat Commun, 2020. **11**(1): p. 6235.
 21. Hinman, V.F. and E.H. Davidson, *Expression of a gene encoding a Gata transcription factor during embryogenesis of the starfish Asterina miniata.* Gene Expr Patterns, 2003. **3**(4): p. 419-22.
 22. Annunziata, R., P. Martinez, and M.I. Arnone, *Intact cluster and chordate-like expression of ParaHox genes in a sea star.* BMC Biol, 2013. **11**: p. 68.
 23. Calestani, C., J.P. Rast, and E.H. Davidson, *Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening.* Development, 2003. **130**(19): p. 4587-96.
 24. McCauley, B.S., E.P. Weideman, and V.F. Hinman, *A conserved gene regulatory network subcircuit drives different developmental fates in the vegetal pole of highly divergent echinoderm embryos.* Dev Biol, 2010. **340**(2): p. 200-8.
 25. McCauley, B.S., et al., *Expression of wnt and frizzled genes during early sea star development.* Gene Expr Patterns, 2013. **13**(8): p. 437-44.
 26. Fresques, T., et al., *The diversity of nanos expression in echinoderm embryos supports different mechanisms in germ cell specification.* Evol Dev, 2016. **18**(4): p. 267-78.
 27. Hinman, V.F. and E.H. Davidson, *Evolutionary plasticity of developmental gene regulatory network architecture.* Proc Natl Acad Sci U S A, 2007. **104**(49): p. 19404-9.
 28. Hinman, V.F., et al., *Developmental gene regulatory network architecture across 500 million years of echinoderm evolution.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13356-61.
 29. Hibino, T., et al., *The immune gene repertoire encoded in the purple sea urchin genome.* Dev Biol, 2006. **300**(1): p. 349-65.
 30. Nishihira, J., *Macrophage migration inhibitory factor (MIF): its essential role in the immune system and cell growth.* J Interferon Cytokine Res, 2000. **20**(9): p. 751-62.
 31. Perillo, M., et al., *Regulation of dynamic pigment cell states at single-cell resolution.* Elife, 2020. **9**.
 32. Sun, Z. and C.A. Ettensohn, *Signal-dependent regulation of the sea urchin skeletogenic gene regulatory network.* Gene Expr Patterns, 2014. **16**(2): p. 93-103.
 33. Rafiq, K., M.S. Cheers, and C.A. Ettensohn, *The genomic regulatory control of skeletal morphogenesis in the sea urchin.* Development, 2012. **139**(3): p. 579-90.
 34. Oulhen, N., et al., *Distinct transcriptional regulation of Nanos2 in the germ line and soma by the Wnt and delta/notch pathways.* Dev Biol, 2019. **452**(1): p. 34-42.
 35. Zhang, F., et al., *Inhibitory phosphorylation of glycogen synthase kinase-3 (GSK-3) in response to lithium. Evidence for autoregulation of GSK-3.* J Biol Chem, 2003. **278**(35): p. 33067-77.
 36. Montross, W.T., H. Ji, and P.D. McCrea, *A beta-catenin/engrailed chimera selectively suppresses Wnt signaling.* J Cell Sci, 2000. **113** (Pt 10): p. 1759-70.

37. Gustavson, M.D., et al., *Tcf binding sequence and position determines beta-catenin and Lef-1 responsiveness of MMP-7 promoters*. Mol Carcinog, 2004. **41**(3): p. 125-39.
38. Geling, A., et al., *A gamma-secretase inhibitor blocks Notch signaling in vivo and causes a severe neurogenic phenotype in zebrafish*. EMBO Rep, 2002. **3**(7): p. 688-94.
39. Meno, C., et al., *Mouse Lefty2 and zebrafish antivin are feedback inhibitors of nodal signaling during vertebrate gastrulation*. Mol Cell, 1999. **4**(3): p. 287-98.
40. Papanayotou, C. and J. Collignon, *Activin/Nodal signalling before implantation: setting the stage for embryo patterning*. Philosophical Transactions of the Royal Society B: Biological Sciences, 2014. **369**(1657): p. 20130539.
41. McCauley, B.S., et al., *Dose-dependent nuclear beta-catenin response segregates endomesoderm along the sea star primary axis*. Development, 2015. **142**(1): p. 207-17.
42. Hikabe, O., et al., *Reconstitution in vitro of the entire cycle of the mouse female germ line*. Nature, 2016. **539**(7628): p. 299-303.
43. Sasaki, K., et al., *Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells*. Cell Stem Cell, 2015. **17**(2): p. 178-94.
44. Juliano, C.E., et al., *Germ line determinants are not localized early in sea urchin development, but do accumulate in the small micromere lineage*. Dev Biol, 2006. **300**(1): p. 406-15.
45. Fujiwara, Y., et al., *Isolation of a DEAD-family protein gene that encodes a murine homolog of Drosophila vasa and its specific expression in germ cell lineage*. Proc Natl Acad Sci U S A, 1994. **91**(25): p. 12258-62.
46. Perillo, M., S.Z. Swartz, and G.M. Wessel, *A conserved node in the regulation of Vasa between an induced and an inherited program of primordial germ cell specification*. Dev Biol, 2022. **482**: p. 28-33.
47. Gildor, T., V. Hinman, and S. Ben-Tabou-De-Leon, *Regulatory heterochronies and loose temporal scaling between sea star and sea urchin regulatory circuits*. Int J Dev Biol, 2017. **61**(3-4-5): p. 347-356.
48. Yajima, M. and G.M. Wessel, *Small micromeres contribute to the germline in the sea urchin*. Development, 2011. **138**(2): p. 237-43.
49. Saitou, M., et al., *Specification of germ cell fate in mice*. Philos Trans R Soc Lond B Biol Sci, 2003. **358**(1436): p. 1363-70.
50. Johnson, A.D., et al., *Regulative germ cell specification in axolotl embryos: a primitive trait conserved in the mammalian lineage*. Philos Trans R Soc Lond B Biol Sci, 2003. **358**(1436): p. 1371-9.
51. Lin, C.Y., J.K. Yu, and Y.H. Su, *Evidence for BMP-mediated specification of primordial germ cells in an indirect-developing hemichordate*. Evol Dev, 2021. **23**(1): p. 28-45.

Figures

A.



B.

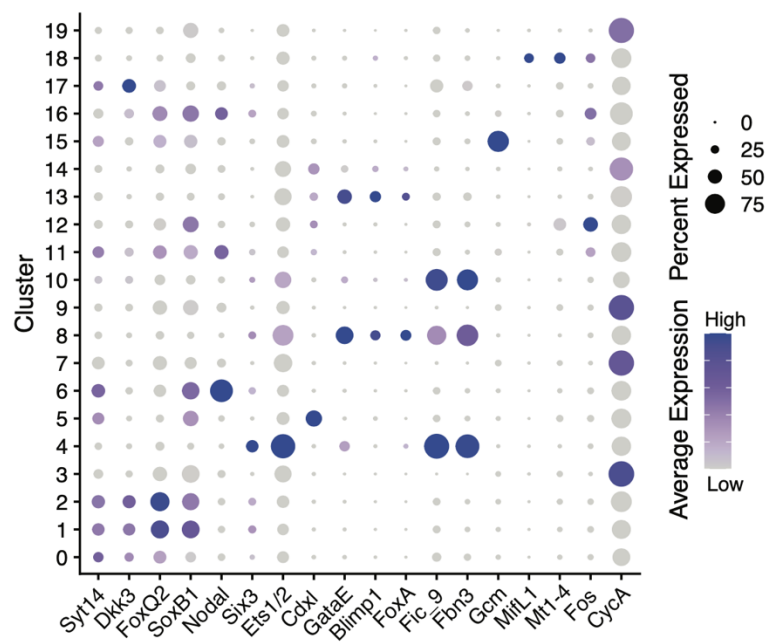


Figure 1. Cell states identified across early sea star development. (A) UMAP visualization of 25,703 cells after integration of six datasets. Cells colored by cell state identity based on marker gene expression. (B) Dotplot showing marker gene expression across clusters. Average gene expression level displayed by color intensity. Percent of cells expressing the marker gene conveyed by circle size.

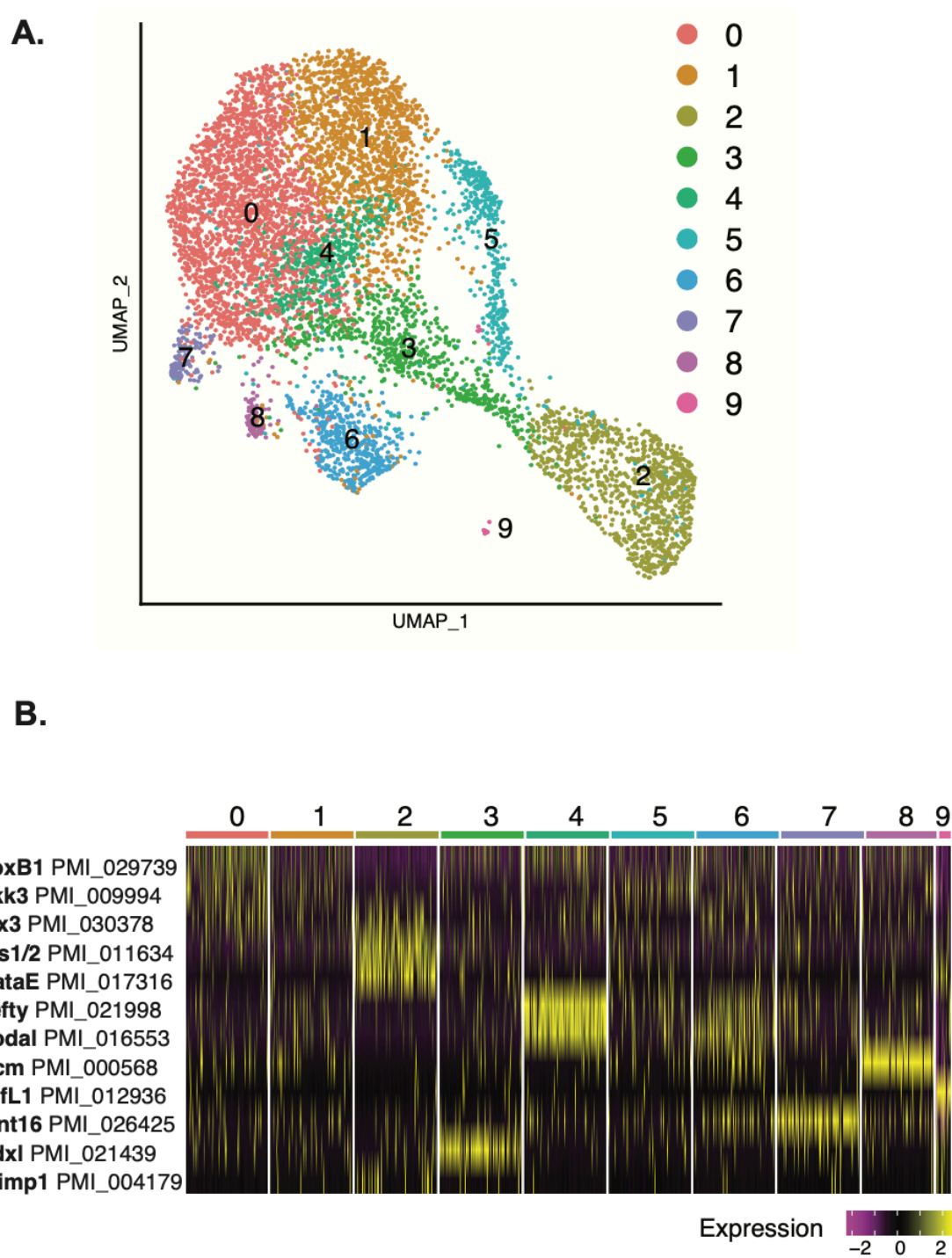


Figure 2. Gene expression at blastula stage. (A) UMAP plot of 7,272 blastula stage cells. Clusters colored by cell state identified by marker gene expression. (B) Heatmap of blastula stage marker gene expression.

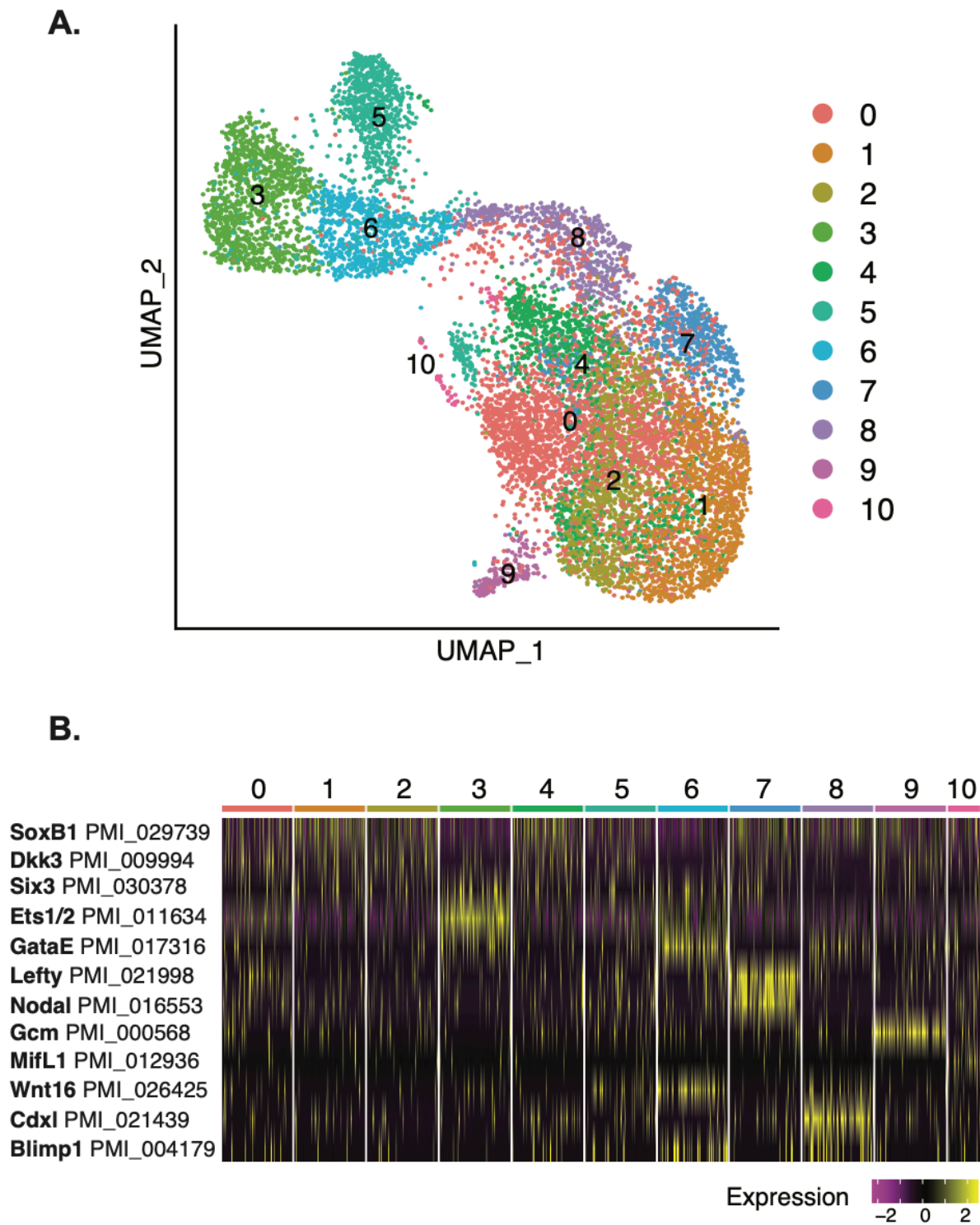


Figure 3. Gene expression at mid-gastrula stage. (A) UMAP plot of 10,448 mid-gastrula stage cells. Clusters colored by cell state identified by marker gene expression. (B) Heatmap of mid-gastrula stage marker gene expression.

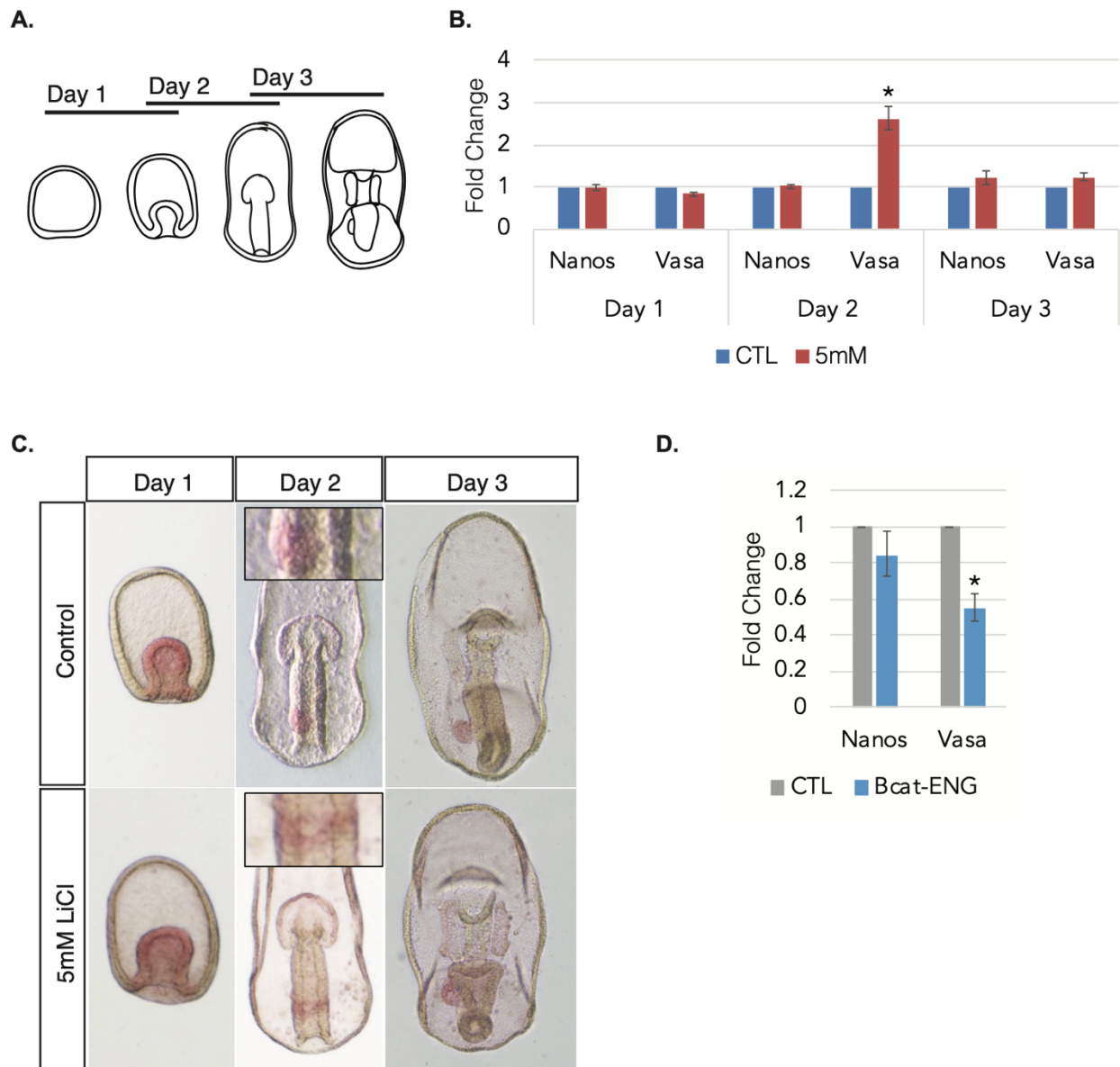


Figure 4. Wnt signaling regulates Vasa expression at gastrula stage. (A) Schematic showing experimental design. Embryos were treated with 5mM LiCl at one, two and three-days after fertilization for 24-hour incubation. (B) qPCR of Nanos and Vasa transcript abundance in control and LiCl treated embryos. Vasa transcript abundance is increased with treatment at day 2, $p=0.0017$. Nanos expression is unaffected. (C) Vasa transcripts (red) accumulate on the left side of the gut in control embryos. Embryos treated with LiCl show Vasa expression as a band across the gut or loss of left-side restriction. (D) Injection with the transcriptional repressor construct β -catenin/engrailed, results in a decrease of Vasa transcript abundance in gastrula stage embryos, $p=0.0148$. Asterisk (*) denotes p -value is at or below 0.05 threshold.

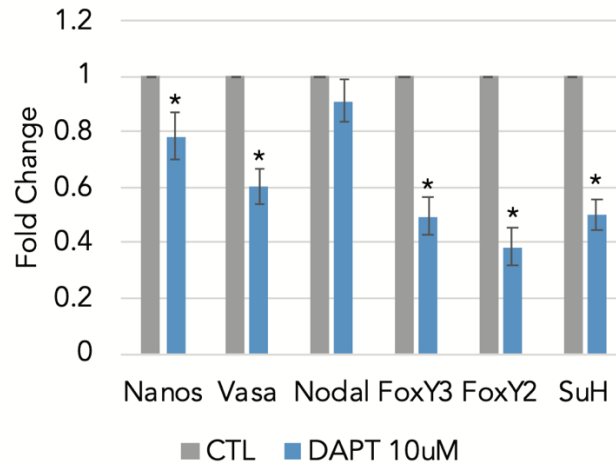


Figure 5. Delta/Notch regulation of germline genes. qPCR following incubation in 10uM DAPT at two-days post-fertilization. Nanos and Vasa expression decreases with DAPT treatment compared to controls. (Nanos $p=0.007$, Vasa $p=0.046$, FoxY3 $p=0.0069$, FoxY2 $p=0.0255$, SuH $p=0.050$). Asterisk (*) denotes p-value is at or below 0.05 threshold.

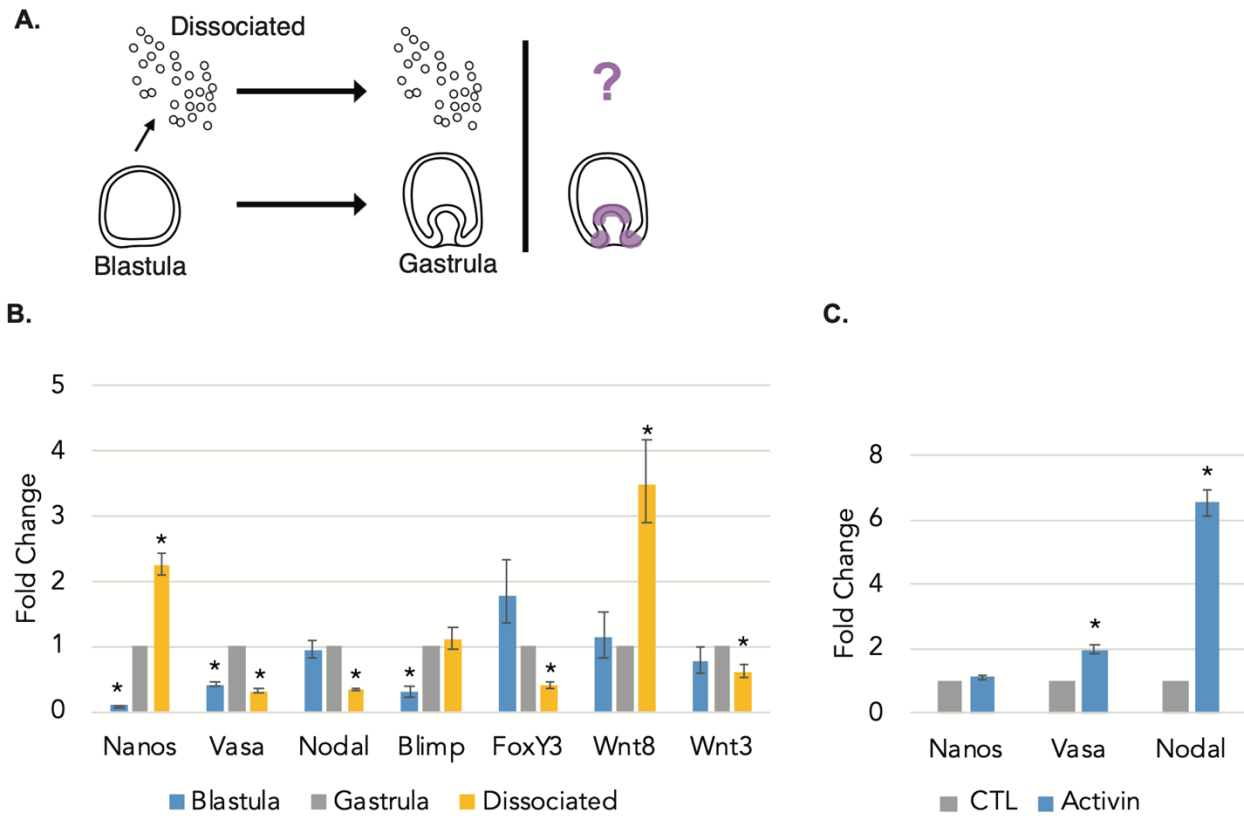


Figure 6. Analysis of cell interactions. (A) Schematic of experimental design. Embryos were cultured until hatched blastula stage at which point they were dissociated (D) and cultured for several hours until gastrula stage. (B) qPCR data showing change in mRNA transcript abundance with dissociation using intact gastrula stage as reference. Nanos expression is upregulated in dissociated cells compared to intact blastula and gastrula stage embryos. Vasa and Nodal expression in dissociated cells is downregulated compared to gastrula stage controls. (Blastula: Nanos $p=1.143E-08$, Vasa $p=0.003$, Blimp $p=0.0007$). (Dissociated: Nanos $p=1.164E-06$, Vasa $p=0.0018$, Nodal $p=4.90E-05$, FoxY3 $p=0.00098$, Wnt8 $p=0.0099$, Wnt3 $p=0.00069$). (C) Treatment of dissociated cells with rhActivinAB results in Nodal upregulation (Nodal $p=8.756E-05$, Vasa $p=0.0034$). Asterisk (*) denotes p-value is at or below 0.05 threshold. Nanos expression is unchanged compared to controls.

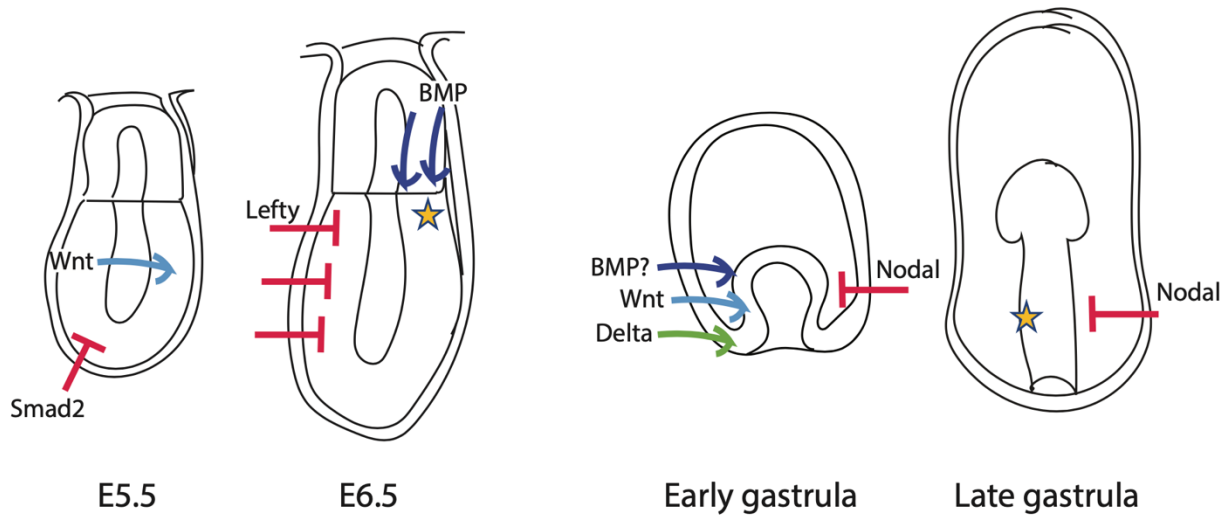
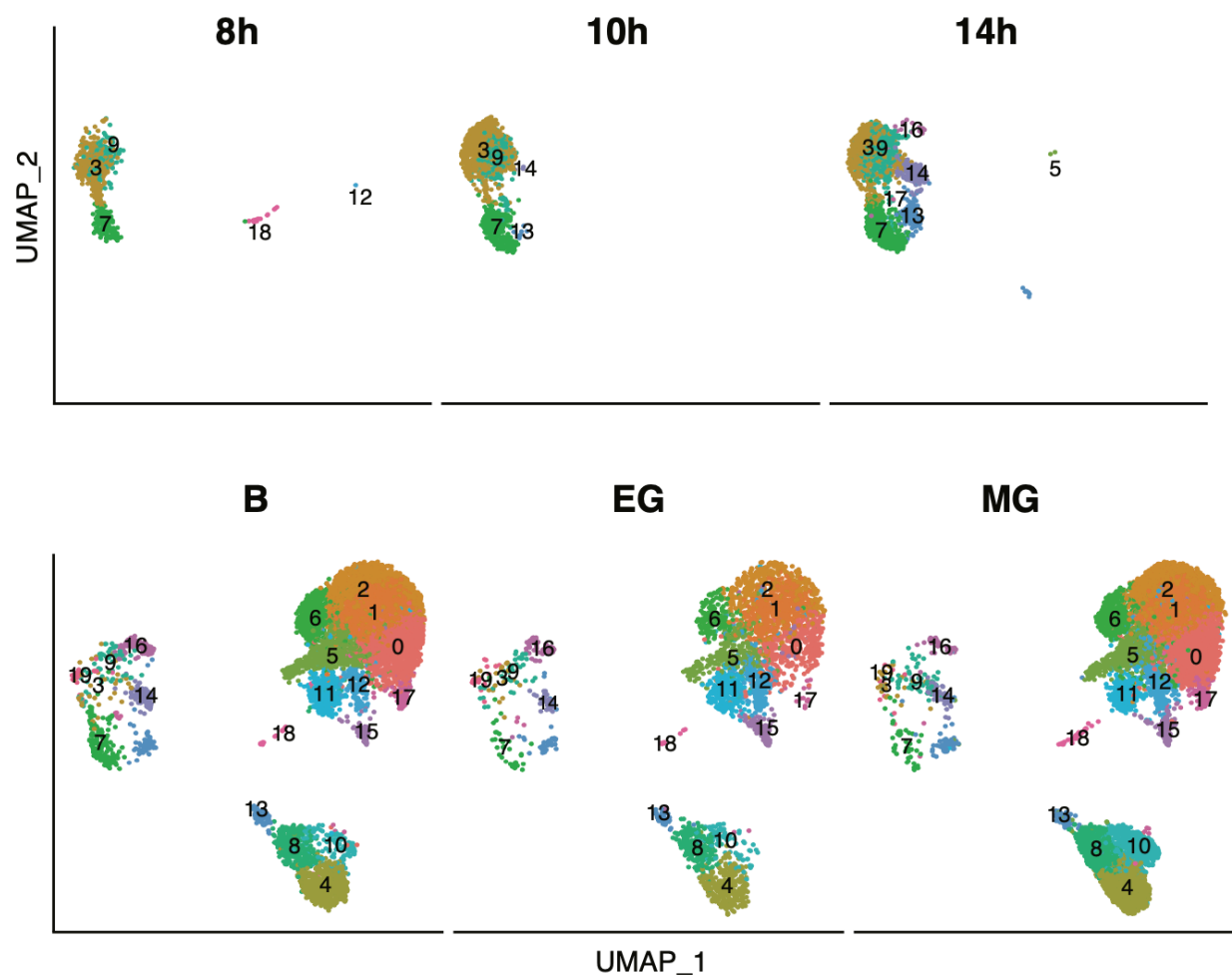
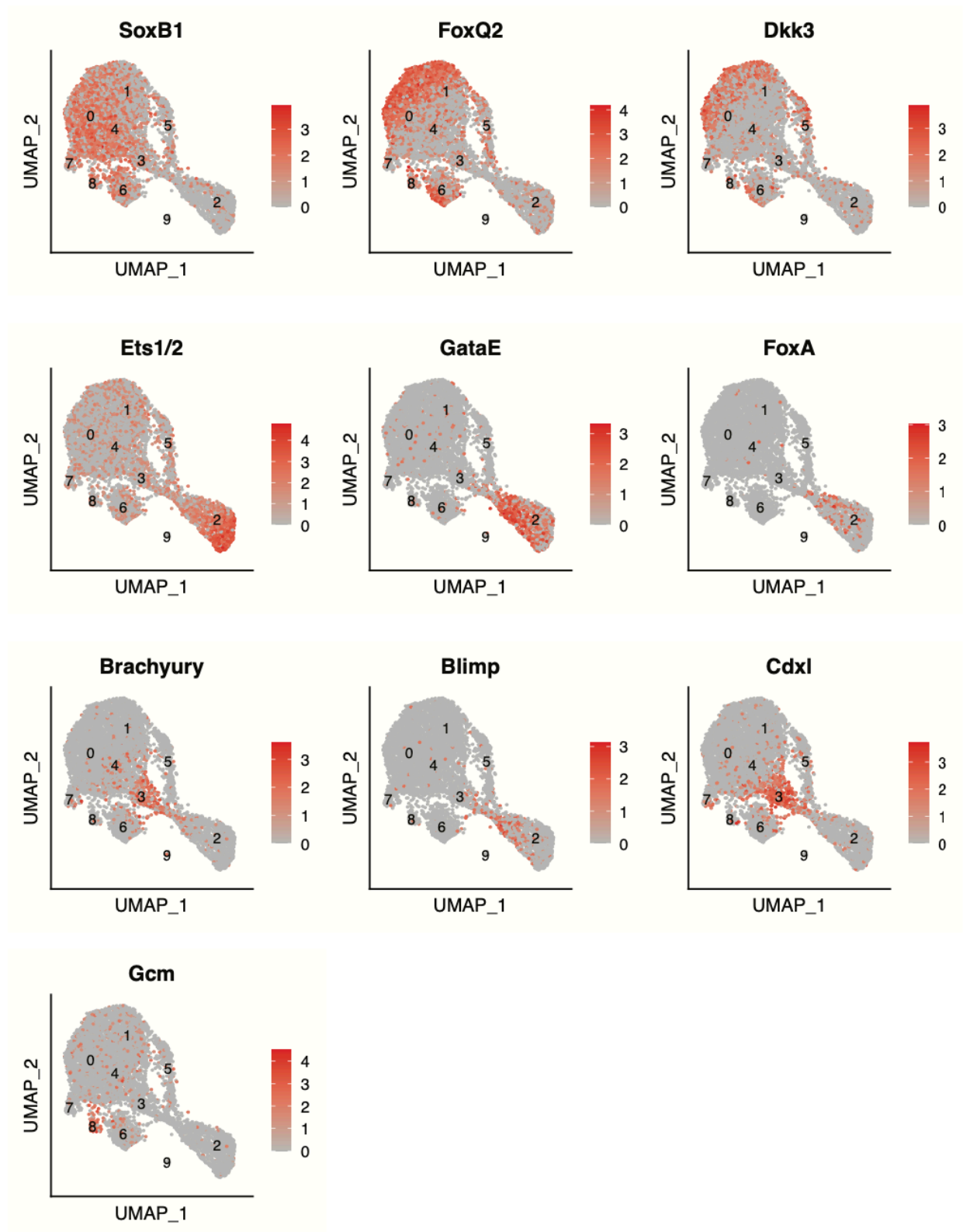


Figure 7. Inductive mechanism in mouse and sea star embryos. Model highlighting similarities and differences between inductive germ line development in mouse and sea star. Early inductive Wnt signaling conserved between mouse and sea star, with a role for Delta/Notch identified in the sea star. Nodal signaling has a conserved role in restricting germ line gene expression in both species. Preliminary data suggests BMP may have a conserved role in activating germ line gene expression.

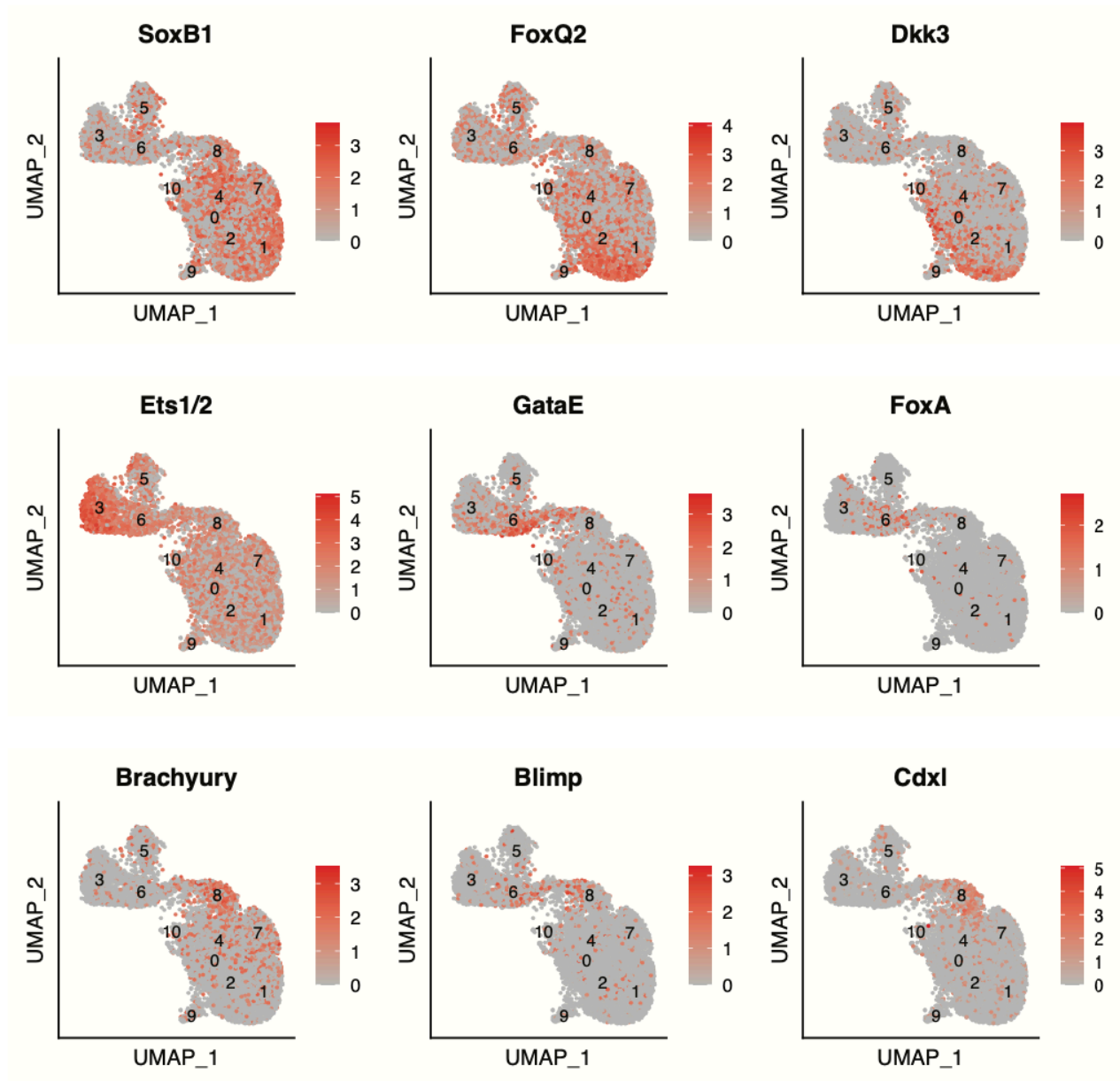
Supplementary Figures



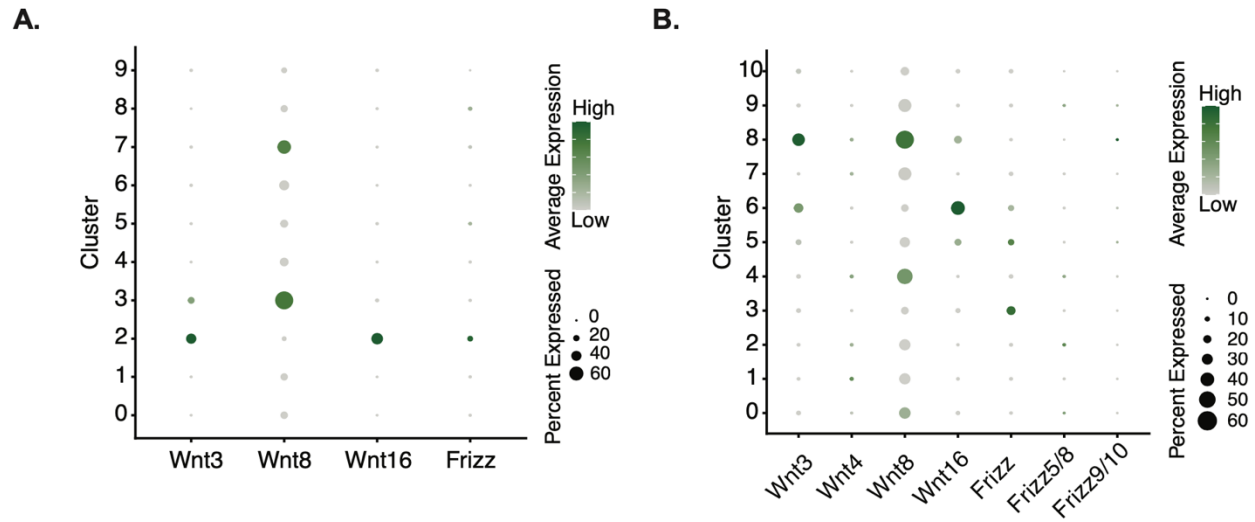
Supplementary Figure S1. Identification of cell states across early sea star development. UMAP visualization of six integrated datasets, separated by embryonic stage: 8-hpf (868 cells), 10-hpf (1,318 cells), 14-hpf (2,448 cells), blastula (7,272 cells), early gastrula (3,349 cells), and mid-gastrula (10,448 cells).



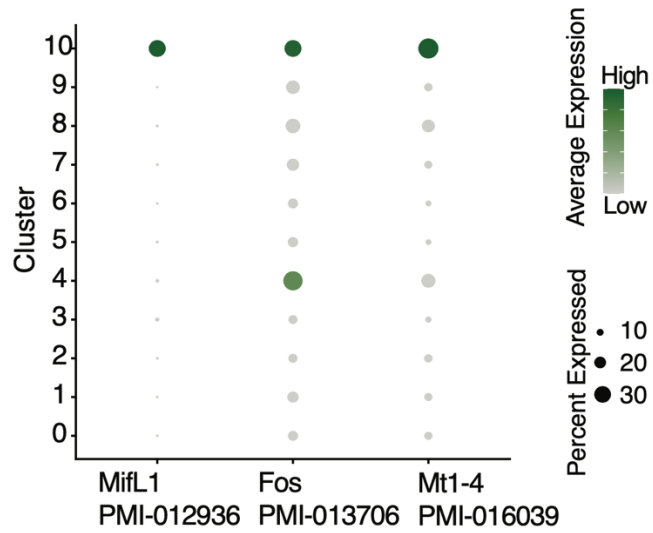
Supplementary Figure S2. Marker gene expression at blastula stage. Feature plots showing expression of ectodermal, mesodermal and endodermal marker genes at blastula stage. Average gene expression level displayed by color intensity.



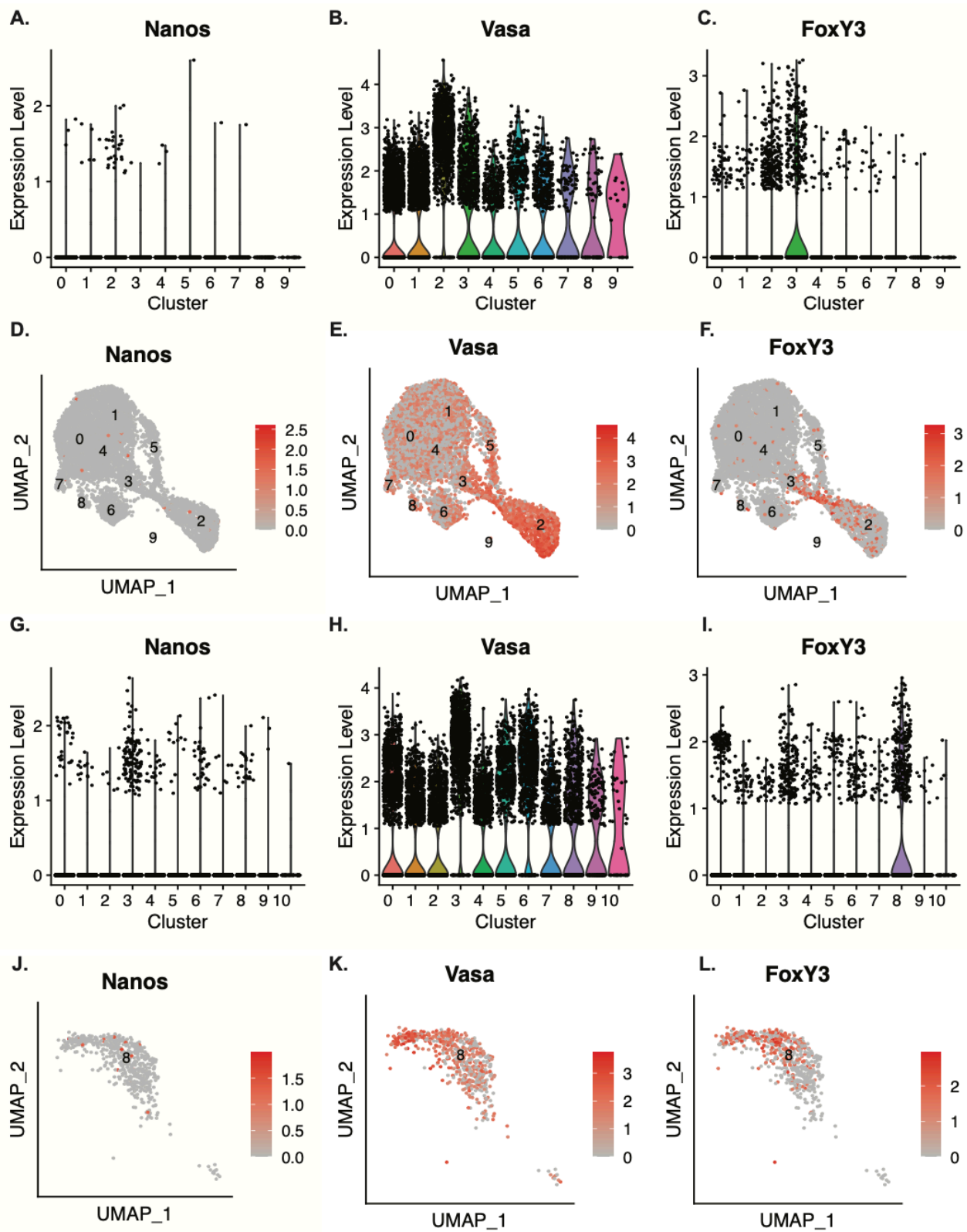
Supplementary Figure S3. Marker gene expression at mid-gastrula stage. Feature plots showing expression of ectodermal, mesodermal and endodermal marker genes at mid-gastrula stage. Average gene expression level displayed by color intensity.



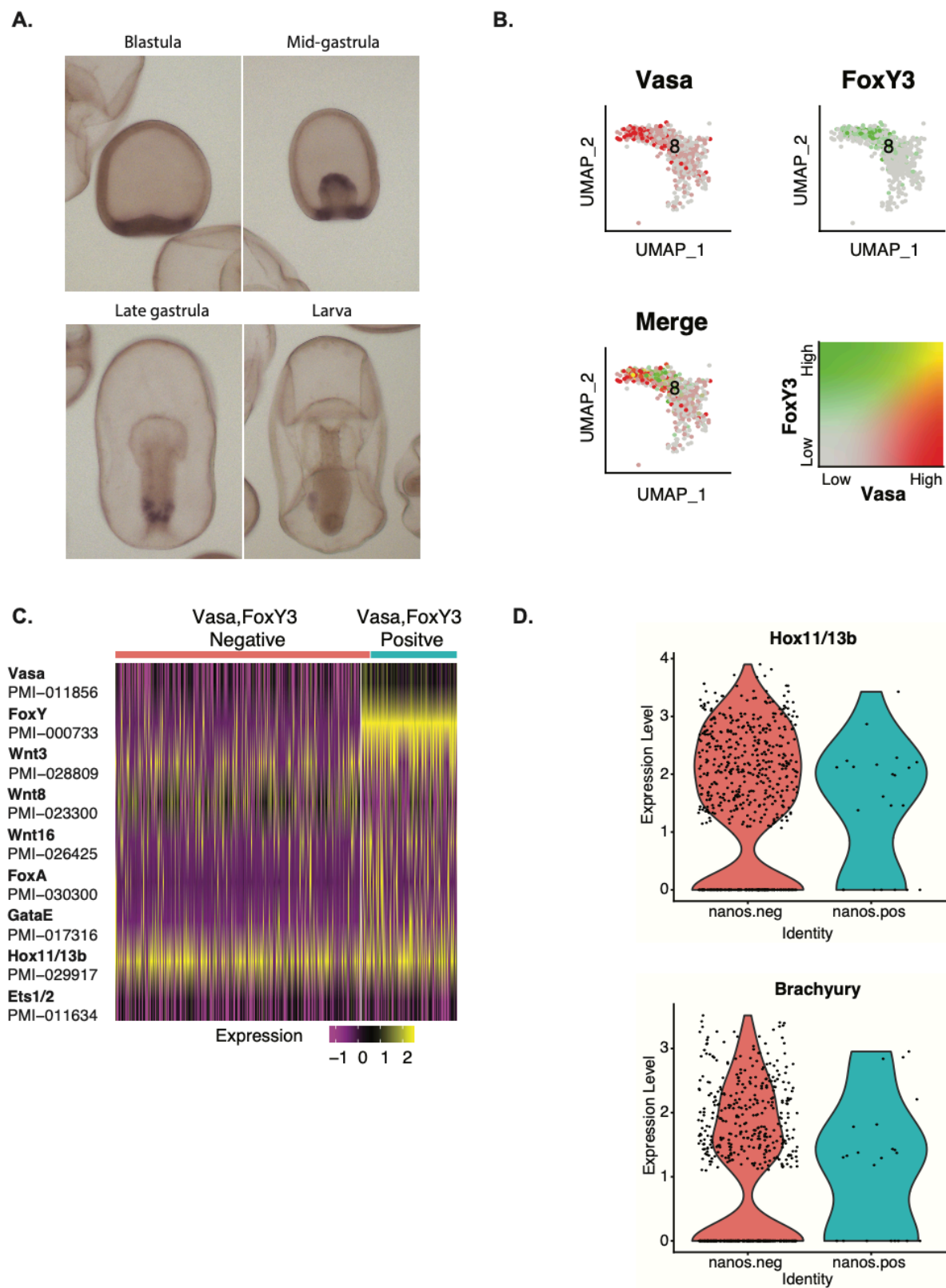
Supplementary Figure S4. Wnt signaling component expression. (A) Dotplot of Wnt signaling pathway component expression at blastula stage. Coexpression of Wnt3, Wnt16, and Frizzled receptor (Frizz) seen in cluster 2. (B) Dotplot of Wnt signaling pathway component expression at mid-gastrula stage.



Supplementary Figure S5. MifL1-enriched cluster expression. Dotplot showing expression of sea urchin PMC marker genes Mt1-4 and Fos in cluster 10, MifL1-enriched cluster, at mid-gastrula stage.

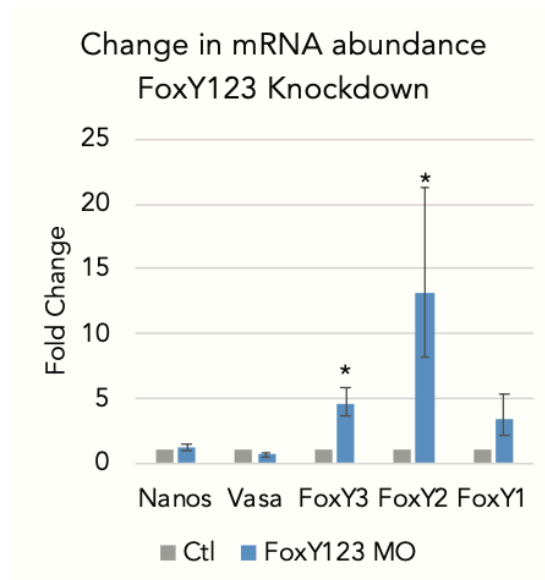
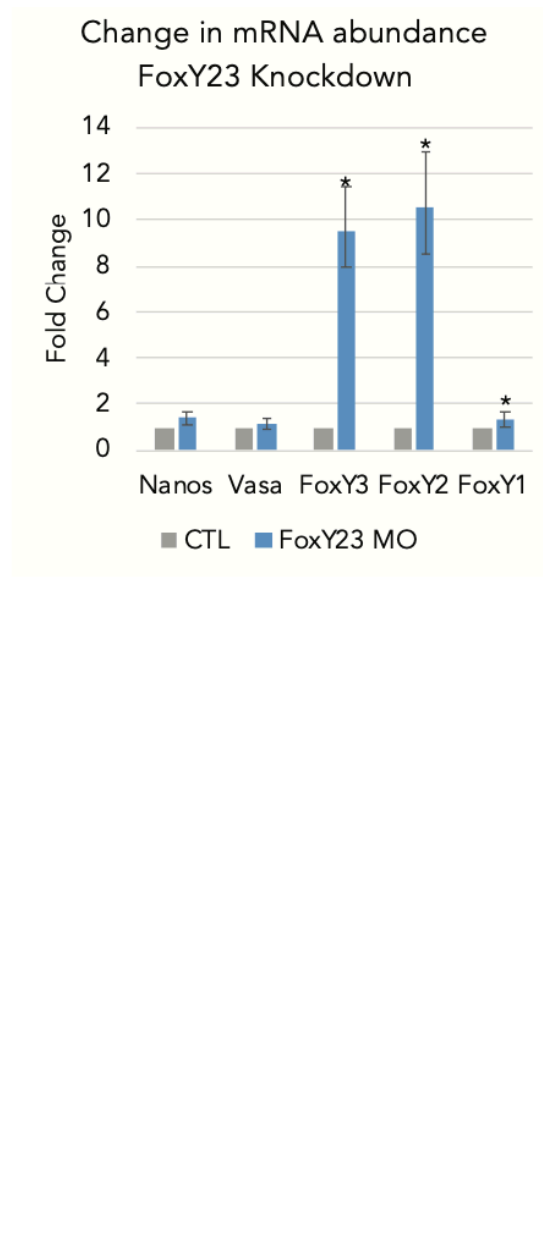
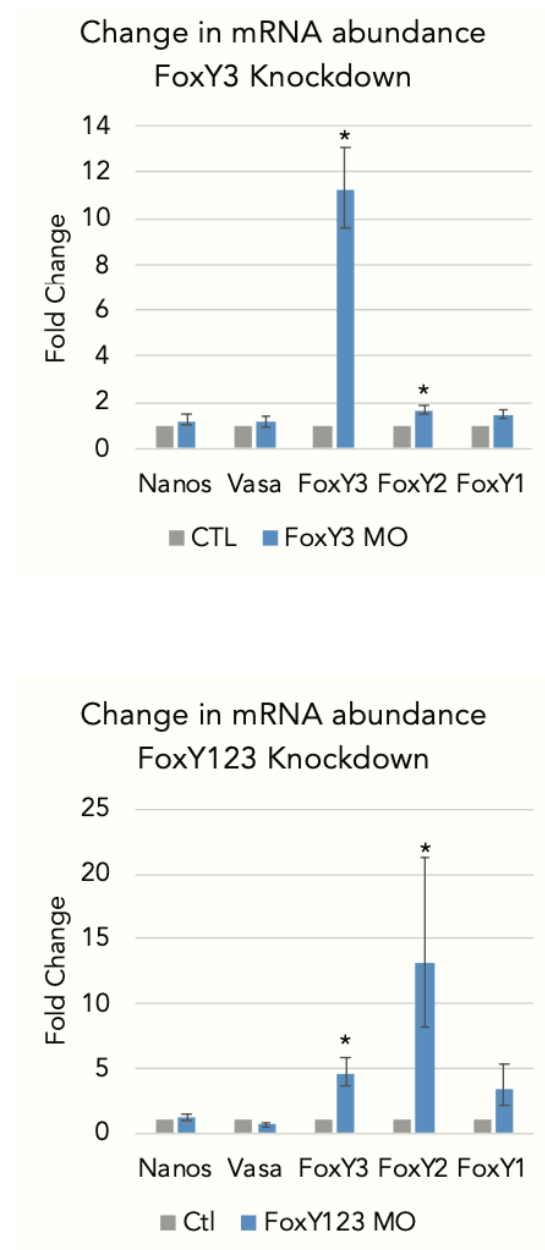


Supplementary Figure S6. Vasa, Nanos, and FoxY3 expression. Violin plots showing Nanos, Vasa, and FoxY3 expression per cluster in blastula stage (A-C) and mid-gastrula stage (G-I). Feature plots showing expression of Nanos, Vasa, and FoxY3 in blastula stage (D-F) and in cluster 8 of mid-gastrula stage (J-L).



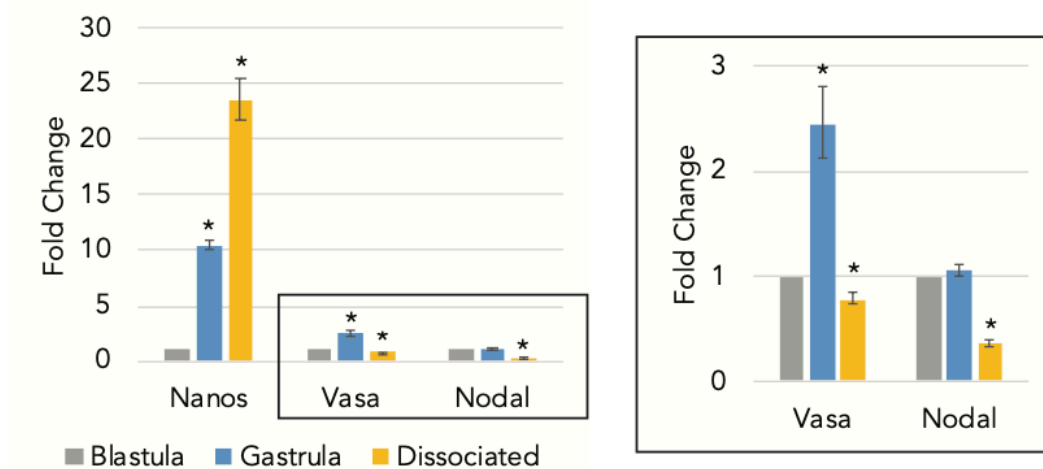
Supplementary Figure S7. Vasa and FoxY3 expression. (A) Expression of FoxY3 in the developing embryo shown by in situ hybridization. (B) Vasa expression in 395 cells of mid-gastrula cluster 8 (hindgut) shown in red (top left). FoxY3 expression in 221 cells

of cluster 8 shown in green (top right). Merge showing co-expression of Vasa and FoxY3 in 167 cells (bottom left). Color scale for expression level (bottom right). (C) Heatmap comparing expression in cells co-expressing Vasa and FoxY3 to those that do not co-express Vasa and FoxY3 in cells of cluster 8 of mid-gastrula stage (hindgut). (E) Violin plot of Hox11/13b and Brachyury expression in Nanos negative and Nanos positive cells of mid-gastrula cluster 8.

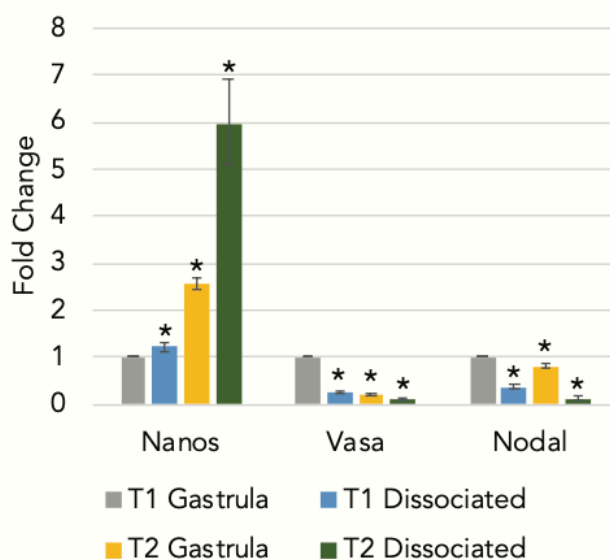


Supplementary Figure S8. FoxY morpholino knockdowns. qPCR data showing morpholino knockdown of FoxY1, FoxY2, or FoxY3 does not affect Nanos and Vasa mRNA abundance. (FoxY3 MO: FoxY3 $p=0.0019$, FoxY2 $p=0.0348$). (FoxY23 MO: FoxY3 $p=0.0038$, FoxY2 $p=0.00002$, FoxY1 $p=0.0272$). (FoxY123 MO: FoxY3 $p=0.0037$, FoxY2 $p=0.0186$). Asterisk (*) denotes p-value is at or below 0.05 threshold.

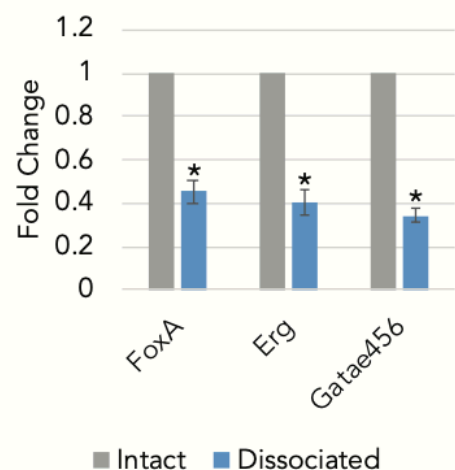
A.



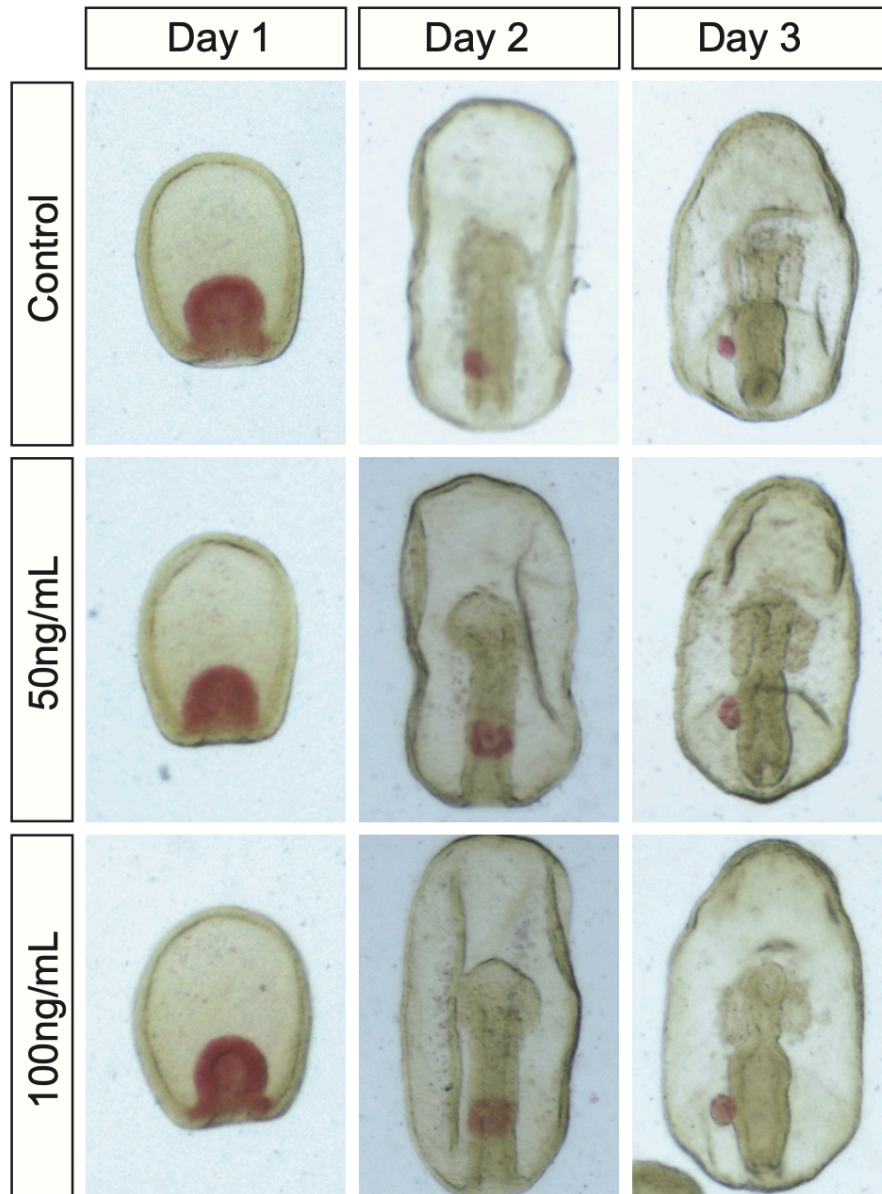
B.



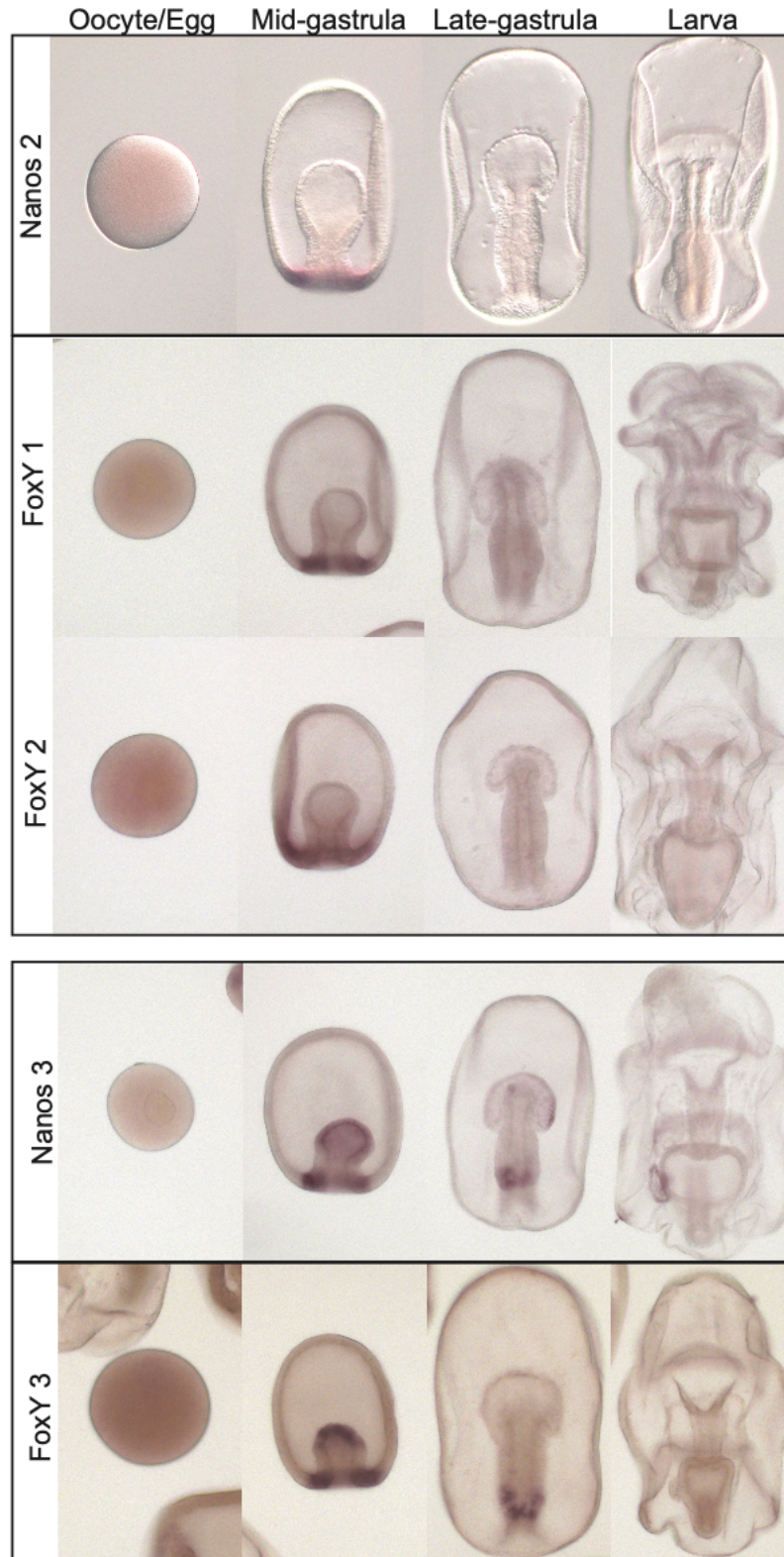
C.



Supplementary Figure S9. Changes in transcript abundance with dissociation. (A) qPCR data showing change in mRNA transcript abundance with dissociation using intact blastula embryos as control. (Right panel is zoomed in.) (Nanos Gastrula $p=1.143E-08$, Dissociated $p=1.281E-07$) (Vasa Gastrula $p=0.003$, Dissociated $p=0.0233$) (Nodal Dissociated $p=0.0012$). (B) Nanos transcript abundance increases with time in dissociated cells compartment to intact gastrula. Embryos and dissociated cells were cultured for 3 hours (T1) and 18 hours after blastula hatching (T2). (Nanos T1D $p=0.0039$, T2G $p=2.85E-05$, T2D $p=0.0005$) (Vasa T1D $p=7.73E-05$, T2G $p=5.55E-05$, T2D $p=3.44E-05$) (Nodal T1D $p=4.19E-05$, T2G $p=0.025$, T2D $p=9.15E-06$) (C) Transcript abundance of gut specific, Wnt target genes are decreased in dissociated cells compared to intact gastrula, (FoxA $p=0.023$), (Erg $p=0.0077$), (Gatae456 $p=0.0023$). Asterisk (*) denotes p-value is at or below 0.05 threshold.



Supplementary Figure S10. BMP4 treatment in early sea star embryos. In situ hybridization for Vasa mRNA expression in embryos treated with 50ng/mL or 100ng/mL recombinant human BMP4 at one, two and three days after fertilization for 24-hour incubation. Controls were incubated with 0.1% HCl BSA. At least 50 embryos from each developmental stage from three parental matings were tested.



Supplementary Figure S11. Nanos and FoxY gene expression. Whole mount in situ hybridizations for Nanos and FoxY transcription factors during early sea star embryogenesis. Nanos3 expression is seen in mid-gastrula at the tip of the archenteron

and the hindgut region. Nanos3 expression persists as a ring in the gut in late gastrula and becomes restricted to the PE; some expression is detected in the anterior pouches. Nanos2 is expressed transiently as a ring around the blastopore in mid-gastrula. FoxY1 and FoxY2 are also transiently expressed at mid-gastrula stage. FoxY3 is expressed in the same domain as Nanos3 at mid- and late gastrula stages. At least 50 embryos from each developmental stage from three parental matings were tested.

Chapter 5: Impact and Future Directions

Research summary

Echinoderms have been placed under the microscope for over a century with observations of sea urchin germ cells and fertilization dating back to 1847 (Briggs and Wessel, 2006; Derbès, 1847). Sea urchins and sea stars, both echinoderm species, have been subject to comparative studies including developmental heterochronies, neurogenesis, pigment cell development, and germ cell specification (Spurrell et al., 2022, submitted) (Gildor et al., 2017; Hinman and Burke, 2018; Wessel et al., 2014a). I too placed these species under a microscope and asked the “simple” question: What is a germline and how do I get one?

What is a germline?

I approached this question by considering the germ cells of both sea star and sea urchin using a single-cell RNA-sequencing (scRNAseq) technique. What is the transcriptional profile of the germ cells and how is it unique from that of somatic cells? Our work identified transcriptional profiles of cell states arising from the three germ layers in sea urchin and sea star, and cell type specific effects after either signaling perturbation or gene knock-outs in sea urchin embryos (Foster et al., 2022; Foster et al., 2020; Foster et al., 2019; Perillo et al., 2020). Analysis of the sea urchin primordial germ cells (PGCs) revealed co-expression of the germ cell genes, Nanos, Vasa, and Seawi, in two distinct cell states, the small micromeres, and the adjacent Veg2 cells. We also detected co-expression of the germ cell genes, Nanos and Vasa, in the sea star embryo. We find the transcription factor FoxY3 enriched in the Vasa-positive cells we identify as the PGC precursors (Foster et al., 2022). FoxY, the only FoxY transcription

factor in the sea urchin, was previously shown to regulate Nanos in the Veg2 cells (Oulhen et al., 2019).

In the sea urchin, the PGCs uniquely co-express the germ cell genes Nanos, Vasa, and Seawi. Additionally, the PGCs lack expression of somatic cell markers (Foster et al., 2020). Our analysis of sea star early embryos matches developmental time points captured for the sea urchin however, restriction of the germ cell genes happens much later in the sea star (Fresques et al., 2014). In sea star, the genes that will eventually become restricted to the PGCs are expressed more broadly and show co-expression with mesodermal markers in early stages. This pattern is also seen in the mouse, suggesting an ancestral, sequential process of germ cell specification (Saitou et al., 2003). The inclusion of these transcriptional landscape changes in our definition of “germline cell” will lead us to better understand their specification in the sea star and other species following the inductive mechanism.

How do I get a germline?

Nanos and Vasa are expressed first in the vegetal pole of the early sea star gastrula, then at the center of the archenteron, and later are restricted to the PE by an opposing Nodal signaling gradient (Fresques et al., 2014; Fresques and Wessel, 2018). With an inhibitory regulation identified, I decided to focus on what signaling may be responsible for activating Nanos and Vasa expression (Figure 1). We discovered a role for Wnt signaling in regulation of Vasa mRNA expression, that Delta/Notch signaling regulates mRNA abundance of both Nanos and Vasa, and we identified potential downstream effectors of this regulation. Using a cell dissociation approach, we identified

that Nanos and Vasa do not exhibit the same dependency on cell interactions (Foster et al., 2022). Our investigation of germ cell specification in the sea star led us to draw comparisons to germ cell specification in other species, both inherited and inductive. We found conservation of a Delta/Notch effect on Nanos expression, and co-expression of the forkhead box transcription factor, FoxY/FoxY3, with germ cell genes in two echinoderm species employing different germ cell specification methods.

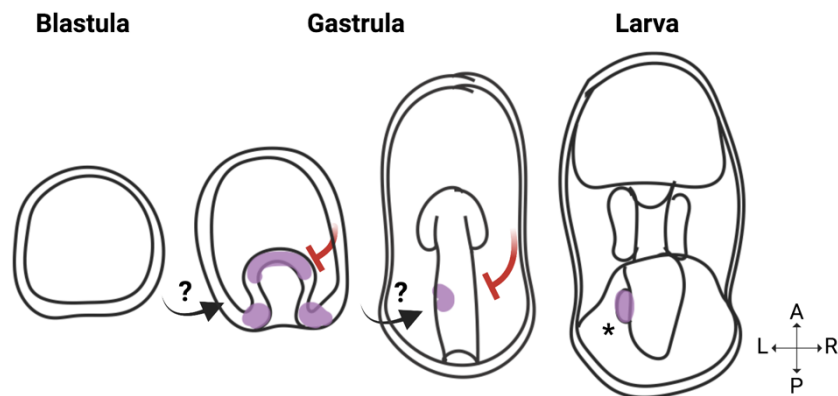


Figure 1. Embryonic context of germ cell specification project. Sequential restriction of Nanos and Vasa mRNA expression (in purple) by Nodal signaling in the ventral and right side of the embryo. Nanos and Vasa are expressed as a ring in the vegetal pole and at the tip of the archenteron in early gastrula. In late gastrula, expression is enriched at the center of the archenteron where it becomes restricted to the left side. Nodal signaling restricts Nanos and Vasa to the dorsal left side where the posterior enterocoel (PE) forms, shown with an asterisk (*). The focus of this research was to identify signaling mechanisms activating the expression of Nanos and Vasa. Anterior/Posterior and Left/Right axes shown in bottom right.

The implications of germ cell specification research in echinoderms are manifold. Studies suggest that the inductive mechanism is ancestral and the inherited mode arose independently several times throughout the evolution of derived species (Extavour and Akam, 2003). The sea star offers the opportunity to study this ancestral mechanism in a sister phylum to chordates. Sea star embryos are highly tractable and rapid development translates to an ability to make and test predictions quickly (Wessel et al.,

2014b). Comparative studies with the closely related sea urchin will grant insight into how germ cell gene regulation differs in an inherited mechanism. Additionally, discoveries in this space will bring further implications for human diseases including infertility and cancer (Persani et al., 2010; Zhang et al., 2020). For the purpose of this chapter, I will posit on the regulation of Nanos and Vasa expression, with an emphasis on future directions in the sea star.

Transcriptional Regulation of Nanos and Vasa

Activation of Nanos and Vasa expression

My research efforts focused on identifying cell signaling pathways that affect germ cell gene expression in the sea star. Our next steps include further characterizing the mechanism of this regulation. Canonical Wnt signaling seems to play a role in the regulation of germ cell gene expression in both sea star and sea urchin. Nanos expression in the sea urchin PGCs was recently found to depend on the canonical Wnt signaling pathway. β -catenin, the downstream signal transducer for Wnt, binds to the Nanos promoter region and directly regulates its transcription in the PGCs (Oulhen et al., 2019). Nanos expression in the sea star is first detected at mid-gastrula (Fresques et al., 2016). This drew forth the question of how Nanos expression is regulated in the sea star and some parallels in the early development of both species supports the hypothesis that it may be conserved. Disheveled, a key canonical Wnt signaling mediator is vegetally localized in the sea urchin oocyte (Gao and Chen, 2010; Peng and Wikramanayake, 2013; Weitzel et al., 2004). Disheveled function prevents β -catenin degradation, and accordingly, nuclear β -catenin is detected in the small micromeres

(Logan et al., 1999). Disheveled protein is also vegetally localized in the sea star oocyte, and knockdown of Disheveled via morpholino results in the downregulation of known Wnt target genes expressed posteriorly in the embryo (Swartz et al., 2021). We can test whether Disheveled disruption also affects Nanos expression in the sea star with the same morpholino approach followed by quantitative RT-PCR (qPCR) and in situ hybridization analysis for Nanos. If Disheveled knockdown results in decreased Nanos expression, we can test whether β -catenin mediates this relationship. Importantly, β -catenin CRISPR knockdown could reveal whether Nanos is a target of canonical Wnt signaling in the sea star either through the activity of maternally deposited Disheveled protein or signaling via Wnt ligands, which are vegetally localized at blastula and gastrula stages (McCauley et al., 2013). Mutation of putative binding sites is another approach for testing β -catenin regulation of Nanos transcription. There are five HMG-box binding motif sequences recognized by β -catenin and its cofactor Tcf/Lef within 8Kb of the Nanos transcription start site (5'-(T/A/G)ACAAAG-3'). It is possible that Disheveled localization initiates a cascade resulting in Nanos transcription. If Disheveled affects Nanos expression in the sea star, then this regulation found in the sea urchin is a conserved regulatory node. If Disheveled and β -catenin disruption does not affect Nanos in the star, then this points to a derived trait of the urchin and suggests an evolutionary change in the Nanos promoter region.

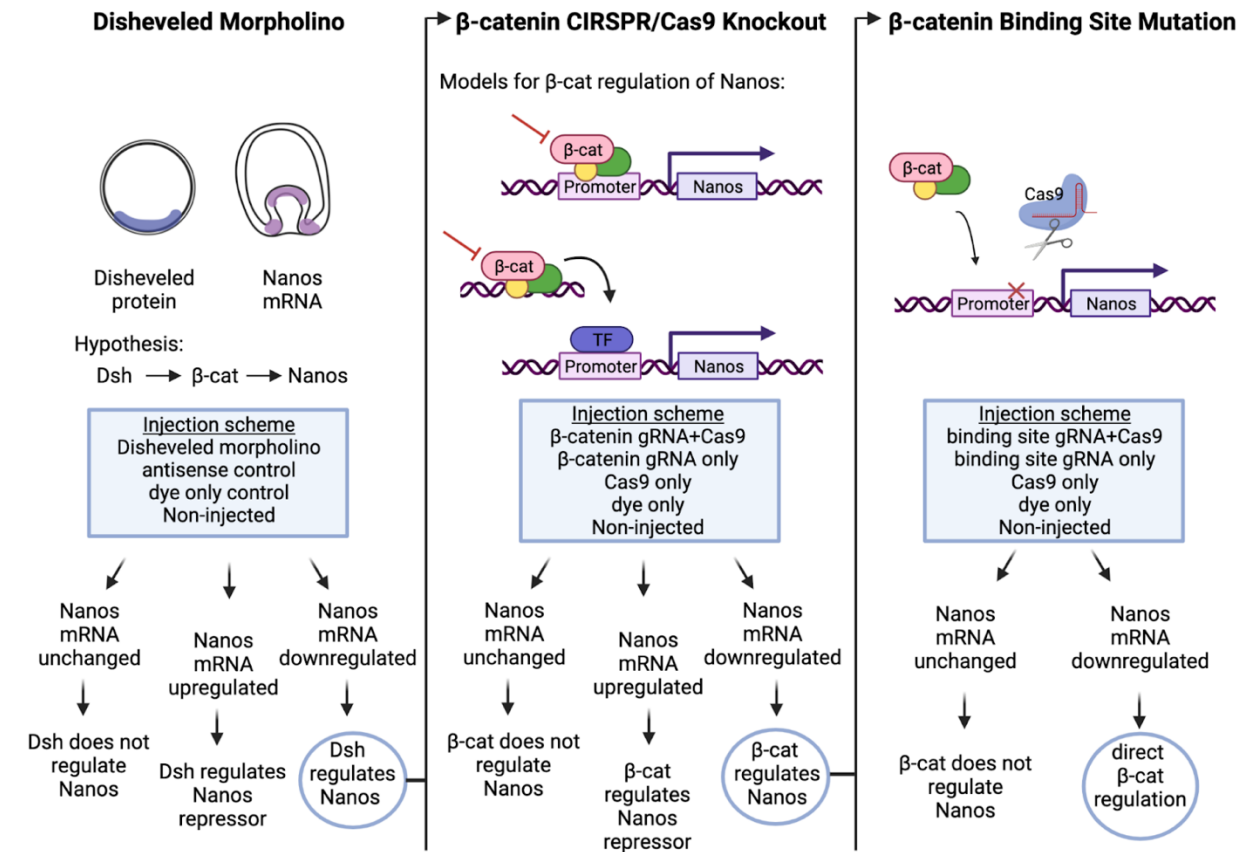


Figure 2. Schematic representation of experimental approach to testing Wnt regulation of Nanos transcription. (Left) Disheveled morpholino approach to test regulatory relationship between maternally deposited Disheveled protein and Nanos transcript expression. (Center) CRISPR/Cas9 mediated β -catenin knockout to test β -catenin regulation of Nanos expression. Whether β -catenin regulation is a direct or indirect relationship remains unknown. (Right) Test for direct regulation of Nanos by β -catenin by mutation of putative β -catenin binding sites in the Nanos promoter region. Figure created with BioRender.com.

Wnt signaling activation was found to regulate Vasa mRNA expression in the sea star gastrula, potentially through β -catenin (Foster et al., 2022). The Wnt receptor Frizzled1 (Fz1) and ligands Wnt18, Wnt8, and Wnt3 are likely candidates for regulating Vasa transcription based on their expression patterns (McCauley et al., 2013). Moreover, scRNAseq showed Wnt3 transcripts present in Vasa, FoxY3-positive cells, as well as in neighboring cells, cells of the same cluster identity that are not Vasa,

FoxY3-positive, along with Wnt8 (Foster et al., 2022). Wnt signals could be coming from the Vasa-positive cells themselves, via Wnt3, or from the neighboring cells expressing Wnt8 and Wnt3. I would begin teasing apart the mechanism of Wnt regulation of Vasa expression by knocking down the Fz1 receptor which is expressed vegetally, overlapping the region of Vasa mRNA expression (Fresques and Wessel, 2018; McCauley et al., 2013). This approach avoids confounding results from potential compensatory signaling through different Wnt ligands if I were to begin with knocking down either Wnt3 or Wnt8. As an alternative to Wnt ligand knockdowns, I could inject message for either Wnt3 or Wnt8 and test whether overexpression of the ligands results in upregulated expression of Vasa in the embryo.

Vasa expression was found to be inducible in somatic cells in the sea urchin. Altering β -catenin levels by LiCl treatment, which promotes β -catenin nuclearization, or injection of a dominant-negative cadherin, which prevents β -catenin nuclearization, showed that Vasa protein accumulation correlated with β -catenin levels, without affecting Vasa expression within the small micromeres (Voronina et al., 2008). Further studies will reveal whether higher Vasa protein levels with increased β -catenin is a result of increased Vasa transcription, translation, or decreased protein turnover. Vasa protein upregulation is also seen throughout the embryo with micromere deletion (Yajima and Wessel, 2011). The mechanism of this accumulation is unknown. How loss of the micromeres is detected by the embryo is an interesting question. Do the micromeres and small micromeres signal to their neighboring cells? Does the loss of this signal initiate new Vasa transcription? Voronina et al. reported no detectable difference in Vasa mRNA accumulation in micromere-depleted embryos compared to

controls. Our group previously hypothesized that micromere deletion represses post-translational Vasa degradation. It would be interesting to investigate how canonical Wnt signaling through β -catenin results in increased Vasa protein accumulation.

We may reach several conclusions through investigating Wnt regulation of Nanos and Vasa in sea star embryos. First, we may find canonical Wnt signaling regulation of Nanos to be conserved in both sea star and sea urchin. We may find that both Nanos and Vasa are regulated by Wnt signaling in the sea star. Wnt signaling activation via LiCl treatment in gastrula did not affect Nanos mRNA abundance (Foster et al., 2022). However, the presence of putative β -catenin binding sites in the Nanos promoter suggests that Wnt signaling could regulate Nanos transcription at another time in development. Finally, Wnt signaling may also play a role in regulating Vasa expression in both species although through different mechanisms, regulating transcription in the sea star and translation in the sea urchin.

While Nanos is enriched in the small micromeres of the sea urchin, Nanos expression expands transiently to the Veg2 (non-skeletogenic mesoderm) cells, adjacent to the small micromeres (Foster et al., 2020; Oulhen et al., 2019). Nanos expression in the Veg2 cells is regulated by the forkhead box transcription factor, FoxY (Oulhen et al., 2019; Song and Wessel, 2012). FoxY expression is first seen in the small micromeres, overlapping with Nanos and Vasa, before expanding to the Veg2 mesodermal cells by in situ hybridization and we see this by scRNAseq as well (Foster et al., 2020; Materna et al., 2013). FoxY is the first transcription factor expressed in cells that give rise to the coelomic pouch and is required for proper coelomic pouch formation in the sea urchin (Materna et al., 2013). The sea star has three FoxY transcription

factors, and we recently reported co-expression of FoxY3 with Nanos and Vasa in blastula and gastrula stages and eventually in the PE (Foster et al., 2022). Knockdown of the FoxYs in sea star does not affect Nanos or Vasa mRNA abundance suggesting that FoxY's role as a regulator of Nanos is not present in the sea star. However, FoxY3's colocalization with the germ cell factors in the PE makes it a target of interest. One possibility to test is that FoxY3 inhibits mesodermal marker expression within the Vasa-positive cells of the hindgut, the presumptive PGC precursors. I would return to the FoxY3 knockdown experiment to test its effect on mesodermal genes. Since FoxY3 does not regulate expression of the germ cell genes Nanos and Vasa, will FoxY3 knockdown or knockout affect formation of the PE and eventually of the PGCs? Another interesting question is whether enrichment of FoxY3 in the PE is also under Nodal signaling control.

Nodal restriction of Nanos and Vasa

The germ cell determinants Nanos and Vasa are expressed broadly in early sea star embryogenesis and Nodal is essential for their restriction to the PE (Fresques and Wessel, 2018). A Nodal morpholino experiment revealed that Nodal is a double block on Nanos and Vasa expression— Nodal inhibits new transcription of Nanos and Vasa and promotes the degradation of existing transcripts (Fresques and Wessel, 2018). The morpholino result does not address whether Nodal's inhibitory effect is direct or indirect, whether Nodal signals directly to the cells expressing the germ cell genes. Treating dissociated cells with Activin, a Nodal signaling agonist, results in upregulation of Nodal mRNA abundance and loss of the Nanos upregulation seen in dissociated controls

(Foster et al., 2022). We interpret this result to mean that the Nanos-positive cells must express the Nodal receptor and therefore Nodal regulation of Nanos and Vasa is direct.

There is an important distinction between how Nodal and Activin bind their TGF β receptors and initiate Smad-dependent signaling. Nodal requires a coreceptor of the EGF-CFC family, to activate the receptor complex, while Activin does not (Shen, 2007; Yeo and Whitman, 2001). The known EGF-CFC factors include Cripto and Cryptic in both mouse and humans (Dono et al., 1993; Yeo and Whitman, 2001). Sea star has a "cripto/FRL-1/cryptic family-like" gene which is seen in some Vasa-positive cells by scRNAseq (Foster, unpublished). We can test for expression of the Nodal receptor and coreceptor in cells expressing the germ cell genes by immunofluorescence in order to conclude with greater confidence that the Nanos and Vasa positive cells express the receptor and respond to Nodal signals.

The transcriptional regulators and RNA decay regulators for Nanos and Vasa downstream of Nodal are still unknown. To further investigate how Nodal inhibits germ cell factor gene expression, we must consider both transcriptional repression and transcript turnover. How does Nodal inhibit new transcription of Nanos and Vasa? Nodal signaling restricts FoxY expression in the sea urchin; the Nodal target gene Pitx2 binds to the FoxY promoter region repressing its transcription (Materna et al., 2013). In the sea star, Pitx2 is expressed at low level in the vegetal region of the early gastrula. At late gastrula stage, Pitx2 is enriched in the right side ectoderm, similar to Nodal, and in the right coelomic pouch (Fresques et al., 2014). We detect co-expression of Pitx2 and Vasa in gastrulae via scRNAseq (Foster, unpublished). Moreover, the nucleotide sequence bound by Pitx2 in the FoxY promoter is present in the Nanos and Vasa

promoter regions (Foster, unpublished) (Materna et al., 2013). For these reasons, the Nodal target Pitx2 is a candidate repressor of Nanos and Vasa transcription in the sea star. We can test whether knockdown of Pitx2 in the sea star results in increased mRNA abundance of Nanos, Vasa, or FoxY3. Also, we can begin to investigate how Nodal signaling promotes turnover of Nanos and Vasa transcripts by looking for changes in RNA degradation proteins including decapping enzymes, endonucleases, exonucleases, and deadenylases, in Nodal activation and inhibition conditions. The CCR4-related deadenylase CNOT6, being a Nanos target gene, is degraded in the small micromeres and this has a protective effect on Vasa and Seawi accumulation in the sea urchin (Swartz et al., 2014). CNOT6 is expressed uniformly in the sea star embryo with the exception of the vegetal-most cells and is later exempt from the PE as well (Fresques et al., 2014). CNOT6 expression does not overlap with expression of the germ cell genes in the sea star, which would be consistent with a role for CNOT6 as a negative regulator of any of those genes. How CNOT6 expression is regulated and whether Nodal signaling affects CNOT6 activity are both interesting directions to pursue.

As we examine the mechanism of action for transcriptional activators and inhibitors of Nanos and Vasa expression we will reveal how their regulation is conserved between two closely related species following distinct methods of germ cell specification. Additionally, we can extend our analysis to the germ cell gene Piwi in sea star and determine how three germ cell genes that initially show different expression patterns become restricted to the same site, the PE (Fresques et al., 2014; Fresques and Wessel, 2018).

Post-Transcriptional Regulation of Nanos and Vasa

My research focused primarily on the expression and regulation of germ cell transcripts in sea star. However, their selective expression in the PGCs does not solely rely on transcriptional control. The function of Nanos as a translational repressor makes its restriction necessary, as ectopic expression can have detrimental effects on the embryo (Lai and King, 2013). In the case of *Drosophila*, Nanos mRNA is present in the mature oocyte, and restriction of Nanos mRNA in the posterior pole of the early embryo is conferred through translational regulation. An element in the Nanos 3'UTR leads to translation within the pole plasm and degradation of the mRNA outside of the pole plasm (Dahanukar and Wharton, 1996). Sea urchin (*S. purpuratus*) Nanos mRNA translation is regulated by elements in both the 5' and 3'UTR regions. Removal of the 5'UTR results in decreased protein synthesis (Oulhen et al., 2013). Importantly, the Nanos 3'UTR is responsible for RNA localization and translational regulation, specifically the Global Nanos Associated RNA Lability Element (GNARLE) element within the 3'UTR. Nanos mRNA is degraded outside of the small micromeres; this regulation is lost when the 388 nucleotide region is removed from the 3'UTR (Oulhen and Wessel, 2014; Oulhen et al., 2013). The GNARLE element is highly conserved between *S. purpuratus* and *H. pulcherrimus* which share a more closely related common ancestor than *S. purpuratus* and *L. variegatus*, which does not have a GNARLE element (Oulhen and Wessel, 2014; Smith et al., 2006). Nanos expression in *L. variegatus* follows a similar pattern to that of *S. purpuratus* by in situ hybridization, localized expression in the vegetal region, likely the small micromeres, expression at

the tip of the archenteron, and restriction to the left coelomic pouch (Fresques et al., 2016). scRNAseq reveals Nanos expression is less specific in *L. variegatus* compared to *S. purpuratus*, with Nanos expression detected in somatic cells of the three germ layers (Morita et al., unpublished). Future studies will reveal the mechanism of Nanos regulation in *L. variegatus* and if differences in the Nanos expression pattern are due to the loss of the conserved GNARLE element. The 3'UTR of sea star (*P. miniata*) Nanos does not have a conserved GNARLE element (Oulhen et al., 2022). In addition, exogenous Nanos-GFP expression was seen ubiquitously in the embryo, suggesting that Nanos accumulation in the PE is regulated transcriptionally (Perillo et al., 2022). This result gives greater emphasis to Nodal's role in restricting Nanos mRNA expression to the putative PGC precursors in the PE (Fresques and Wessel, 2018). Removing the GNARLE element in the sea urchin results in widespread Nanos transcript expression (Oulhen and Wessel, 2014; Oulhen et al., 2013). The sea star displays the opposite scenario, there is no GNARLE element in the Nanos 3'UTR, and the transcript is not ubiquitously expressed. This suggests that the expression of Nanos in specific cells, although initially broad compared to its eventual restriction, is regulated transcriptionally.

Vasa transcripts are present throughout early cleavage and blastula stages in the sea urchin before becoming enriched in the small micromeres at gastrula stage (Juliano et al., 2006). Gustavus (Gus), an E3-ubiquitin ligase, regulates Vasa post-translationally in the sea urchin; Vasa is turned over in somatic cells by Gus function and as a result, Vasa protein is enriched and maintained in the PGCs (Gustafson et al., 2011). In the sea star, Vasa is present in the oocyte and broadly in the early embryo before it is

restricted to the region of PE formation (Fresques et al., 2014; Fresques and Wessel, 2018). Gus was found to regulate Vasa protein turnover outside of the PE; lack of Gus expression in the PE is protective of Vasa protein in the PE and thus the putative PGC precursors (Perillo et al., 2022). Residues within the Vasa open reading frame (ORF) required for proper Vasa enrichment in the small micromeres have been identified, and it was found that Gus binds directly to Vasa (Gustafson et al., 2011). Future studies will identify whether sequences required for Vasa turnover in the sea urchin are also present in the sea star.

There are many steps from the initial expression of a transcript to the proper expression of a functional protein. A deep understanding of germ cell specification will come from studies spanning the journey from transcript to protein. Post-transcriptional regulation is essential for ensuring that Vasa is expressed in the right place in both sea star and sea urchin. Gus-Vasa binding was also reported in *Drosophila* and seems to be a regulatory interaction conserved with sea urchin although the binding residues are not conserved (Gustafson et al., 2011; Styhler et al., 2002). Is Gus-mediated regulation of Vasa protein conserved throughout evolution, present in both inductive and inherited germ cell specification mechanisms?

Transcription versus translation

Both transcriptional and post-transcriptional regulatory mechanisms have been identified for the germ cell determinants, Nanos and Vasa, and we see differences in the regulation of these genes when comparing different species. How has regulation of these highly conserved germ cell genes evolved, does it rely more heavily on

transcriptional or post-transcriptional events? Is there a balance between translational and transcriptional regulation throughout evolution? Changes within the gene sequence or the 3'UTR can point to changes in post-transcriptional regulation, such as changes or loss of the Nanos GNARLE sequence within the 3'UTR. The evolution of transcriptional regulation can be very dynamic, with changes occurring in the promoter or cis-regulatory sequences, transcription factors, cofactors, or in the structure of gene regulatory networks (Wray et al., 2003). A gene's promoter region can have multiple transcription factor binding sites, some of which are necessary for transcription while others are dispensable, and the effect of transcription factor binding is context dependent, so the identification of binding sites must be investigated under different conditions like different stages of development, for example. In the case of *Drosophila*, Nanos is expressed in the germarium of the ovary, during oogenesis. Nanos is expressed in the nurse cells which then deposit their contents to the oocyte (Wang et al., 1994). Knowing how Nanos transcription is regulated in the ovary, whether it depends on cell signaling, could help us deduce how transcriptional regulation of Nanos has evolved across species and how specificity is conveyed. Would we see an emphasis on post-transcriptional regulation in organisms that follow the inherited mechanism? Or is that opposed by the need to restrict germ cell gene expression, as we see in the sea star? Is the regulation of Nanos transcription conserved between *Drosophila* and the sea star even though its expression is seen under very different developmental conditions?

An analysis of genomic data from six animals including fly, fish, frog, worm, chicken, and humans sought to compare maternal and zygotic transcripts and any

trends in how they are regulated. Reliance on transcriptional or post-transcriptional regulation was assumed by lengths of the 3'UTR and 5'IGR (intergenic regions). The study concluded that mammalian maternal genes have evolved to have complex regulatory regions likely for increased specificity. On the other hand, maternal genes of egg-laying animals, the fly, zebrafish, and frog, displayed short 5'IGR relating to their lack of transcriptional specificity. Additionally, maternal genes presented long 3'UTRs consistent with a reliance on post-transcriptional regulation for proper expression (Shen-Orr et al., 2010). Maternal deposits, like yolk proteins, supplement energy requirements for the embryo and an emphasis on post-transcriptional regulation allows for developmental progression before zygotic expression begins (Hamm and Harrison, 2018). Nanos and Vasa are maternally deposited in the fly, frog, and zebrafish but not in mouse or humans. We might consider whether differences in the regulation of the germ cell genes, Nanos and Vasa, in particular, are due to their being maternal in some organisms or due to the mechanism of germ cell specification. The fly, frog, and zebrafish employ the inherited mechanism and Nanos is maternally deposited. In the sea urchin, which also follows the inherited mechanism, Nanos is not a maternal transcript. Does Nanos regulation in the sea urchin more closely resemble its regulation in the sea star or other organisms following the inherited mechanism? Discovery of β -catenin regulation of Nanos transcription in another species following the inherited mechanism would help answer this question. Can we trace a progressive reliance on post-transcriptional regulation of the germ cell genes in species that employ an inherited mechanism branching from an ancestor that follows an inductive mechanism? A focused study comparing the 3' and 5' regulatory regions of germ cell genes, in different

species spanning both inductive and inherited mechanisms, might reveal highly conserved regulatory regions, essential for proper expression, and what regulatory differences have evolved to be unique for a particular species.

Do all cells have the capacity to become germ cells?

The sea star embryo dissociation experiments reveal an interesting result; removing cell signaling does not result in a loss of Nanos transcription. Rather, Nanos mRNA abundance is upregulated compared to intact embryos of the same developmental stage and its expression level continues to increase with time in that dissociated state (Foster et al., 2022). This suggests that within the intact embryo Nanos is under inhibitory control, possibly to prevent overexpression. It also suggests that all cells have the necessary machinery to express the germ cell gene Nanos (Figure 3A). In the mouse, mutants for Smad2 and FoxH1, both Nodal signaling mediators, all epiblast cells take on a posterior fate and express the early germ cell marker Blimp1 (PRDM1) (Ohinata et al., 2009). This suggests that all epiblast cells are receptive to BMP signals of the extraembryonic ectoderm and have the ability to take on a germ cell fate (Figure 3B). In fact, Bmpr, the receptor for BMP2/4, is expressed ubiquitously during early embryogenesis (Mishina et al., 1995). Here Nodal signaling activity is critical, restricting germ cell gene expression and posterior fate.

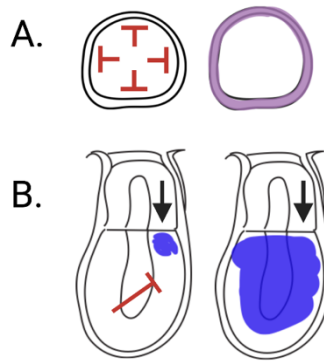


Figure 3. Capacity for germ cell gene expression in early embryo. (A) Schematic of sea star embryo hypothesis in which inhibition prevents widespread Nanos transcript expression at blastula stage. Nanos expression shown in purple. (B) Schematic representation of mouse early embryo. Activating BMP signals from the embryonic extoderm drive Blimp1 expression in the PGC precursors. Loss of inhibitory Nodal signals in the anterior embryo results in widespread expression of the germ cell marker Blimp1 in epiblast cells.

The germ cells must maintain pluripotency and an undifferentiated state as the embryo develops. For the sea star, the germ cell genes are not maternally deposited so their expression must be turned on and maintained. There are three potential conditions for the Nanos promoter region: inaccessible due to the chromatin being condensed, accessible and bound by a transcription factor (on), or accessible and bound by a repressor (off) (Figure 4A). Single-cell sequencing assay for transposase accessible chromatin (scATAC-seq) identifies accessible regions in the chromatin at single-cell resolution. This data can be displayed as sequencing peaks and can be used to cluster cell states, similar to scRNAseq data (Buenrostro et al., 2015; Yan et al., 2020). A scATAC-seq analysis of blastula and gastrula stage embryos will shed light on the chromatin landscape of the Nanos promoter region at two key developmental stages (Figure 4B,C). Analysis of blastula stage embryos would reveal whether the Nanos promoter is accessible at this stage before transcript accumulation is detected. An

absence of sequencing peaks in the Nanos promoter suggests a closed chromatin state and that Nanos expression is regulated at the epigenetic level in certain cell types. A potential outcome of scATAC-seq of blastulae is lack of sequencing peaks for Nanos in apical ectoderm and accessibility of the Nanos promoter in vegetal ectoderm or presumptive mesoderm. If we find the Nanos promoter region to be accessible in all cell types, Nanos expression could be specific to the vegetal embryo through one of two models— localized expression of the activator in the vegetal region, or a vegetally localized activator opposed by an inhibitor in the apical region (Figure 4B). By analyzing gastrula stage we can compare the chromatin landscape of the germ cell precursors to the rest of the embryo. I would expect a population of cells to show sequencing peaks in the promoter region for Nanos, Vasa, and the mesodermal gene Ets1, corresponding to cells in the archenteron, and another with peaks for Nanos, Vasa, and Brachyury, based on their co-expression in the hindgut region. Cells in these two regions of mesoderm development would also show accessibility of Vasa and FoxY3 promoters (Foster et al., 2022). For cells of the gut on the right side of the embryo, where Nodal signaling restricts Nanos expression, I expect the Nanos promoter region to be accessible as well as the Nodal promoter, this will differentiate left and right regions of the gut. There are cells of the endomesodermal region that, like the ectoderm, do not express Nanos. It will be interesting to find whether the Nanos promoter is accessible in these cells. Does Nanos expression not expand along the gut across the anterior/posterior axis because of regulation at the chromatin level or is the promoter accessible and occupied by a repressor? With Nodal morpholino treatment we saw an expansion of the Nanos expression domain in the gut, from left-sided restriction to forming a ring around the gut,

an expansion was not seen along the anterior/posterior axis (Fresques and Wessel, 2018). scATAC-seq will reveal whether endomesodermal cells corresponding to regions above and below Nanos-positive cells have condensed chromatin in the Nanos promoter region or whether Nanos expression is restricted by something other than Nodal. If the analysis of chromatin accessibility of early blastula shows accessible Nanos promoter throughout the embryo, this would suggest all cells of the early embryo have the potential to express Nanos. Could this be a hallmark of the inductive mechanism for which restriction is an important element? To follow up on the presence of an apically localized inhibitor in blastulae or an inhibitor besides Nodal in gastrulae, we can return to testing signaling pathway perturbation and assess effects on Nanos expression patterns by in situ hybridization. In summary, these efforts would reveal whether all cells of the blastula have the capacity to express Nanos, it will shed light on how Nanos expression is regulated in blastulae and how that may change with development, whether it relies on epigenetic and genetic methods.

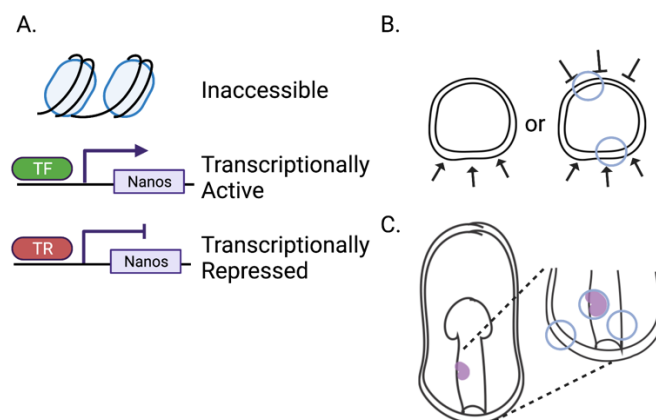
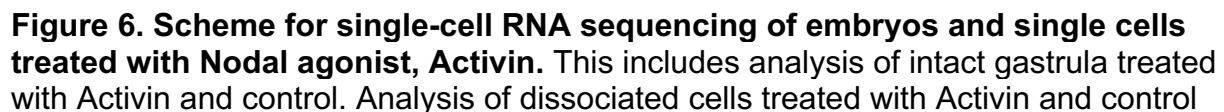
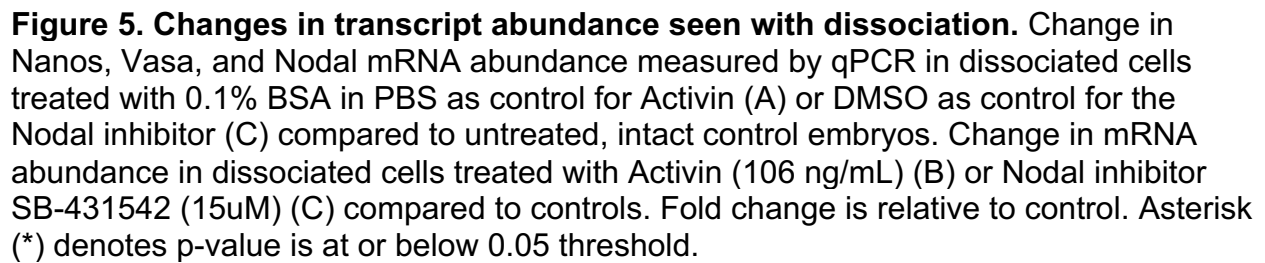


Figure 4. Analysis of chromatin landscape in the sea star embryo. (A) Potential transcriptional states for Nanos. The promoter region is inaccessible due to a closed chromatin state, the promoter is accessible and bound by a transcription factor (TF), or the promoter is accessible and bound by a transcriptional repressor (TR). (B) Models for regulation of localized Nanos expression at blastula stage. (Left) Activating signals in the vegetal embryo. (Right) Activating signals in the vegetal embryo contrasted with

inhibitory signals in the apical embryo. Gray circles represent embryonic cells that will be compared using scATAC-seq analysis. (C) Mid-gastrula stage embryo, Nanos mRNA expression, restricted to the left side of developing gut, shown in purple. Gray circles highlight embryonic cells that will be compared using scATAC-seq analysis, ectodermal cell types, Nanos-positive mesoderm, and Nanos-negative mesoderm. Created with BioRender.com

Another notable result from dissociation is an opposing trend seen between Nanos and Vasa mRNA abundance. We consistently see an increase in Nanos mRNA abundance in dissociated cells compared to intact controls (Figure 5A,C). This occurrence is paired with a drop in both Vasa and Nodal expression. Dissociated cells treated with Activin, a Nodal agonist, significantly upregulate Nodal and Vasa (Figure 5B). To further test this Nodal effect, we treated dissociated cells with an inhibitor, SB-431542, targeting the type I Nodal receptor. With this treatment, we saw a slight yet significant decrease in Nanos and now an upregulation of Vasa compared to controls, the opposite trend seen with control dissociation samples (Figure 5C,D). We did not see a reduction in Nodal signaling with inhibitor treatment. It could be that further inhibition of Nodal in dissociated cells resulted in upregulation of Nodal expression. The dissociation experiment led us to conclude that Nodal is acting directly on Nanos-positive cells yet there are several questions we need to address. First, does a small population of cells upregulate Nanos or is a greater population of cells expressing Nanos? We have two approaches for addressing this question. We can perform in situ hybridization for Nanos in single cells and compare the percent of cells expressing Nanos immediately after dissociation to that in dissociated cells cultured for several hours and to intact embryos. Another approach is to perform this same quantification of Nanos-positive cells using scRNAseq (Figure 6). To quantify Nanos positive cells in

dissociated cells, we can follow the same protocol used for the dissociation experiments. That is, we collect blastula stage embryos as a control, dissociate blastula and culture these cells in Activin and control, and collect intact embryos cultured for the same timespan as dissociated samples. In this way we have benchmarks for expression levels at blastula and gastrula stage. We can then compare the percent of Nanos-positive cells detected in blastula, gastrula, and dissociated controls. Additionally, we can measure gene expression changes in Nanos-positive dissociated cells compared to Nanos cells in an intact control embryo. A scRNAseq approach will also let us characterize the change in Vasa abundance with dissociation. I particularly want to determine whether there is a regulatory relationship between Nanos and Vasa in individual cells. Does Nanos overexpression inhibit or downregulate Vasa transcript expression? At this point, there are multiple interpretations we can make from this dissociation experiment. It is possible that removal of Nodal inhibition results in more cells of the embryo expressing Nanos (capacity for germ cell gene expression). It is also possible that in individual cells, the Nanos overexpression that results from loss of Nodal inhibition then downregulates Vasa. In sum, repeating this dissociation experiment followed by scRNAseq analysis rather than qPCR will teach us more about how Nanos and Vasa are regulated. A parallel experiment with intact embryos will complement this approach as we could detect the impact of Nodal signaling overexpression and disruption of Nanos/Vasa expression on other cell types of the embryo (Figure 6).



have an intact blastula stage and intact gastrula stage controls for benchmark gene expression levels. Nanos mRNA expression shown in purple.

How does cell contact regulate Nanos gene expression?

Another potential explanation for the significant Nanos upregulation seen with embryo dissociation is that cell contact plays a role in regulating germ cell factor expression and thus germline specification. Delta/Notch perturbation in early gastrula resulted in a loss of Vasa and Nanos mRNA abundance compared to untreated controls (Foster et al., 2022). This knowledge led us to hypothesize that dissociation, a state with reduced cell interactions, would result in a similar trend however, this was not the case for Nanos. One possibility is that signaling through cell contact, like through Delta/Notch or cadherins, inhibits Nanos transcription until gastrula stage at which point Nanos begins to accumulate. β -catenin is known to have dual functions in cell adhesion and signaling and could play a role in this sort of regulation. In the plasma membrane, β -catenin promotes cell adhesion by forming a complex with E-cadherin and α -cadherin. Key phosphorylation events regulate the loss of β -catenin-cadherin binding and promote binding to the nuclear cofactor BCL9 (B-cell lymphoma 9), which results in β -catenin activating transcription with its coactivator TCF/Lef, as in canonical Wnt signaling (Brembeck et al., 2006; MacDonald et al., 2009). A test for whether cell contact inhibits Nanos expression is to overexpress cadherin through injection of the transcript into oocytes, performing the dissociation experiment, and testing for change in mRNA abundance. Does cadherin overexpression sequester β -catenin and undo the Nanos upregulation seen with dissociation? To test a role for Delta/Notch instead, we could overexpress the Notch intracellular domain (NICD) which is cleaved upon Delta

binding to the receptor, translocates to the nucleus, and regulates gene expression (Kopan, 2012). The goal of this analysis is to simulate cell contact in dissociated cells and uncover whether the upregulation of Nanos mRNA abundance is associated with the loss of the repressive Nodal signal or an effect of reduced cell adhesion or both.

The importance of mesoderm

A mesodermal precursor state seems to be a common theme across studies in the mouse, axolotl, and cricket, all of which employ inductive germ cell formation (Chatfield et al., 2014; Ewen-Campen et al., 2013; Saitou et al., 2003). In the mouse, the PGC precursors become less responsive to BMP signals, which are known to advance mesoderm development (Saitou et al., 2003). BMP RNAi experiments in the cricket differentiated BMP-induced PGC formation from mesoderm formation (Nakamura and Extavour, 2016). In both of these examples what we see is a split between mesoderm and germ cell specification late in embryonic development. Is PGC development initially coupled with mesoderm development in the embryo? One test we have not performed is knocking down mesodermal genes to test if germ cell gene expression is affected. Low level of the mesodermal factor T/Brachyury was found to regulate a state of residual pluripotency essential for germ cell development in the mouse (Aramaki et al., 2021). This work found that T binds to regulatory regions of pluripotency genes such as Oct4 and Nanog, which are expressed in the mouse PGCs. High T expression inhibits PGC fate and thus the PGCs are induced in posterior epiblast cells with lower T levels. While this research reports binding of T to several pluripotency genes, it is unclear whether T/Brachyury expression remains when Nanos

is present in the mouse embryo. It is possible that a similar mechanism exists in the sea star and Nanos expression is activated or maintained in a subset of mesodermal cells based on levels of mesodermal factors. One way to test this is to perform a CRISPR/Cas9 knockout of endogenous Nanos and inject a Bacterial Artificial Chromosome (BAC) including the Nanos promoter region and ORF and perform an in situ hybridization for Nanos to test if we only see Nanos expression in mesodermal cell types (Figure 7A,Bi). This would tell us whether the transcriptional activators for Nanos are only present in mesodermal cells. As controls, I would produce BACs pairing the promoter region for the hindgut gene *Cdx1* or the ectodermal gene *SoxB2* with the Nanos ORF. I predict embryos injected with those BAC constructs would express Nanos in the hindgut or ectoderm respectively (Figure 7Bii,Biii). I would also inject a BAC with the GFP ORF under control of the Nanos promoter and in this way also test whether GFP expression is driven to the mesoderm in a pattern similar to that of endogenous Nanos transcripts (Figure 7Biv). While a scATACseq approach would uncover chromatin accessibility at the Nanos promoter in single cells, it will not tell us about the presence or absence of transcriptional regulators. To test whether T/Brachyury plays a conserved role in regulating germ cell specification in the sea star, I propose taking a morpholino approach to reduce expression of the protein and see by in situ hybridizations in late gastrula whether Nanos expression expands along the anterior/posterior axis of the gut (Figure 7Bv).

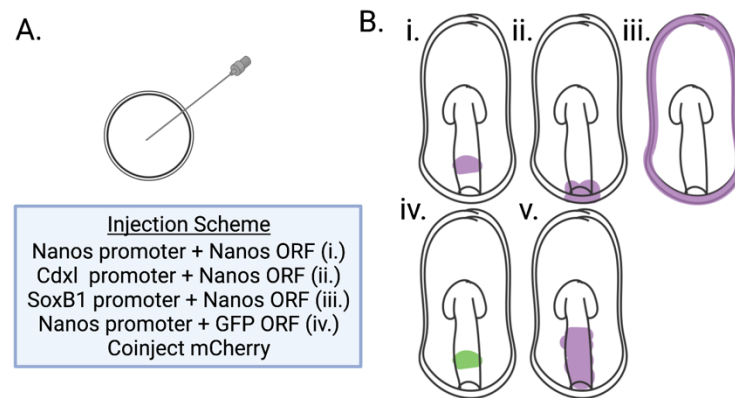


Figure 7. BAC injection approach for testing mesoderm-driven expression of Nanos. (A) Oocytes will be injected with BAC constructs to drive expression of the Nanos ORF under the control of Nanos, Cdx1 (hindgut), or SoxB1 (ectoderm) gene promoter regions. A BAC containing the GFP ORF under control of the Nanos promoter region will also be tested. All BAC will be coinjected with mCherry to identify injected oocytes. (B) Predicted expression outcomes for in situ hybridization with a probe for the Nanos ORF following BAC injection. Nanos enrichment in the center of the archenteron (i.), hindgut (ii.), or ectoderm (iii.) GFP expression in the center of the archenteron (iv.). Nanos probe detected throughout mesoderm, expanding across anterior/posterior axis (v.). Created with BioRender.com.

It would also be interesting to test whether cells that lose Nanos and Vasa expression go on to take a mesodermal or endodermal fate and whether loss of germ cell gene expression results in broad transcriptomic shifts for these cells. We could accomplish this with a time course scRNAseq experiment and reconstruct developmental trajectories of the embryonic cell types in sea star (Trapnell et al., 2014). This technique has been used to map mouse organogenesis and transcriptomic changes in mouse male germ cells (Cao et al., 2019; Zhao et al., 2021). Of course, these experiments need to be followed up through lineage tracing analyses. Although this direction may not enhance our knowledge of germ cell specification, it may lead to a greater understanding of mesodermal development and whether transient Nanos expression is functional in these cells.

Additional considerations for germ cell specification

There are several questions I did not get to address but believe are essential to pursue beginning with what role BMP signaling plays in specifying the sea star germline. A role for BMP regulation of germ cell specification has been reported in many species following the inductive mechanism including cricket, mouse, the acorn worm, and axolotl (Donoughe et al., 2014; Johnson et al., 2003; Lin et al., 2021; Ohinata et al., 2009). Preliminary evidence suggests that BMP signaling may regulate Nasa mRNA expression in the sea star. Treating early gastrula with recombinant human BMP4 results in an expansion of the Vasa mRNA expression region, from left-sided restriction to present across the gut (Foster et al., 2022). Opposing BMP and Nodal signals direct left/right axis formation in the urchin and this regulatory relationship may be conserved in the sea star (Luo and Su, 2012). Since germ cell gene expression is regulated via cell signaling in the sea star, it is necessary to determine whether signaling crosstalk is an element of germ cell specification. One occurrence to examine further is Vasa regulation at gastrula stage. We know that Nodal restricts Vasa expression from the right side (Fresques and Wessel, 2018). When we upregulate Wnt signaling or BMP signaling we see an expansion of Nanos signal from left to right (Foster et al., 2022). Although it seems β -catenin may bind and regulate Vasa, we have yet to test what happens to Nodal signal with Wnt overexpression. Immunolabeling for the downstream Nodal effector, phospho-Smad2/3, in LiCl treated and untreated controls will help us determine if Wnt signaling opposes Nodal inhibition. The mechanism of BMP influence on Vasa expression is not yet known but we can take a similar approach to test whether Nodal and BMP signals oppose each other in the sea star and if it affects the region of

germ cell gene expression. Pursuing BMP experiments in the sea star will reveal whether BMP signaling is an ancestral, highly conserved regulator of germ cell specification.

Descriptions of the inherited and inductive methods of germ cell specification can seem like two extremes— PGC specification from mesodermal precursors (induction) or early sequestering of maternally-derived germ cell determinants, separating PGCs from somatic cells (inherited). The difference between the inductive and inherited mechanisms is nuanced, likely without a clear divide. Not only do we see conservation of regulatory nodes between the two mechanisms of germ cell specification, but we also see elements of both mechanisms in germ cell specification in *Drosophila* (Colonna et al., 2022). Specification of the germ cells in *Drosophila* is dependent on the inheritance of pole plasm, which includes maternally derived Nanos and Vasa transcripts (Santos and Lehmann, 2004). A recent study reported that BMP signaling is required for proper germ cell specification, regulating Vasa accumulation in the pole cells, transcriptional quiescence, and proper germ plasm localization through interaction with the maternally deposited germ cell determinant, Oskar (Colonna et al., 2022). The presence of an inherited mechanism of germ cell specification does not obviate a role for cell signaling. The sea urchin may be one example where we see a signaling-dependent compensatory mechanism for specifying germ cells. Will the convergence of induction/cell signaling and inheritance methods prove to be more common with future research?

There are inputs outside of cell signaling that affect germ cell specification, metabolism and environmental factors such as pollutants and toxicants are a couple of

examples. Studies have shown a link between metabolic state and epigenetic regulation which in turn regulates cell fate transitions (Keating and El-Osta, 2015). Metabolomic and proteomic analysis of mouse PGCs compared to embryonic stem cells (ESC) and gonadal somatic cells showed decreased glycolysis and increased oxidative phosphorylation as a hallmark of PGC fate (Hayashi et al., 2017). Notably, a change in energy metabolism is required for the transition from naive pluripotency in epiblast-like cells (EpiLCs) to a PGC-like cell (PGCLC) state. The metabolite alpha-ketoglutarate (α -KG) promotes germ cell fate in an *in vitro* mouse model by enhancing the expression of key demethylases that regulate the loss of DNA methylation in the PGCs. Pseudotime analysis of EpiLC and PGCLC identified differential gene expression of enzymes that produce and convert α -KG which led to testing a role for α -KG (Tischler et al., 2019). We can begin to explore whether there are metabolic changes in the presumptive PGC precursors by looking for metabolic regulators in our single-cell datasets.

Bisphenol A (BPA), a chemical used to manufacture certain plastics, has been documented as a reproductive toxicant in mouse and human studies but whether it disturbs the early stages of germ cell formation has been understudied (Matuszczak et al., 2019). One study sought to document the effects of environmentally relevant exposure to BPA before PGC specification occurs. The researchers used an *in vitro* approach to target and measure the results of BPA exposure with greater sensitivity than a whole embryo approach in the mouse. In the study, induced mouse Epiblast-like cells (ELC) were treated with varying concentrations of BPA, and cultured into PGC-like cells (PGCLC). BPA exposure did not inhibit PGCLC specification; however, PGCLC exposed to BPA showed critical abnormalities in the transcription of metabolic genes,

expression of retrotransposons and gametogenesis genes, and showed signs of DNA damage (Ooi et al., 2021). While PGC development is not halted there are significant differences noted in these cells compared to untreated controls and the potential consequences of these abnormalities must be further analyzed. A sea urchin study reports spermiotoxicity, concentration-dependent growth reduction, and skeletal malformation seen with BPA exposure (Ozlem and Hatice, 2008). A directed study of BPA exposure and germ cell gene expression will fill an essential gap in understanding reproductive consequences in echinoderm species. Testing the outcomes of BPA exposure in sea star specifically could shed light on whether the cell signals involved in germ cell specification are altered and inform mammalian studies as the inductive mechanism is shared in these species.

The experimental approaches for studying germ cell specification presented in this chapter cover a range of proximity to direct transcriptional regulation of the germ cell genes. As we progressively focus on the details of transcriptional regulation for these genes, the context of an embryo's environment remains equally significant. Analysis of metabolic changes during germ cell specification and development confers information about the nutritive needs of the embryo and we can learn how inadequate nutrient availability affects reproductive development. Certainly, a link between nutritional conditions during early development and epigenetic markers inherited by offspring has been reported in the past, a well-known example of this is the Dutch Hunger Winter study (Heijmans et al., 2008; Lacal and Ventura, 2018). Chemicals and toxicants are included under the umbrella of environmental factors that influence reproduction. BPA exposure has been linked to disruption of oocyte maturation and of

the ovarian reserve, and even the appearance of lesions in the oviduct (Matuszczak et al., 2019; Richter et al., 2007). Those studies highlight the importance of environment in reproduction across several stages of development. Echinoderms spawn millions of gametes into the sea water; these animals are exposed to toxicants throughout their entire life cycle. Since our studies include animals arriving from variable environments, we anticipate seeing a range of effects of their environment on germ cell development. This aspect of our experimental model in itself can lead to great discovery of what germ cell development events are intrinsic to the embryo and which are dependent on the embryo's surroundings.

Future experimental plans

Analysis of Nanos protein expression

We have focused on the regulation of Nanos transcripts but must consider regulation of the protein to understand germ cell specification in greater depth. For example, we don't yet know whether cell interactions regulate protein levels of the germ cell genes. The dissociation experiment revealed that inhibition is an important aspect of Nanos expression regulation in the early embryo. It would be interesting to test whether the dissociated cells, which continuously upregulate Nanos transcript expression also express Nanos protein. Immunohistochemistry on dissociated cells using a Nanos antibody will answer this question. In the future, we could also test for presence of the Nanos protein in dissociated cells treated with the Nodal agonist Activin or other small-molecule cell signaling perturbants.

Evolution of Nanos transcript localization methods

The sea star Nanos 3'UTR does not have the GNARLE element conserved in the *S. purpuratus* and *L. variegatus* sea urchin species. The GNARLE element in the *S. purpuratus* Nanos 3'UTR is required for specific enrichment in the small micromeres (Oulhen et al., 2013). An interesting question to test is whether the machinery targeting Nanos transcripts for degradation outside of the small micromeres by the GNARLE sequence is conserved in the sea star. The GNARLE element is not present in the sea star but we don't yet know whether the machinery that binds this element and targets Nanos for degradation in urchins is present in the sea star. To better understand how the regulation of Nanos evolved in echinoderms we can test whether the sea urchin Nanos 3'UTR becomes localized in the sea star embryo. First, I would inject the sea star Nanos ORF with β -globin 3'UTR to test whether the 3'UTR of sea star Nanos plays a role in its localization. If the injected embryos show broader Nanos expression by in situ hybridization compared to controls it would suggest that the Nanos 3'UTR in sea star plays a role in its localized expression. An important future experiment will be to inject a construct of sea star (*P. miniata*) Nanos ORF with the sea urchin (*S. purpuratus*) Nanos 3'UTR and assess expression of the RNA construct by in situ hybridization. If the Nanos mRNA expression pattern for this construct differs from that of the endogenous Nanos, it would suggest that the urchin GNARLE element is functional in the sea star. Because in the sea urchin, GNARLE regulates Nanos specific enrichment in the small micromeres and its degradation everywhere else, I would expect widespread degradation of the construct in the sea star embryos. This experiment will shed light on whether the evolution of Nanos localization from sea star to sea urchin relies on the

acquisition of a GNARLE element alone or broader changes in the mechanism of selective transcript degradation.

A reporter for Nanos protein activity

A particular question I have is whether there is functional Nanos protein present in the early sea star embryo considering the expression pattern for Nanos mRNA before it is restricted to the PE. While an antibody for the sea star Nanos was made previously in the lab, immunolabeling of embryos has proven challenging because the antibodies do not effectively permeate the ectoderm. Further troubleshooting and adjusting of the immunolabeling protocol may enhance permeability for antibodies to label proteins within the embryo. However, immunohistochemistry will only report of presence of the Nanos protein. Nanos and Pumilio work together to repress translation of their target genes. Pumilio binds the Nanos Response Elements (NRE) in the 3'UTR of target RNAs and recruits Nanos to inhibit translation (Asaoka-Taguchi et al., 1999; Cho et al., 2006). Knowledge of where Pumilio is expressed would point towards potential regions of Nanos protein activity. Pumilio transcripts are ubiquitous in the sea star embryo (Fresques et al., 2014). This information does not tell us anything about where both Nanos and Pumilio may be acting. To determine whether the Nanos protein is active, we can design a reporter for Nanos activity. We can do this by creating a GFP construct with Nanos Response Elements (NRE) in the 3'UTR. If we inject oocytes with a construct of GFP with NREs in the 3'UTR then we can see in live embryos where GFP is expressed and more importantly if the protein is absent in cells of mid-gastrulae as Nanos mRNA expression is first seen in this stage (Fresques et al., 2016). If there is no

Nanos protein activity in gastrula stages I would expect to see GFP protein in all cells of the embryo. If there is active Nanos and Pumilio protein where we see Nanos transcript expression, I would expect GFP protein to be absent from the vegetal region of the early gastrula and the center of the archenteron in later gastrula stages. If Nanos protein is not present until after PE formation, I would expect GFP protein to be selectively excluded from the PE and present in the rest of the embryo. Injection of a GFP construct with a sea star Nanos 3'UTR in which the NRE sequences have been mutated would serve as a control. In embryos injected with this construct I would expect GFP protein to be present throughout the embryo. This experiment informs us about presence and activity of the Nanos protein in early sea star development, which we don't yet understand. Moreover, we can use this reporter for Nanos activity to test whether cell signaling or other cues regulate Nanos at the protein level.

Quality control measures in scRNAseq datasets

My research produced the first reported single cell analysis of a sea urchin and sea star species (Foster et al., 2022; Foster et al., 2019). The quality control (QC) measures of the data analysis closely followed the protocols composed for the Seurat V2 and V3 with some notable differences based on the fact that we used a new species (Butler et al., 2018; Stuart et al., 2019). A quality control step described in other scRNAseq projects is the removal of low-quality cells which can mean outliers having gene counts well above the average, known as multiples (more than one cell per droplet), or well below average meaning empty droplets. Additionally, broken cells are removed before subsequent analysis based on their higher levels of mitochondrial gene

counts (Ilicic et al., 2016). Removal of low-quality cells is meant to produce more accurate results and profile of the embryo. While I was able to set cut-offs to remove outliers, unfortunately, the removal of broken cells is currently difficult. The current nomenclature and annotations of the genomes does not allow us to simply extract a list of mitochondrial genes. In the future, approaches like re-annotating mitochondrial genes would greatly simplify filtering low-quality cells resulting from the capture of broken or lysed cells. Re-annotation of the echinoderm genomes will also be useful in matching the nomenclature of genes conserved across different species. This would significantly facilitate the comparison of gene expression patterns across species and standardizing removal of low-quality cells across datasets from different species. Additionally, filtering low-quality cells results in more accurate clustering and subsequently more accurate integration of datasets. Integration is essential for the identification of gene expression changes of a particular cell identity through trajectory analysis. Trajectory analysis will be an important part of future studies to track the putative germ cells in the sea star and identify which of the mesodermal Nanos-positive cells make up the PE.

References

- Aramaki, S., Kagiwada, S., Wu, G., Obridge, D., Adachi, K., Kutejova, E., Lickert, H., Hubner, K., Scholer, H.R., 2021. Residual pluripotency is required for inductive germ cell segregation. *EMBO Rep* 22, e52553.
- Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K., Kobayashi, S., 1999. Maternal Pumilio acts together with Nanos in germline development in *Drosophila* embryos. *Nat Cell Biol* 1, 431-437.
- Brembeck, F.H., Rosario, M., Birchmeier, W., 2006. Balancing cell adhesion and Wnt signaling, the key role of beta-catenin. *Curr Opin Genet Dev* 16, 51-59.
- Briggs, E., Wessel, G.M., 2006. In the beginning...animal fertilization and sea urchin development. *Dev Biol* 300, 15-26.
- Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., Greenleaf, W.J., 2015. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486-490.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., Satija, R., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* 36, 411-420.
- Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., Trapnell, C., Shendure, J., 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566, 496-502.
- Chatfield, J., O'Reilly, M.A., Bachvarova, R.F., Ferjentsik, Z., Redwood, C., Walmsley, M., Patient, R., Loose, M., Johnson, A.D., 2014. Stochastic specification of primordial germ cells from mesoderm precursors in axolotl embryos. *Development* 141, 2429-2440.
- Cho, P.F., Gamberi, C., Cho-Park, Y.A., Cho-Park, I.B., Lasko, P., Sonenberg, N., 2006. Cap-dependent translational inhibition establishes two opposing morphogen gradients in *Drosophila* embryos. *Curr Biol* 16, 2035-2041.
- Colonna, M.M., Goyal, Y., Johnson, H.E., Syal, S., Schedl, P., Deshpande, G., 2022. Preformation and epigenesis converge to specify primordial germ cell fate in the early *Drosophila* embryo. *PLoS Genet* 18, e1010002.
- Dahanukar, A., Wharton, R.P., 1996. The Nanos gradient in *Drosophila* embryos is generated by translational regulation. *Genes Dev* 10, 2610-2620.
- Derbès, A.A., 1847. Observations sur le Mécanisme et les Phénomènes qui Accompagnent la Formation de l'Embryon chez l'Oursin Comestible. *Ann Sci Nat Zool* 8, 80-98.
- Dono, R., Scalera, L., Pacifico, F., Acampora, D., Persico, M.G., Simeone, A., 1993. The murine *cripto* gene: expression during mesoderm induction and early heart morphogenesis. *Development* 118, 1157-1168.
- Donoughe, S., Nakamura, T., Ewen-Campen, B., Green, D.A., 2nd, Henderson, L., Extavour, C.G., 2014. BMP signaling is required for the generation of primordial germ cells in an insect. *Proc Natl Acad Sci U S A* 111, 4133-4138.
- Ewen-Campen, B., Donoughe, S., Clarke, D.N., Extavour, C.G., 2013. Germ cell specification requires zygotic mechanisms rather than germ plasm in a basally branching insect. *Curr Biol* 23, 835-842.

- Extavour, C.G., Akam, M., 2003. Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* 130, 5869-5884.
- Foster, S., Oulhen, N., Fresques, T., Zaki, H., Wessel, G., 2022. Distinct mechanisms of germ cell factor regulation for an inductive germ cell fate. *bioRxiv*, 2022.2002.2004.479164.
- Foster, S., Oulhen, N., Wessel, G., 2020. A single cell RNA sequencing resource for early sea urchin development. *Development* 147.
- Foster, S., Teo, Y.V., Neretti, N., Oulhen, N., 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.
- Fresques, T., Swartz, S.Z., Juliano, C., Morino, Y., Kikuchi, M., Akasaka, K., Wada, H., Yajima, M., Wessel, G.M., 2016. The diversity of nanos expression in echinoderm embryos supports different mechanisms in germ cell specification. *Evol Dev* 18, 267-278.
- Fresques, T., Zazueta-Novoa, V., Reich, A., Wessel, G.M., 2014. Selective accumulation of germ-line associated gene products in early development of the sea star and distinct differences from germ-line development in the sea urchin. *Dev Dyn* 243, 568-587.
- Fresques, T.M., Wessel, G.M., 2018. Nodal induces sequential restriction of germ cell factors during primordial germ cell specification. *Development* 145.
- Gao, C., Chen, Y.G., 2010. Dishevelled: The hub of Wnt signaling. *Cell Signal* 22, 717-727.
- Gildor, T., Hinman, V., Ben-Tabou-De-Leon, S., 2017. Regulatory heterochronies and loose temporal scaling between sea star and sea urchin regulatory circuits. *Int J Dev Biol* 61, 347-356.
- Gustafson, E.A., Yajima, M., Juliano, C.E., 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.
- Hamm, D.C., Harrison, M.M., 2018. Regulatory principles governing the maternal-to-zygotic transition: insights from *Drosophila melanogaster*. *Open Biol* 8, 180183.
- Hayashi, Y., Otsuka, K., Ebina, M., Igarashi, K., Takehara, A., Matsumoto, M., Kanai, A., Igarashi, K., Soga, T., Matsui, Y., 2017. Distinct requirements for energy metabolism in mouse primordial germ cells and their reprogramming to embryonic germ cells. *Proc Natl Acad Sci U S A* 114, 8289-8294.
- Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E., Lumey, L.H., 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A* 105, 17046-17049.
- Hinman, V.F., Burke, R.D., 2018. Embryonic neurogenesis in echinoderms. *Wiley Interdiscip Rev Dev Biol* 7, e316.
- Illicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C., Teichmann, S.A., 2016. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol* 17, 29.
- Johnson, A.D., Crother, B., White, M.E., Patient, R., Bachvarova, R.F., Drum, M., Masi, T., 2003. Regulative germ cell specification in axolotl embryos: a primitive trait

- conserved in the mammalian lineage. *Philos Trans R Soc Lond B Biol Sci* 358, 1371-1379.
- Juliano, C.E., Voronina, E., Stack, C., Aldrich, M., Cameron, A.R., Wessel, G.M., 2006. Germ line determinants are not localized early in sea urchin development, but do accumulate in the small micromere lineage. *Dev Biol* 300, 406-415.
- Keating, S.T., El-Osta, A., 2015. Epigenetics and Metabolism. *Circulation Research* 116, 715-736.
- Kopan, R., 2012. Notch signaling. *Cold Spring Harb Perspect Biol* 4.
- Lacal, I., Ventura, R., 2018. Epigenetic Inheritance: Concepts, Mechanisms and Perspectives. *Frontiers in Molecular Neuroscience* 11.
- Lai, F., King, M.L., 2013. Repressive translational control in germ cells. *Molecular Reproduction and Development* 80, 665-676.
- Lin, C.Y., Yu, J.K., Su, Y.H., 2021. Evidence for BMP-mediated specification of primordial germ cells in an indirect-developing hemichordate. *Evol Dev* 23, 28-45.
- Logan, C.Y., Miller, J.R., Ferkowicz, M.J., McClay, D.R., 1999. Nuclear beta-catenin is required to specify vegetal cell fates in the sea urchin embryo. *Development* 126, 345-357.
- Luo, Y.J., Su, Y.H., 2012. Opposing nodal and BMP signals regulate left-right asymmetry in the sea urchin larva. *PLoS Biol* 10, e1001402.
- MacDonald, B.T., Tamai, K., He, X., 2009. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell* 17, 9-26.
- Materna, S.C., Swartz, S.Z., Smith, J., 2013. Notch and Nodal control forkhead factor expression in the specification of multipotent progenitors in sea urchin. *Development* 140, 1796-1806.
- Matuszczak, E., Komarowska, M.D., Debek, W., Hermanowicz, A., 2019. The Impact of Bisphenol A on Fertility, Reproductive System, and Development: A Review of the Literature. *Int J Endocrinol* 2019, 4068717.
- McCauley, B.S., Akyar, E., Filliger, L., Hinman, V.F., 2013. Expression of wnt and frizzled genes during early sea star development. *Gene Expr Patterns* 13, 437-444.
- Mishina, Y., Suzuki, A., Ueno, N., Behringer, R.R., 1995. Bmpr encodes a type I bone morphogenetic protein receptor that is essential for gastrulation during mouse embryogenesis. *Genes Dev* 9, 3027-3037.
- Nakamura, T., Extavour, C.G., 2016. The transcriptional repressor Blimp-1 acts downstream of BMP signaling to generate primordial germ cells in the cricket *Gryllus bimaculatus*. *Development* 143, 255-263.
- Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., Saitou, M., 2009. A signaling principle for the specification of the germ cell lineage in mice. *Cell* 137, 571-584.
- Ooi, S.K.T., Jiang, H., Kang, Y., Allard, P., 2021. Examining the Developmental Trajectory of an in Vitro Model of Mouse Primordial Germ Cells following Exposure to Environmentally Relevant Bisphenol A Levels. *Environ Health Perspect* 129, 97013.
- Oulhen, N., Morita, S., Wessel, G.M., 2022. Post-transcriptional regulation of factors important for the germ line. *Curr Top Dev Biol* 146, 49-78.

- Oulhen, N., Swartz, S.Z., Wang, L., Wikramanayake, A., 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.
- Oulhen, N., Wessel, G.M., 2014. Every which way--nanos gene regulation in echinoderms. *Genesis* 52, 279-286.
- Oulhen, N., Yoshida, T., Yajima, M., Song, J.L., Sakuma, T., Sakamoto, N., Yamamoto, T., 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.
- Ozlem, C.A., Hatice, P., 2008. Effects of bisphenol A on the embryonic development of sea urchin (*Paracentrotus lividus*). *Environ Toxicol* 23, 387-392.
- Peng, C.J., Wikramanayake, A.H., 2013. Differential regulation of disheveled in a novel vegetal cortical domain in sea urchin eggs and embryos: implications for the localized activation of canonical Wnt signaling. *PLoS One* 8, e80693.
- Perillo, M., Oulhen, N., Foster, S., Spurrell, M., Calestani, C., Wessel, G., 2020. Regulation of dynamic pigment cell states at single-cell resolution. *Elife* 9.
- Perillo, M., Swartz, S.Z., Wessel, G.M., 2022. A conserved node in the regulation of Vasa between an induced and an inherited program of primordial germ cell specification. *Dev Biol* 482, 28-33.
- Persani, L., Rossetti, R., Cacciatori, C., 2010. Genes involved in human premature ovarian failure. *J Mol Endocrinol* 45, 257-279.
- Richter, C.A., Birnbaum, L.S., Farabollini, F., Newbold, R.R., Rubin, B.S., Talsness, C.E., Vandenberg, J.G., Walser-Kuntz, D.R., vom Saal, F.S., 2007. In vivo effects of bisphenol A in laboratory rodent studies. *Reprod Toxicol* 24, 199-224.
- Saitou, M., Payer, B., Lange, U.C., Erhardt, S., Barton, S.C., Surani, M.A., 2003. Specification of germ cell fate in mice. *Philos Trans R Soc Lond B Biol Sci* 358, 1363-1370.
- Santos, A.C., Lehmann, R., 2004. Germ cell specification and migration in *Drosophila* and beyond. *Curr Biol* 14, R578-589.
- Shen, M.M., 2007. Nodal signaling: developmental roles and regulation. *Development* 134, 1023-1034.
- Shen-Orr, S.S., Pilpel, Y., Hunter, C.P., 2010. Composition and regulation of maternal and zygotic transcriptomes reflects species-specific reproductive mode. *Genome Biol* 11, R58.
- Smith, A.B., Pisani, D., Mackenzie-Dodds, J.A., Stockley, B., Webster, B.L., Littlewood, D.T., 2006. Testing the molecular clock: molecular and paleontological estimates of divergence times in the Echinoidea (Echinodermata). *Mol Biol Evol* 23, 1832-1851.
- Song, J.L., Wessel, G.M., 2012. The forkhead transcription factor FoxY regulates Nanos. *Mol Reprod Dev* 79, 680-688.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., Satija, R., 2019. Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888-1902 e1821.
- Styhler, S., Nakamura, A., Lasko, P., 2002. VASA localization requires the SPRY-domain and SOCS-box containing protein, GUSTAVUS. *Dev Cell* 3, 865-876.

- Swartz, S.Z., Reich, A.M., Oulhen, N., Raz, T., Milos, P.M., Campanale, J.P., Hamdoun, A., Wessel, G.M., 2014. Deadenyase depletion protects inherited mRNAs in primordial germ cells. *Development* 141, 3134-3142.
- Swartz, S.Z., Tan, T.H., Perillo, M., Fakhri, N., Wessel, G.M., Wikramanayake, A.H., Cheeseman, I.M., 2021. Polarized Dishevelled dissolution and reassembly drives embryonic axis specification in sea star oocytes. *Curr Biol* 31, 5633-5641 e5634.
- Tischler, J., Gruhn, W.H., Reid, J., Allgeyer, E., Buettner, F., Marr, C., Theis, F., Simons, B.D., Wernisch, L., Surani, M.A., 2019. Metabolic regulation of pluripotency and germ cell fate through alpha-ketoglutarate. *EMBO J* 38.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., Rinn, J.L., 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 32, 381-386.
- Voronina, E., Lopez, M., Juliano, C.E., Gustafson, E., Song, J.L., Extavour, C., George, S., Oliveri, P., McClay, D., 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.
- Wang, C., Dickinson, L.K., Lehmann, R., 1994. Genetics of nanos localization in *Drosophila*. *Dev Dyn* 199, 103-115.
- Weitzel, H.E., Illies, M.R., Byrum, C.A., Xu, R., Wikramanayake, A.H., Ettensohn, C.A., 2004. Differential stability of beta-catenin along the animal-vegetal axis of the sea urchin embryo mediated by dishevelled. *Development* 131, 2947-2956.
- Wessel, G.M., Brayboy, L., Fresques, T., Gustafson, E.A., Oulhen, N., Ramos, I., Reich, A., Swartz, S.Z., Yajima, M., Zazueta, V., 2014a. The biology of the germ line in echinoderms. *Mol Reprod Dev* 81, 679-711.
- Wessel, G.M., Fresques, T., Kiyomoto, M., Yajima, M., Zazueta, V., 2014b. Origin and development of the germ line in sea stars. *Genesis* 52, 367-377.
- Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., Romano, L.A., 2003. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* 20, 1377-1419.
- Yajima, M., Wessel, G.M., 2011. Small micromeres contribute to the germline in the sea urchin. *Development* 138, 237-243.
- Yan, F., Powell, D.R., Curtis, D.J., Wong, N.C., 2020. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol* 21, 22.
- Yeo, C., Whitman, M., 2001. Nodal signals to Smads through Cripto-dependent and Cripto-independent mechanisms. *Mol Cell* 7, 949-957.
- Zhang, F., Liu, R., Liu, C., Zhang, H., Lu, Y., 2020. Nanos3, a cancer-germline gene, promotes cell proliferation, migration, chemoresistance, and invasion of human glioblastoma. *Cancer Cell International* 20, 197.
- Zhao, J., Lu, P., Wan, C., Huang, Y., Cui, M., Yang, X., Hu, Y., Zheng, Y., Dong, J., Wang, M., Zhang, S., Liu, Z., Bian, S., Wang, X., Wang, R., Ren, S., Wang, D., Yao, Z., Chang, G., Tang, F., Zhao, X.Y., 2021. Cell-fate transition and determination analysis of mouse male germ cells throughout development. *Nat Commun* 12, 6839.

Appendix A: Identifying gene expression from single cells to single genes.

Oulhen N., **Foster S.**, Wray G., Wessel G.M. Identifying gene expression from single cells to single genes. *Methods Cell Biol* 2019, **151**:127-158.

I contributed to writing the single cell sequencing portion of this methods chapter. This methods chapter was a culmination of our efforts to test different single-cell RNA sequencing platforms for analysis of early echinoderm embryos.

Summary

Gene regulatory networks reveal how single transcription factors contribute to a dynamic cascade of cellular information processing. Necessary to understand such a network within cells, between cells, and in cells over time, is a set of different technologies to expose the mechanism effectively. Recent advances in technologies have broadened the toolkit for testing GRN mechanisms and connections. Here we will emphasize three approaches that we have found important for interrogating transcriptional mechanisms in echinoderms: single cell mRNA sequencing (Drop-seq), nascent RNA detection and identification, and chromatin immunoprecipitation (ChIP). We present these applications in order since it is a logical experimental protocol as well. With preliminary information from bulk mRNA transcriptome analysis and differential gene expression studies (DE-seq), one may need to test in what specific cells important genes may be expressed and use single cell sequencing to define such links. Nascent RNA analysis with the Click-iT chemistry allows the investigator to deduce when the RNA was transcribed, not just identify its presence, and ChIP allows the investigator to study direct interactions of putative transcriptional regulators with the gene promoter of interest. This flow of thinking, and the technologies to support it, are presented here for echinoderms. While many of the procedures are general and applicable to many organisms and cell types, we emphasize unique aspects of the protocols for consideration in using echinoderm embryos, larvae, and adult tissues.

A. Single cell sequencing

1. Transcriptomics and utility of single cell mRNA sequencing

All cells of the body come from a single cell, and the many siblings resulting from cell divisions diversify in every way imaginable. Understanding the diversity in what runs each cell means examining each cell apart from its neighbors. Current technology is limited in how we can see proteins, lipids, sugars, and metabolites, but polymers of nucleic acids can be replicated experimentally in vitro, making them more abundant and visible. Investigators recently have been able to do so from a single, isolated cell. Thus, we can now visualize one part of a cell's fingerprint that drives their form, fate, and function, and begin to understand how diverse a cell type, a tissue, and an organism really is.

This section will summarize the procedure of single cell mRNA sequencing in echinoderms. Now, instead of generating bulk transcriptomes from whole populations of embryos, one can generate transcriptomes from individual cells representing either a specific germ layer, a specific developmental stage, or a tissue type.

The protocol for mRNA-drop sequence analysis is common for many cells from diverse organisms. These protocols have been applied to organs, tissues, and cell types as diverse as human blood, mouse brains, fly embryos, and adult Hydra. The controllers for making individual cell/drops are diverse, but all share common principles. The ways in which each population of mRNA from a single cell is processed is also diverse, but again with shared principles. We will summarize these various considerations here for single cell RNA-seq, and emphasize how echinoderms have been processed for drop-seq analysis in order to maximize utility of the technology. Overall, echinoderm

embryos, larvae, and adult tissues are wonderfully well suited for such analysis, and the results from these analyses should be transformative over the years to come.

2. Summary of current instrumentation available for single cell sequencing

a. Comparison of Cost, Benefit, Yield

Single-cell RNA sequencing (scRNA-seq) allows mRNA transcriptome wide analysis of individual cells. This involves three major steps; isolating individual cells, extracting RNA and converting it to cDNA, and finally creating a cDNA sequencing library.

Several methods are available to perform single cell sequencing analysis. These methods vary in the way they isolate cells and generate cDNA, and as a result they vary in their sensitivity and accuracy (among other things i.e. cost).

A recent review (Ziegenhain et al. 2017) summarizes and analyzes six methods for performing scRNA-seq discussing accuracy, sensitivity, power, and cost efficiency.

- CEL-seq2
- DROP-seq
- MARS-seq
- SCR-seq
- Smart-seq
- Smart-seq2

All single cell methods require tissue dissociation followed by cell isolation. Cell isolation can be plate based, as with CEL-seq2 and Smart-seq. Cells can be sorted via Fluorescence Activated Cell Sorting (FACS) into a 96-well or 384-well plate, which is

the case for MARS-seq, SCRB-seq, and Smart-seq2. In the Drop-seq method, controlled fluid flows interact such that cells are isolated in microdroplets within a microfluidic chamber.

Sensitivity and Accuracy

Sensitivity refers to the ability to capture an mRNA transcript for inclusion in the cDNA sequencing library. To address sensitivity, the number of genes detected per cell is measured and this value is compared across the six methods. Smart-seq2 showed the highest sensitivity measure and Drop-seq the lowest.

Accuracy refers to how well the read quantification matches the concentration of mRNA transcripts. To measure accuracy, synthetic RNA transcripts of known concentration from the External RNA Control Consortium (ERCC) were spiked into the samples and the expression values of this RNA was calculated. Accuracy was comparable between the scRNA-seq methods, with Drop-seq having a slightly higher accuracy measure over other techniques.

Cost efficiency

The authors calculated the minimal cost for creating sequencing libraries for the number of cells needed for a power of 80% at three sequencing depths, 1, 0.5, and 0.25 million reads. Drop-seq is most cost effective for sequencing at a depth of e.g. 250,000 reads.

Table 1 (from Ziegenhain et al. 2017)

Method	Cell Number per Group*	Minimal Cost (\$)
CEL-seq2/C1	85/100/110	~2,420/2,310/2,250

Drop-seq	99/135/254	~1,010/700/690
MARS-seq	110/135/160	~1,380/1,030/820
SCRB-seq	64/90/166	~900/810/1,080
Smart-seq/C1	150/172/215	~9,010/9,440/11,290
Smart-seq2 (commercial Tn5)	95/105/128	~10,470/11,040/13,160
Smart-seq2 (in-house Tn5)	95/105/128	~1,520/1,160/1,090

*Cell groups listed for sequencing depth of 1, 0.5, and 0.25 million reads.

Differences in cDNA generation

The cDNA molecules generated using Smart-seq and Smart-seq2 are not barcoded and instead emphasize full length sequencing coverage of transcripts. Cel-seq, MARS-seq, SCRB-seq, and Drop-seq all produce barcoded cDNA transcripts and add a unique molecular identifier (UMI) sequence. Barcoding and UMI sequences are a particular advantage of these single cell methods and enables tracking the well or cell of origin for transcripts. The use of UMIs improves mRNA quantification because you can differentiate between original mRNA molecules and those derived from the amplification step.

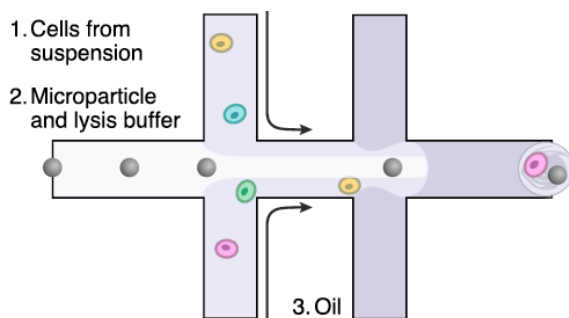
In sum, the nature and goals of your experiment will drive your choice of single cell RNA sequencing protocol. In cases where you have a sample with a low cell count, perhaps enriched by FACS, MARS-seq and Smart-seq2 would be advised. Drop-seq allows you to analyze a large number of cells and is the most cost-effective for generating sequencing libraries. Therefore, Drop-seq is recommended when trying to identify a particular, small population of cells. In our case, the goal was to identify a rare population of cells (primordial germ cells PGC/small micromeres) in the sea urchin and sea star.

Here we provide detail on two Drop-seq instruments we have utilized, the 10X Chromium and the Dolomite Nadia instrument.

Drop-seq captures single cells by use of a microfluidic chamber. A flow of beads suspended in lysis buffer, a flow of cells, and a flow of oil meet to form thousands of combined emulsion droplets. The beads are coated with oligo-dT primers, which include a PCR priming sequence, a specific barcode unique to each bead, and a unique molecular identifier (UMI). The number of droplets formed greatly outnumbers the number of beads and cells in order to avoid multiplets, cases of more than one bead or cell in a droplet. Once cells are isolated within the droplets, the cells are lysed and the beads capture polyadenylated mRNAs.

10X Chromium

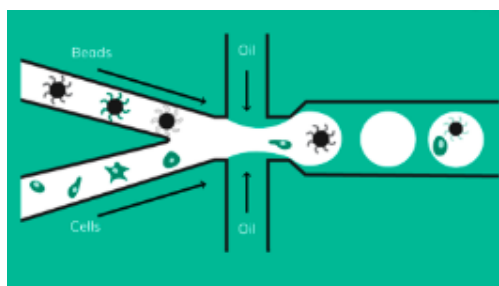
10X Genomics reports approximately 65% cell processing efficiency for the 10X Chromium instrument. The 10X can capture between 500 and 10,000 cells per channel, with a capability of up to 8 channels per run. Cells up to 30µm in diameter have been tested on this device. The microfluidic technology of the 10X Chromium follows that of the Macosko et al 2015 drop-seq publication. The resulting gel beads in emulsion are referred to as GEMs.



(Macosko et al., 2015)

Nadia

The Nadia, like the 10X Chromium, is a microfluidic-based system. It has a 10% cell processing efficiency when following recommended concentrations of 300 cells/ μL and 600 beads/ μL . The Nadia can capture up to 50,000 cells in a single run because it can run a larger volume and a larger number of cells in one run compared to the 10X. The Nadia can accommodate cell sizes up to 50 μm .



(from <https://www.dolomite-bio.com/support/downloads/>)

Analysis

Approximately 5,000 cells are generally processed through the 10X Genomics Chromium instrument for embryos. After UMI filtering of the sequencing data, 3,348 (66%) cells are recovered for sequence data analysis. In a comparable Nadia experiment, the recommended concentrations for the cell and bead aliquots are different, but for the same embryo experiments approximately 75,000 cells are run according to manufacturer's recommendations.

3. Preparation of single cells – tissue dissociation – and CRISPR embryos

a. Obtaining single cells

Dissociating cells from embryos, tissues, or organs requires several important considerations. First, complete dissociation of cells is essential. Doublets, or higher ordered aggregates of cells that go through the drop-seq protocol will be averaged in their transcripts, thereby diminishing the strength of the single-cell procedure. Second, it is essential to have all cells represented proportionally in the dissociation. If the tissue is not completely dissociated, the clumps of cells remaining may represent a distinct population of cells that now will not be present in the analysis. Third, many of the drop-seq protocols can read 5,000 – 10,000 cells per run. If the stage of interest for single cell sequencing is a larva with several thousand cells, and many different cell types, each of them may not be represented at sufficient depth in the final analysis. In such cases it is important to perform multiple runs of the same embryos, and/or to enrich for cells of interest. For example, if one were interested in pigment cells in the ectoderm of a sea urchin larva, one can dissociate the ectoderm off the larvae, spin out the remainder of the larvae, and drop-seq only the separated ectodermal cells. In this case one has decreased the diversity of cell types, and enriched many fold the cells of interest. Where possible, fluorescence activated cell sorting is also of great utility, but does require specific fluorescence labeling of the cells of interest. Protocols for differential dissociation of embryos have been documented in this series of manuals (see Chapter XXX (McClay)).

b. Embryos vs larvae vs adult tissues

Many drop-seq instruments and microfluidic chips can handle cells upwards of 40 - 50 microns in diameter. Larger cells, eggs, early blastomeres, however, will get clogged in

the chip and abort the run. By morula stage and later though in many echinoderm species, each cell is less than ~25 microns and is able to flow through a microfluidic device. If the sample of interest is prior to hatching, make sure to treat the embryos at fertilization to inhibit crosslinking of the fertilization envelope so that the embryo can be dissociated. The protocol for this is detailed in Chapter XXX of this series (reference) , and usually uses 2mM 3-aminotriazole or 10mM p-aminobenzoic acid (pH7.8).

Stripping the non-crosslinked envelopes off the embryos usually involves pouring the embryos through a nylon mesh, the size of which is just smaller than the size of the embryos e.g. 64 micron mesh for embryos of *S. purpuratus* ~80 microns.

To dissociate embryos, larvae and even some adult tissues, we use procedures based on treatments with calcium – free media, followed by gentle dissociation using a loose fitting Dounce homogenizer. For *S. purpuratus* embryos and early larvae we use the following protocol:

- i. All procedures are on ice, using media that has been stored at 4°C and is kept on ice during the dissociation.
- ii. Wash embryos 2x with CaFSW. Pellet embryos using either settling at 1x g, centrifugation in a hand centrifuge for 15 seconds, or 8 seconds in a microfuge (for small numbers e.g. <1000; simply push the manual start button on the microfuge and release it after 8 seconds. The speed will gradually increase from 0 rpm to approximately 8,000 during this time). Aspirate off media and repeat, resuspending embryos each time, and filling the tube each time with new media.

- iii. Resuspend embryos in hyalin extraction media (HEM; see below for recipe) for 10 minutes on ice, for embryos prior to gastrulation, 20 minutes on ice, for later embryos (post-gastrulation and larvae), inverting every few minutes.
- iv. Pellet embryos and resuspend in 1-2 mls of 0.5M NaCl. The exact volume will depend on the number of embryos and the concentration of cells needed for the drop-seq analysis. This level of this salt does not appear to alter subsequent steps in the procedures.
- v. Place the resuspended embryos in a cold Dounce homogenizer and if possible, perform the douncing under a stereo microscope, watching the effectiveness of dissociation with each stroke. If visualizing under the microscope directly in the Dounce tube is not possible, take a small sample out after a few strokes and visualize embryos and dissociated cells on a microscope slide. Cease douncing when the vast majority of embryos have been dissociated.
- vi. Pour the dissociated cells through a 40 micron cell strainer to remove any clumps of cells remaining, and evaluate and count the single cells on a hemacytometer. Depending on the drop-seq controller used, cell concentrations in the range of 300/microliter are targeted. If you need to concentrate the cells, spin in a microfuge tube for 12 seconds (as above, from manually starting) and resuspend the cell pellet in 0.5 M NaCl to the desired concentration. In general, pellet the single cells as little as possible. Even though they are robust, one should minimize breakage in cells of a population that may be more fragile than others, and thereby skew the dataset.

- vii. Although embryos and early larvae from sea urchins dissociate well with calcium-free protocols, some adult tissues do not dissociate to single cells effectively using only this approach. Many mammalian tissues are dissociated using trypsin treatment, and similar approaches work well for echinoderms e.g. adult ovaries. We have used 0.2% trypsin (use trypsin formulations made for tissue dissociations which are less pure, less expensive, and may have collagenase as well) on ovaries e.g. from the sea star *Patiria miniata*.
- viii. Resuspend trypsin to 0.2% in 0.5M NaCl. Dissect ovary and mince it in CaFSW into small pieces using a scalpel or small scissors. Allow the tissue to settle, remove the CaFSW and add the trypsin solution.
- ix. Incubate the minced tissue with the trypsin at room temperature for 20 minutes, with gentle mixing.
- x. Settle the tissue, and aspirate off the trypsin solution. Resuspend in 0.5 M NaCl and homogenize with a loose fitting Dounce homogenizer.
- xi. We have also used 0.5M urea (really) to dissociate embryos and ovaries. Make 1M urea in distilled water, and dilute 1:1 with CaFSW at 4°C. Add embryos or adult tissues to the urea mix for 10 minutes (4°C) and then replace the media with 0.5M NaCl. Homogenize gently with a Dounce homogenizer as described above.
- xii. Note: the reverse transcriptase and cDNA synthesis steps require magnesium ions. Therefore, do not use EDTA in the dissociation steps, only EGTA.

c. Single cell analysis from perturbation experiments:

A major value of single cell mRNA – seq is to test for changes in transcript profiles following a perturbation. Some perturbations may be highly impactful on certain lineages, whereas other cell types may be insensitive to the same perturbant. Whole embryo transcriptome analysis, or qPCR analysis from whole embryos would minimize these differences, whereas single cell mRNA-seq would likely reveal these differences and define the cells influenced the most. The perturbation may be from small molecule/drug inhibitors, from environmental changes of the culture condition, from morpholino antisense oligonucleotide (MASO) knock-downs, and/or from CRISPR/Cas9 gene targeting experiments. The conditions for dissociating embryos remain the same as described above. For small molecule inhibitor/drug experiments the numbers of embryos treated are often in vast excess (thousands, generating millions of single cells following dissociation) whereas microinjection experiments often instead yield hundreds of embryos. We found that having ~500 microinjected embryos meant that we would have ample single cells from dissociated embryos, depending on the stage of dissociation. The dissociation of small numbers of embryos is enhanced by observing the process under a stereo microscope, both the homogenization and resuspension process.

4. Summary of protocols for obtaining drop-seq emulsions

Bead counting

For the Nadia protocol, the first step is to order the barcoded beads from ChemGenes (Catalog number: Macosko-2011-10(V+)). These beads are expensive (\$3,524US for 10 umole), and it takes about a month to receive them after placing the order.

However, this amount of beads will be enough to process around 50 drop-seq experiments. When the beads are received, they need to be kept at 4°C until use. It is important to use low retention tips during the entire experiment to avoid losing beads, RNA or cDNA. The investigator needs to wash the beads one time with 1 ml of 100 % ethanol plus 1 ml of TE/Tween (10mM Tris pH8, 1mM EDTA, 0.01% Tween) in nuclease free water. This mix is transferred to a 50 ml falcon tube and spun for 1 minute at 1000 g at 4°C. The supernatant is discarded and the beads are washed 2 times with 30 ml of 100% ethanol (using the previous centrifuge settings), and 2 times with 30 ml of TE/Tween. After adding the TE/Tween for the second time, the beads are kept in this 30 ml suspension by either pipetting or inverting the tube, and 20 ul of this mix can be loaded to a Fuchs Rosenthal C-chip haemocytometer to count the beads. It's important to count at least 4 different regions of the chip to have a more accurate average (The chip contains 16 of these regions). As an example, users here obtained an average of 52 beads per region of the Fuchs Rosenthal. This number is multiplied by 16 to represent the number of beads in the entire Fuchs Rosenthal grid. This grid volume is equivalent to 3.2 ul. To obtain the number of beads per ul, the investigator needs to divide his total number per 3.2. Using the previous example: $52 \times 16 = 832$ beads in the grid, and $832 \text{ beads} / 3.2 \text{ ul} = 260 \text{ beads per ul}$. The stock of beads then contains 30 ml of a solution at a concentration of 260 beads per ul, meaning a total of 7.8 million of beads. This solution has to be aliquoted in Eppendorf tubes of 150 000 beads (number required for each single cell sample). In our example, the investigator will obtain 51 aliquots ($150\,000 \text{ beads} / 260 \text{ beads per ul} = 577 \text{ ul per aliquot}$). These aliquots can be stored at 4°C for several months.

For the 10x Chromium protocol, the beads come with the kit, ready to be used. They only need to be kept at room temperature for 30 minutes before loading the chip.

Bead preparation

For the Nadia, on the day of the experiment, the bead aliquot is centrifuged for 1 minute at 1000 g at 4°C. The supernatant is carefully removed (with a p200 followed by a p10 pipette) without touching the pellet, and the beads are then re-suspended in 250 ul of cold Lysis buffer. The Lysis buffer is described in the Nadia protocol

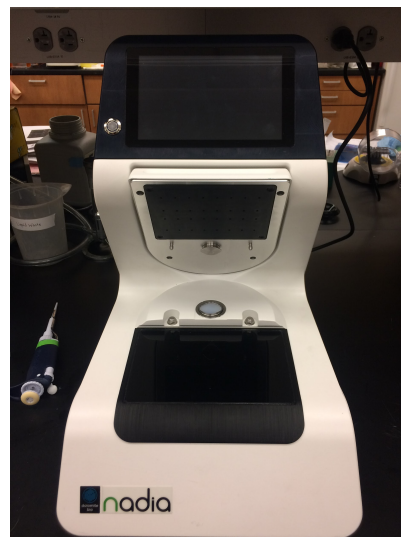
(<https://www.dolomite-bio.com/support/downloads/>). We recommend to make this buffer in advance in a large volume (e.g. 25ml), without DTT, filter it through 0.2 uM filter and store it at 4°C, it will be good to use for several weeks. The DTT needs to be added freshly on the day of the run.

Cell Encapsulation

The Nadia instrument (represented on the image below) is really easy to use . After turning on the device, the cartridge can be placed on the Nadia. These cartridges need to be bought separately, they are sold as either 1, 2, 4 or 8 identical chips per cartridge. Even though it seems more efficient to directly process 8 samples at once, it requires one to have all the dissociated cell samples on the same day, and more time will be spent in each step to process all of them at once. To improve the quality of the samples, the investigator should use the 1, 2 or 4 chip cartridges. It's also important to notice that these cartridges are expensive and cannot be reused. If the investigator only needs one chip in a 2-chip cartridge, the second chip is then lost. A set of cartridges for 8 samples, including 2 cartridges of 2 chips and one cartridge of 4 chips costs about 1,500 dollars. For the 10x Chromium, 8 chip cartridges are included in the

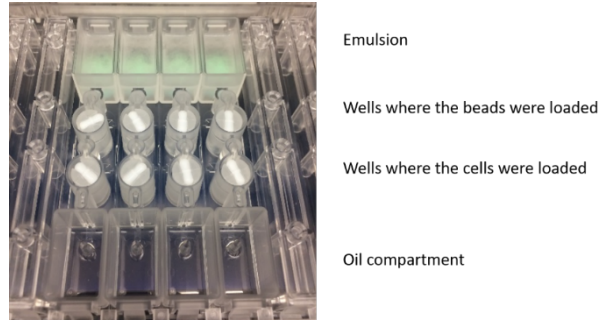
kit, but unused chips are lost if the investigator only needs to use 4 chips out of the 8 in one run for example.

For the Nadia, all the instructions appear on the screen at the required time. The first step is to add 3 ml of oil (Biorad 1864006) in the oil compartment. The Nadia will then cool down for a few minutes. Following the on-screen instructions, mix and load 250 μ l of beads (previously re-suspended in the lysis buffer) and 250 μ l of the cells (previously prepared) in their

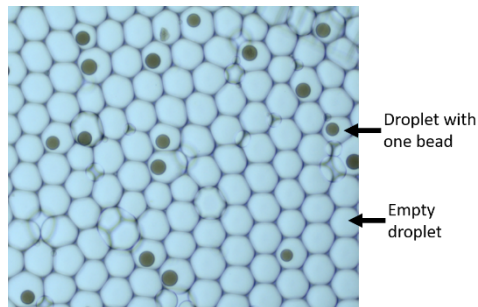


respective compartments (use the p200 and load 2 times 125 μ l for both the beads and the cells). The Nadia will form droplets for about 20 minutes at 6 °C, followed by 10 minutes at room temperature to improve cell lysis. The default parameters of the Nadia work perfectly for sea urchin and sea star dissociated cells. However, an additional device from Dolomite, the Nadia Innovate, enables the investigator to control diverse parameters (such as droplet size, droplet frequency, temperature,...) and to take live videos of the droplet formation. This device is especially helpful if a sample constantly results in an abnormal emulsion. The user will be able to change the parameters and observe the droplets in the same time. A run with the Nadia takes 28 minutes, with more cells run and an extended lysis step, whereas the 10x Chromium Controller takes 6.5 minutes for fewer cells and no additional lysis incubation.

At the end of the Nadia run, the chip will look like this: (example for a 4 chip cartridge).



The resulting emulsion should look dense and white with both instruments. With the



Nadia, a green light above the well will indicate to the user which compartment contains the emulsion. The investigator will discard the underlying oil (about 2 ml), and use a low retention tip (as a reminder) to pipette the remaining 1 ml that contains the droplets and transfer it to a low DNA binding tube. It is not a problem if some left over oil is also added to this tube. The beads will be washed in the following steps. Before processing with the samples, it is crucial to evaluate the quality of the emulsions under a microscope. 8.5 ul of each emulsion can be loaded into a Neubauer haemocytometer. Droplets and beads will be visible, but the cells have already been lysed in the Nadia. Sometimes, 2 beads can be found in one droplet but they should be a lot less abundant than the single bead droplet. This figure below is an example of the expected observation.

For both instruments, if the emulsion does not look dense or white, something wrong happened, and the samples need to be run again. The droplets are probably bigger than expected and contain many beads. If the emulsion looks good but in lower volume than expected, the system probably clogged during the run. The investigator can then decide to re-run the sample, or to process this sample normally especially if the samples are precious. The droplets are probably perfect, but less abundant than in a good run.

Bead isolation

With the Nadia, the next step is to extract the beads from the droplets. The emulsion is carefully added to the center of a falcon tube containing 30 ml of 6xSSC (Sigma 20x SSC, S6639-1L) diluted in nuclease free water. After addition of perfluorooctanol (Sigma 370533), the tube has to be manually shaken vigorously 3 to 4 times in order to break the droplets. This falcon tube is then centrifuged for 1 minute at 1000 g at 4°C and kept on ice. The isolated beads will be sitting on top of the oil fraction. The supernatant (SSC 6x) can be removed carefully using a 10 ml pipette with a pipette controller. It is better to leave a few ml on top of the beads instead of getting too close to the beads with the pipette. Using the same pipette controller, with a 25 ml pipette, 30 ml of SSC 6x can be strongly released in the same tube to re-suspend the beads. The oil will fraction itself back to the bottom of the tube. This entire bead suspension needs to be quickly pipetted back to a new 50 ml falcon tube using a 10 ml pipette. If this step is not performed fast enough, the beads will start to settle back on top of the oil. It is important to avoid pipetting oil in this step because it will prevent the beads from

forming a nice pellet in the next steps. This bead suspension can be centrifuged for 1 minute at 1000 g at 4°C to obtain a pellet of beads. After removing most of the supernatant, the beads are then transferred to a 1.5 ml low DNA bind tube and washed 2 times with SSC 6x (1 minute at 1000 g at 4°C). The beads can be gently re-suspended between these 2 washes, by pipetting them 5 times up and down. With the 10x Chromium, the emulsion is directly used for reverse transcription, without isolating the beads, following the 10X protocol.

Reverse transcription

For the Nadia, 300 ul of the 5x Reverse Transcription (RT) buffer are added to the pelleted beads to equilibrate them before starting the reaction. They can be re-suspended again by pipetting 5 times up and down before centrifugation. The Reverse Transcription Buffer (made according to the Nadia handbook) can be prepared during the droplet formation in the Nadia and kept on ice until this step is reached. All the components can be mixed and kept on ice, except the reverse transcriptase itself that needs to be added right before the reaction. The reverse transcription will require an incubation of 30 minutes at room temperature followed by 90 minutes at 42°C. It is important in both of these incubations to maintain the beads in suspension (if they settle down, this will prevent the complete reaction). It is recommended to use a disk-style or a rotisserie-style rotator. This protocol requires an unusual amount of enzyme, a regular tube of reverse transcriptase (ThermoFisher scientific, EP0751) will not even be enough for one reaction. We advise investigators to buy the large tube right away (EP0753) and to evaluate the left over volume before starting a new run on the Nadia. It is crucial to do the reverse transcription right after the run to avoid RNA degradation.

The investigator can then wash the beads according to the Nadia protocol and store the tubes at 4°C overnight.

For the 10X Chromium, the emulsions are incubated at 53°C for 45 minutes, and at 85°C for 5 minutes. The tubes can be stored at 4°C for up to 3 days, or -20°C for a week. In contrast to the Nadia, the bead isolation happens after the reverse transcription in this case. 125 ul of the recovery agent, used to disrupt the droplets is added to each tube. After 60 seconds, and without mixing, a biphasic mixture will appear. The pink fraction contains the recovery agent and the oil, the clear fraction contains the cDNA. After removing most of the pink fraction, without disrupting the clear fraction, 200 ul of Dynabeads Clean up mix are added to the clear fraction. The cDNA is then eluted according to the 10x Chromium protocol

(<https://www.10xgenomics.com/>).

Exonuclease treatment

This step is specific to the Nadia instrument. Some beads may have not been in contact with a cell in some droplets, and have not captured any RNA molecule. This exonuclease treatment will remove the primers on these empty beads. The beads are washed with 1 ml of 10 mM Tris pH 8, mixed by pipetting 5 times up and down and centrifuged for 1 minute at 1000 g at 4°C. The bead pellet is re-suspended in 200 ul of exonuclease mix: 22 ul of exonuclease I buffer, 11 ul of exonuclease (New England Biolabs, M0293L) and 187 ul of nuclease free water per sample. (Prepare the mix first in a new tube, the volumes indicated here assume 10% extra volume for pipetting losses. Then pipette 200 ul of this mix to the tube containing the beads). After 45

minutes at 37°C, the beads are washed one time with TE/SDS (10 mM Tris pH 8, 1 mM EDTA pH 8, 0.5% SDS) and one time with TE/TW.

5. Summary of protocols for making sequencing libraries

PCR

For the Nadia instrument, before making the cDNA library, the beads need to be washed 2 times with 1 ml of nuclease free water, and counted using a Fuchs-Rosenthal haemocytometer (as previously described in the bead counting section). The average number of beads per region (for example: 14) will be multiplied by 16 and divided by 3.2 to obtain the number of beads per ul (for example: $14 \times 16 / 3.2 = 70$). The beads were previously re-suspended in 1 ml of nuclease free water, giving a total number of beads of $70 \times 1,000 \text{ ul} = 70,000$ beads per ul in our example. The Nadia protocol recommends to adjust the bead concentration to 200 beads per ul and to use 10 ul of this mix (2,000 beads) to make the cDNA library for each sample. In this protocol, the left over (68,000) beads will be lost and rare cell types might also be lost if they were not represented in the chosen 2,000 bead aliquot. We suggest to amplify the entire amount of beads to obtain a more accurate representation of the sample. In this case, the beads can be diluted to 500 beads per ul instead of 200, resulting in 5,000 beads per PCR reaction. Instead of doing a single PCR per sample, we suggest to aliquot all the beads into PCR tubes (for example, $70,000 / 5,000 = 14$ PCRs for one sample). These PCRs will be pooled together in one single tube at the end of the reaction, following their purification with the AMPure XP beads (Beckman Coulter, A63880) as explained in the Nadia protocol.

For the 10X Chromium, the entire volume of purified cDNA (35 ul) is amplified in one single tube after addition of 65 ul of the Amplification reaction mix. It is important to use a Thermo cycler that can accommodate a reaction volume of 100 ul. This PCR is then washed with the SPRIselect reagent.

For both instruments, the quality of the amplified DNA need to be tested with the Fragment bioanalyzer (size and concentration of the DNA).

Tagmentation

The fragments resulting from the PCR are expected to be larger than 1kb, since they often represent the full length cDNA. However, these fragments cannot be directly sent for Illumina sequencing. They need to be cleaved in smaller fragments and tagged for sequencing purpose in order to be sequenced on an Illumina flow cell. Fragments larger than ~700 bps cause cluster overlap, and decreased signal, reducing the sequence effectiveness.

The Nadia protocol suggests using the Nextera XT kit (Illumina, FC-131-1024) for the tagmentation. Only 600 pg of full length cDNA will be tagmented for each sample. The Nextera XT kit relies on the transposase Tn5 that will cleave the DNA and tag it with adaptors. However, in our hands, this Nextera XT kit does not lead to a sufficient amount of DNA for sequencing purposes and it required trouble shooting each sample in order to obtain the correct fragmentation size. A new kit type is now available, the Nextera Flex (20018704); it delivers consistent insert sizes regardless of the DNA input amount or size. This new bead-based technology minimizes bias and opportunities for error, resulting in highly reproducible sequencing data. We recently tested this new Nextera Flex kit, and obtained perfect fragmentation for all our samples. Moreover, this

kit enables investigators to fragment larger amounts of full length cDNA (1 ng to more than 500 ng) and to obtain a reproducible size of fragments (~500 bps). We strongly recommend the use of the Nextera Flex kit since it also leads to tagmented libraries that are about 10 times more concentrated than with the Nextera XT kit.

The 10x Chromium samples are fragmented with a specific fragmentation buffer included in the kit. In contrast to the Nadia, here almost the entire sample is tagmented (35 ul out of the 40 ul eluted in the previous step). This fragmentation step is followed by a purification with the SPRIselect reagent. The adaptors required for the Illumina sequencing are then ligated to the fragments. These ligated fragments are purified once again with the SPRI reagent.

Sample Indexing

Before sending the samples for sequencing, they need to be indexed. These indexes are especially important if the investigator decides to run several libraries in the same sequencing lane. The samples that will be run together need to have a different index, and the samples run separately can have the same index. The indexes will be used to identify the origin of the transcripts. For the Nadia, the samples can be indexed using the Nextera XT index kit (Illumina, FC-131-1001) if the libraries were made using the Nextera XT kit or using the Nextera DNA CD index kit if the Nextera Flex kit was used (Illumina, 20018707). It is not recommended to use XT indexes with Flex libraries as Nextera DNA CD kit has indexes of higher purity that have been carefully optimized for improved performance. The indexes are added to the tagmented libraries by PCR. With the Nextera XT index kit (presented in the Nadia handbook), only one variable

index (N7XX) is added to each library. The second one is constant in each library (using the New P5 smart PCR hybrid oligo). Each PCR is purified with the previously used AMPure XP beads and analyzed on the Fragment bioanalyzer. With the Nextera DNA CD kit, two variable indexes are added by PCR; the indexed DNA is then purified using the purification beads provided in the kit.

For the 10x Chromium, the reagents are provided in the kit. Only one variable index is also added in each library (I7 index). Each PCR is purified with the previously used SPRIselect reagent and analyzed on the Fragment bioanalyzer. In each case, the fragments should now be around 500 bp.

The 10x Chromium offers an additional step to better evaluate the library concentration. In addition to the concentration obtained with the Fragment bioanalyzer, the investigator can also evaluate the library concentrations using qPCR (<https://www.10xgenomics.com/>).

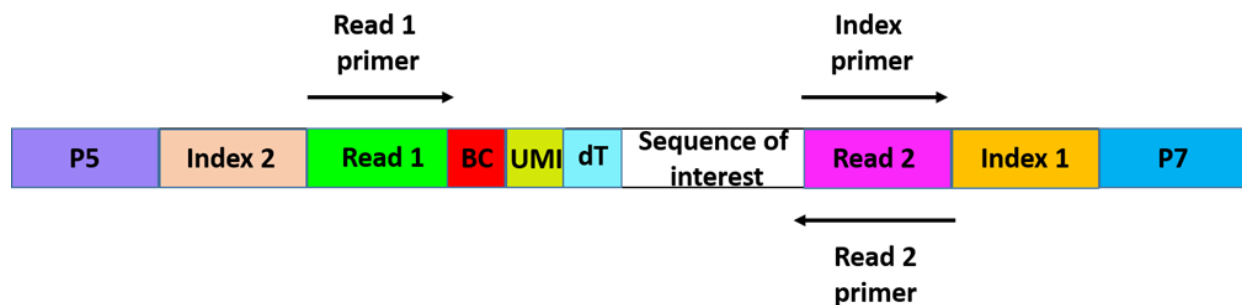
Sequencing

Regardless of the construction methods described above, the libraries contain the sequence of interest surrounded by adapter constructs on each side. These adapters have the flow cell binding sites, P5 and P7, to allow the library to attach to the Illumina flow cell surface. A primer complementary to Read 1 will allow the sequencing of the bead barcode (BC), as well as the UMI (Unique Molecular Identifier). When Read 1 is finished, everything from Read 1 is removed and an index primer is added. A primer complementary to Read 2 is then used to sequence the cDNA fragment. Paired end

sequencing is essential to obtain these sequences: 26bp read1, 8bp sample index, 98bp read2.

Illumina generally asks libraries to be submitted at 15 nM in 20 ul for the HiSeq2500 instrument. Experiments looking to get an in-depth view of the transcriptome will require about 200 million reads per sample.

Example of a single cell library:



6. Overview of procedure pipelines for analysis of results

Cell Ranger R Kit (10X Genomics resource)

10X Genomics offers several software pipelines for analysis of sequencing data. The Cell Ranger Single Cell Software Suite 2.1.0, run on Linux, demultiplexes the Illumina base call files (BCL), the file containing the nucleotide base calls for each sequencing cycle, into FastQ files, which contain both the base calls and a quality measure for all reads. Cell Ranger also performs UMI counting. In our case, Cell Ranger was also used to generate the *S.purpuratus* reference genome. Secondary analysis of gene expression data can be accomplished using Cell Ranger or using R or Python.

Clustering analysis for these samples was accomplished using Seurat, an R package for scRNA-seq analysis developed by the Satija Lab [<https://satijalab.org/>].

The first step in the Cell Ranger analysis pipeline is *cellranger mkfastq*, this pipeline produces your FASTQ files. Cellranger count performs genome alignment and produces UMI counts in the form of a matrix, this is done individually for each sample. Cellranger aggr combines the matrices from individual runs of cellranger count to normalize the data based on sequencing depth. This is an essential step in creating a gene-barcode matrix for an entire experiment.

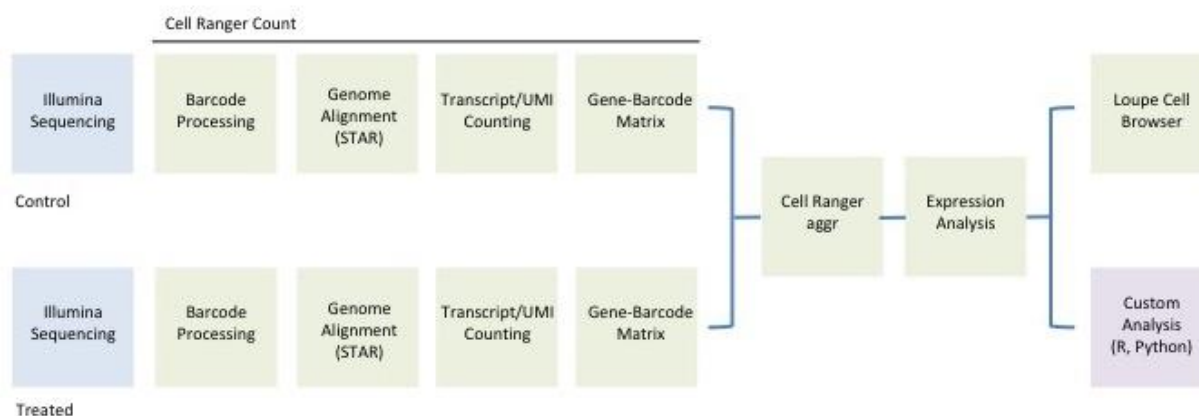
The gene-barcode matrix file produced can be uploaded to R for downstream analysis. The first column in the matrix is the list of genes that were sequenced and mapped. In the row for one gene, the information that follows is the barcode that was linked to that gene transcript.

LOC585686 TTTGGTTAGAACTCGG TTTGGTTAGCTCAACT
TTTGGTTAGTACGCGA

LOC764781 TTTGGTTAGTGACATA TTTGGTTGTACCAGTT TTTGGTTGTATAGTAG

LOC579747 TTTGGTTTCACCCTCA TTTGGTTTCCAAATGC
TTTGGTTTCCACGACG

Cellranger reanalyze is a pipeline that allows you to change the default parameters for visualizing the gene expression data of your experiment. For example, here is a pipeline to compare gene expression between a control sample and a treated sample:



Expression Analysis Using Seurat

Loupe Cell Browser, a 10X Genomics application, is used to visualize the clustering and gene expression data output from running the Cell Ranger pipelines. Cell Ranger pipeline output files contain results for t-SNE projection, principal component analysis (PCA) and k means clustering. K means clustering assigns cells to k number of groups based on their characteristics. This form of clustering is helpful when the identity of your cell types, in terms of gene expression, is unknown. The 10X Genomics website offers a tutorial (PBMC data), which walks you through seeing the pipeline results in R: <http://cf.10xgenomics.com/supp/cell-exp/cellrangerrkit-PBMC-vignette-knitr-2.0.0.pdf>.

The Satija Lab offers different tutorials for the Seurat R package based on the nature of your experiment. The tutorial titled, “Integrating stimulated versus control PBMC datasets to learn cell type specific responses”

(https://satijalab.org/seurat/immune_alignment.html) matches the sea urchin

experiment best when the goal is to analyze changes in particular cell types between the control and treated samples or compare gene expression changes in particular cell types over time. We used this code for our analysis with a couple of adjustments to cater to our model organism.

The first step in the analysis using Seurat is quality control, or QC. The QC step includes filtering cells and normalizing the data. In this line of code, the cells within the 10X data stored in the variable “C1” are filtered based on numbers of genes.

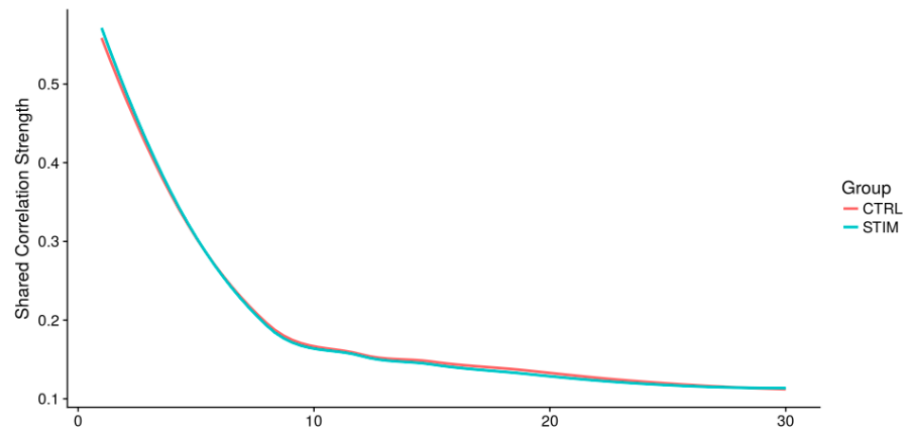
```
C1 <- FilterCells(C1, subset.names = "nGene", low.thresholds = 400, high.thresholds = 2500)
```

As part of the QC, we filtered cells by setting a low threshold of 400 genes and a high threshold of 2500 genes. These thresholds are set to exclude droplets where a complete cell was not captured and multiplets, where more than one cell lysed in a droplet would present as having many more genes.

Following the given code, we are able to find variable genes in the dataset and employ the RunTSNE command. Creating a principal component analysis plot is common for analysis of single cell sequencing data. The use of canonical correlation analysis (CCA) is similar to the of principal component analysis, both of these analyses look for sources of variation between your control and experimental datasets. The Seurat tutorial includes a CCA step but others may make use of PCA plots.

Canonical correlation analysis subsets cells based on correlation in the gene expression data and produces many correlation values numbered CC1, CC2... We next choose CC's for downstream analysis like clustering by looking at the plot of

shared correlation strength to identify the point at which the correlation strength plateaus.



In the Seurat tutorial, based on this graph, CC1-20 are chosen for analysis.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) plots are based on variation in the sequencing data with the x and y axis being responsible for the greatest source of variation. You can imagine a case where this would be one experimental condition against the control. PCA is originally plotted in 3D.

There are numerous resources for understanding PCA plots, an interactive example may be helpful in seeing how variability in the dataset affects a 3D plot. Analyzing a 3D plot poses its own challenges, however, and reducing the data to two dimensions can help with visualization.

Interactive PCA example online: <http://setosa.io/ev/principal-component-analysis/>

After choosing the appropriate correlation values for analysis, the tutorial walks you through generating t-distributed stochastic neighbor embedding plots.

T-Distributed Stochastic Neighbor Embedding

T-Distributed Stochastic Neighbor Embedding (t-SNE) plots allow you to visualize patterns in scRNA-seq data in two dimensions. Each point on the plot is representative of one cell, defined by two t-SNE components. Proximity on the two dimensional plot corresponds to similarity in the gene expression profiles of the cells. Clusters therefore correspond to a group of cells with similar expression and likely function. The identity of the clusters must be drawn from the unique gene profiles of the clusters. The expression of marker genes can also be shown on a t-SNE plot—this can be shown as expression greater than zero or as a range.

The *FindClusters* command, which is employed after the *RunTSNE* command, has a “resolution” parameter. Setting resolution to a value greater than 1 will give more clusters in the t-SNE plot, and a value below 1 will give fewer. While this can be modulated, the goal is to visualize distinct, unique cell clusters.

Once the clusters are visualized on a t-SNE plot, you can extract what genes are conserved and differentially expressed in the clusters. The job of identifying cell types either based on the presence of particular genes, sets of genes, or overrepresentation analysis begins here.

Gene Ontology

Gene Ontology (GO) term enrichment is used to interpret sets of genes based on annotations for the reference gene set. Once you have compiled your list of conserved

and differentially expressed genes for each cluster, you can use this list for overrepresentation analysis. The *S.purpuratus* reference is present on the GO website.

There is also information on how to interpret results on the GO website:

<http://geneontology.org/page/go-enrichment-analysis>.

Definitions:

Sensitivity: ability to isolate an mRNA transcript from an individual cell and incorporate it to a cDNA library

Accuracy: refers to the read quantification reflecting the concentration of extracted mRNA

Multiplet: microdroplet with more than one cell or more than one droplet

GEM: gel bead in emulsion

7. Alternative single cell technologies

The technology in single cell RNA sequencing is changing rapidly. On the horizon in mid-2018 is sequencing of RNA from nuclei (not cytoplasmic) using drop-seq technologies, so called Single nuclei RNA-sequencing (DRONC-seq). This is particularly useful for tissues that are extremely hard to dissociate e.g. mammalian brain tissue, in which the cells are lysed or damaged in the dissociation process. This is less essential for cells that instead dissociate readily as in echinoderms, but the technology could help investigators distinguish between mature mRNA and newly transcribed RNAs for example, or in characterizing tissues from adult e.g. tube feet that might be difficult to dissociate. The 10x company also announced recently an effort for

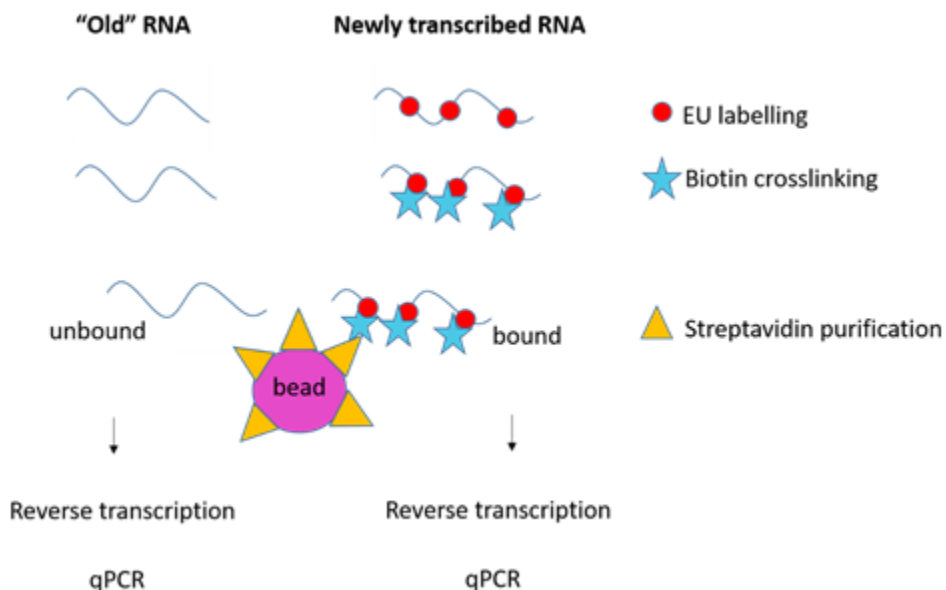
single cell ATAC-seq, to identify open chromatin sites in genomic DNA on a per cell basis. Such an experiment could reveal the heterogeneity in cellular function or fate, even before the phenotype was apparent. Hopefully the echinoderm community is quickly able to make use of these approaches to study important biological functions, and to take advantage of the many wonderful characteristics that these embryos and animals have to offer.

B. Nascent RNA capture

Many procedures for studying embryonic development rely on testing steady state levels of mRNAs. That is, we normally measure the balance of mRNA synthesis with mRNA degradation. These assays usually consist of in situ hybridization, qPCR, RNA seq, or single cell transcriptome analysis for example. Each mRNA, however, has its own dynamics. Some mRNAs have a rapid turnover, so that the investigator is measuring genes recently transcribed, whereas other mRNAs can be prolonged in their half-life, misleading the investigator into thinking the gene was transcriptionally active at the time of the measurement. The protocol here documents how investigators test for newly transcribed mRNAs and how to distinguish them from aged, long-lived transcripts.

The Click-iT Nascent RNA Capture Kit (ThermoFisher scientific, C10365) employs a modified nucleotide that is incorporated into growing, nascent mRNAs. The nucleotide analog is tagged so that the investigator can then measure how much transcript was made during the incubation window of time. RNA expression also changes with specific perturbations (protein inhibitors, morpholinos, CRISPR, etc.), this nascent RNA

procedure is used to determine the effects of these perturbations on specific, or general transcriptional activity. Briefly, the newly transcribed RNAs are labelled with EU, crosslinked to Biotin using the Click-It technology, purified with Streptavidin beads, reverse transcribed and tested by qPCR, or RNA-seq (Figure XX).



Protocol:

1. The stock of EU is re-suspended in DMSO at a final concentration of 200 mM according to the manufacturer's instructions.
2. We suggest to perform these incubations in small volumes e.g. 12 well plates (final volume of 2 ml per well) to reduce the volume of EU used. This reagent is costly. We recommend to use more than 250 embryos per well.
3. Add the EU to a final concentration of 1mM EU for desired time: 5 hours to several days (Fresques et al, 2018). This time will depend on the abundance of the transcript of interest and the developmental time point(s) that are being analyzed. It is important to note that the shorter the incubation, the better the

temporal resolution, but the lower the signal of labeled mRNAs. Embryos do have a large amount of free nucleotides within the cytoplasm (Davidson, 1976) and use of the analog is concentration and time dependent.

4. At the end of the labeling period, the embryos are transferred into Eppendorf tubes and centrifuged for 1 minute at 3000 rpm. Keep tubes and wash solutions on ice to reduce further transcriptional activity and swimming of the embryos back into the supernatant. After the centrifugation, the supernatant is removed and the embryos are re-suspended in 75 μ l of RLT buffer provided in the Rneasy micro Kit (Qiagen, 74004). The samples can be stored at -80°C or the RNA can be extracted right away following this Rneasy micro kit protocol. The purified RNA can be eluted in 20 μ l of nuclease free water, nanodropped and stored at -20°C. Unincorporated nucleotides are washed out during the RNA isolation procedure so free nucleotides are not a complication in the labeling procedure.
5. One might be interested in measuring the half-life of various transcripts that were transcribed during the labeling period. In that case, after the pulse period of labeling, wash the embryos at least 3 times and add 10mM buffered UTP for the duration of the incubation to compete effectively with residual nucleotide analog.
6. Crosslink the Biotin Azide reagent in the kit to the nascent EU labelled RNA using the click-iT reaction buffers provided. Prepare the Click-iT reaction cocktail according to the protocol. The concentration of Biotin Azide depends on the

concentration of the purified RNA as explained in the manufacturer's protocol.

As an example, here is the protocol that we used for most of the samples:

25 ul of Click-iT EU buffer,

4 ul of CuSO₄,

2.5 ul of Biotin azide (0.5 mM final),

15.75 ul of EU labelled RNA (at 100 ng/ul corresponding to a total amount of 1.575 ug),

1.25 ul of Click-iT reaction buffer additive 1,

1.5 ul of Click-iT reaction buffer additive 2.

7. The mix will turn dark brown after addition of the buffer additive 2. The reaction is performed for 30 minutes at room temperature under gentle vortexing. We recommend using a Thermo mixer with the following settings: 350 rpm at 22°C. The RNA (Biotin EU labelled and unlabeled RNA) can then be precipitated and re-suspended in 50 ul of nuclease free water, following the protocol given. Using the sea star larvae , we obtained about 30 ng/ul.
8. The Biotin EU labelled RNA then is purified with 15 ul of Dynabeads MyOne Streptavidin T1 magnetic beads (ThermoFisher Scientific, 65601) for each sample. If the investigator has several samples, we recommend washing the total amount of beads (150 ul for 10 samples) all together before aliquoting them into separate reactions. For each sample, the RNA binding mix is prepared as follows:

49 ul of Click-iT RNA binding buffer 2x,

2 ul of RNaseOut Recombinant ribonuclease inhibitor,

49 ul of purified RNA (it was previously re-suspended in 50 ul but 1 ul was used for nanodrop).

9. This RNA binding mix is heated at 68 to 70°C for 5 minutes before adding 15 ul of bead suspension in each reaction. The binding reaction is performed in the Thermo mixer using the previous parameters (30 minutes at room temperature, ~ 22°C).

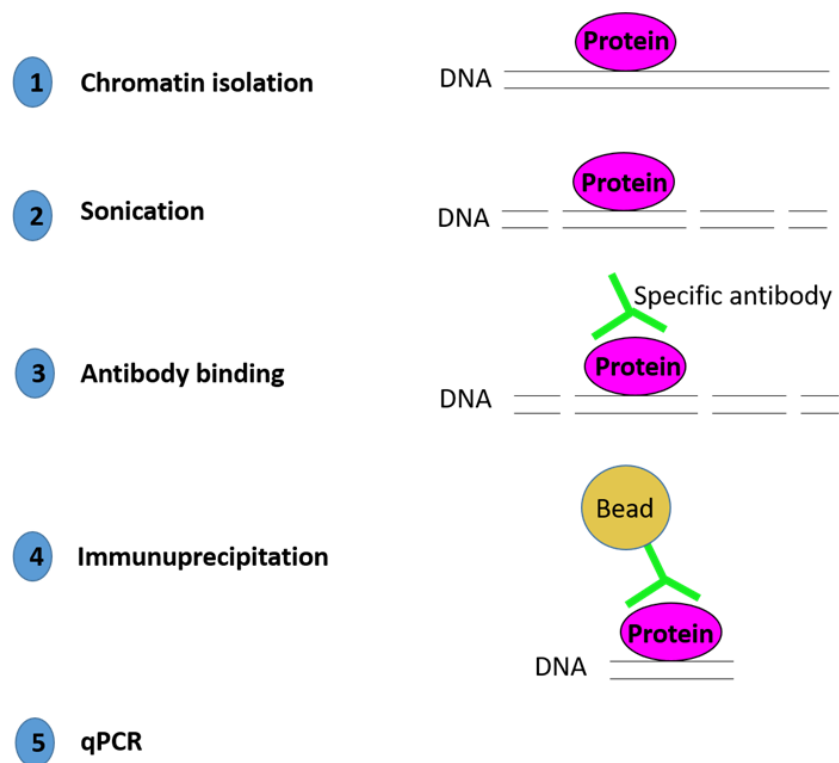
10. The beads are immobilized using the tube magnet. We recommend to keep the unbound fraction (approximately 100 ul) before starting the wash, it represents the RNAs that were already present in the embryos but that were not being transcribed during the time of the EU incubation.

11. The beads are washed according to the protocol, and re-suspended in 15 ul of the Click-iT reaction wash buffer 2. We suggest to precipitate the 100 ul of unbound fraction (using the previous protocol: glycogen, ammonium acetate and ethanol) and re-suspend it in the same volume as the labelled RNA: 15 ul of nuclease free water.

Both fractions need to be reverse transcribed right away to avoid RNA degradation. We routinely use the Maxima first strand cDNA synthesis kit (ThermoFisher Scientific, K1641) for reverse transcription. Specific transcripts can then be analyzed in each fraction, using qPCR. For example, we use Maxima SYBR Green/ROX 2X qPCR Master Mix (ThermoFisher Scientific, FERK0222).

C. Chromatin Immunoprecipitation

Chromatin Immunoprecipitation (ChIP) is a powerful method to test specific protein - DNA interaction. Briefly, the chromatin is isolated and sonicated to shear the DNA in fragments of approximately 200 bp. An antibody against your favorite DNA binding protein is used to immunoprecipitate the DNA-protein complex. After purification of this complex with magnetic beads, the bound DNA is analysed by qPCR for promoter specific studies or by ChIP-seq to identify new targets. Here, we focus on the ChIP protocol followed by specific qPCR analysis.



Cell dissociation and Fixation (based on a protocol performed on *Drosophila* (Larschan et al,2007))

1. The embryos are cultured in 50 ml falcon tubes (0.25% v/v, meaning 125 ul of pelleted embryos in 50 ml) until the desired developmental time point.

2. After centrifugation, wash the pelleted embryos 3 times with 25 ml of calcium free sea water (CaFSW, see below).
3. Re-suspend the embryos in 10 ml of dissociation buffer (1M Glycine, 25mM EDTA, pH 8.0) and keep them on ice for 10 minutes (Juliano et al, 2014), disaggregate them by pipetting a few times up and down.
4. Wash the dissociated cells 3 times with 5 ml of calcium free sea water.
5. Fix these 5 ml of dissociated cells by adding 5 ml of formaldehyde 2% (stock made in calcium free sea water), to obtain a final concentration of 1%.
6. Rotate the sample at room temperature for 10 minutes.
7. Add glycine to a final concentration of 125 mM, and rotate at room temperature for 5 minutes.
8. Transfer the cells to a 15 ml falcon tube and keep it on ice.
9. Spin at 1500g for 5 minutes at 4°C.
10. Discard the supernatant, and re-suspend the pellet in 5 ml of PBS- 0.5mM EDTA pH 7.2 + 5 ul of PMSF 0.2 M
11. Spin at 1500g for 5 minutes at 4°C.
12. Discard the supernatant, and re-suspend the pellet in 1 ml of ChIP wash buffer A.
13. Transfer the sample to a 1.5 ml Eppendorf tube.
14. Rotate the tube for 10 minutes at 4°C.
15. Spin at 1500g for 5 minutes at 4°C.

16. Discard the supernatant, and re-suspend the pellet in 1 ml of ChIP wash buffer B

17. Rotate the tube for 5 minutes at 4°C.

Cell counting

18. Dilute 10 ul of the sample with 90 ul of ChIP buffer B. Continue to rotate at 4°C while counting.

19. Count the cells using a Haemocytometer.

20. Spin the rest of the sample (tube that was rotating), at 1500g for 5 minutes at 4°C.

21. Remove the supernatant entirely.

22. Freeze the pellet in liquid nitrogen, and store it at - 80°C.

Cell lysis and sonication

23. Re-suspend the pellet in 300 ul of ChIP Lysis buffer.

24. Lyse the cells by rotating at 4°C for 10 minutes.

25. Transfer the sample to a Protein low Bind tube.

26. Sonicate with the Bioruptor: 6 rounds of 5 minutes (each round is programmed to alternate between 30 seconds on and 30 seconds off). Add ice in the Bioruptor after each round.

27. Spin at 13,000 rpm for 10 minutes at 4°C.

28. Transfer supernatant to a 15 ml Falcon tube, and add 9 volumes of ChIP dilution buffer (for example, add 2.7 ml of buffer in 300 ul of sample)

29. Remove 400 ul for DNA input, transfer it to a new Protein Low Bind tube.

30. The rest of the sample (2.6 ml) is aliquoted into 3 Protein Low Bind tubes (860 ul per tube) and frozen in liquid nitrogen. Store the samples at - 80°C.

Input DNA Clean-up

31. To the 400 ul of input DNA, add

21.5 ul of 20% SDS

15 ul of 5M NaCl,

1 ul of RNase A (10/mg/ml stock)

32. Reverse crosslink the DNA at 65°C (hot plate) overnight

33. On the next day, add 1 ul of RNase A and incubate at 37°C for 30 minutes.

34. Add 20 ul of 1M Tris-HCL pH 6.8

10 ul of 0.5 M EDTA pH 8

3 ul of 20 mg/ml proteinase K

35. Incubate at 42°C for 90 minutes

36. Perform phenol chloroform isoamyl alcohol extraction

Add 500 ul of phenol chloroform isoamyl alcohol to the sample and shake gently
(no vortex)

Spin at 14,000 rpm for 5 minutes at room temperature.

Save the supernatant in a new tube.

37. Repeat the extraction with chloroform only

38. Precipitate the DNA (about 500 ul of supernatant) with ethanol, add:

10 ul of glycogen

1/10 volume of 3M NaOAc (about 50 ul)

2 volumes of 100 % ethanol (about 1 ml)

39. Incubate at - 80°C overnight.
40. Spin at 14,000 rpm for 10 minutes at 4°C.
41. Discard the supernatant and wash the pellet with 1 ml of fresh cold 70% ethanol
42. Spin at 14,000 rpm for 10 minutes at 4°C.
43. Discard the supernatant and spin down again for 1 minute to remove the residual ethanol.
44. Open lid and leave the tube at room temperature for 1 minute to dry the pellet.
45. Re-suspend the pellet in 100 ul of ddH₂O
46. Nanodrop the sample
47. Run 500 ng of the sample on a 2% agarose gel. The DNA should form a smear around 200 bp.
48. Store this input DNA sample at - 80°C, it will be used to normalize the qPCR data.

Immunoprecipitation

49. Thaw the tubes of diluted chromatin prep on ice
50. Add your specific antibody at appropriate dilution (usually 1/1000 to 1/500) in one tube

Add control IgG to a second tube (same concentration as your specific antibody)

(if you use a rabbit affinity purified antibody, use rabbit affinity purified IgG as a control)

(your third aliquot can be used to test a second specific antibody in the same experiment)
51. Rotate overnight at 4°C.

52. The next day, add 50 ul of magnetic protein A dynabeads to each IP sample
(optional: pre wash the beads with 500 ul of ChIP dilution buffer before adding them to the samples)
53. Rotate for 2 hours at 4°C.
54. Use the magnet to immobilize the beads and discard the supernatant between each of the following washes.
55. Wash 2 times with 1 ml of cold RIPA 150. Rotate at 4°C for 5 minutes between each wash.
56. Wash 1 time with 1 ml of cold RIPA 300. Rotate at 4°C for 5 minutes.
57. Wash 1 time with 1 ml of cold LiCl/TE buffer. Rotate at 4°C for 5 minutes.
58. Wash 2 times with 1 ml of cold TE buffer. Rotate at 4°C for 5 minutes between each wash.
59. To elute, add 250 ul of Sodium bicarbonate buffer to the beads. Rotate at room temperature for 15 minutes and transfer the supernatant to a new Protein low bind tube.
60. Repeat the elution in the same volume, and add the supernatant to the same tube.
61. Add 20 ul of 5M NaCl
62. Incubate the samples at 65°C overnight.

IP DNA Clean-up

63. On the next day, add 1 ul of RNase A and incubate at 37°C for 30 minutes.
64. Add 20 ul of 1M Tris-HCL pH 6.8
10 ul of 0.5 M EDTA pH 8

3 ul of 20 mg/ml proteinase K

65. Incubate at 42°C for 90 minutes

66. Perform phenol chloroform isoamyl alcohol extraction:

Add 500 ul of phenol chloroform isoamyl alcohol to the sample and shake gently
(no vortex)

Spin at 14,000 rpm for 5 minutes at room temperature.

Save the supernatant in a new tube.

67. Repeat the extraction with chloroform only

68. Precipitate the DNA (about 500 ul of supernatant) with ethanol, add:

10 ul of glycogen

1/10 volume of 3M NaOAc (about 50 ul)

2 volumes of 100 % ethanol (about 1 ml)

69. Incubate at - 80°C overnight.

70. Spin at 14,000 rpm for 10 minutes at 4°C.

71. Discard the supernatant and wash the pellet with 1 ml of fresh cold 70% ethanol

72. Spin at 14,000 rpm for 10 minutes at 4°C.

73. Discard the supernatant and spin down again for 1 minute to remove the
residual ethanol.

74. Open lid and leave the tube at room temperature for 1 minute to dry the pellet.

75. Re-suspend the pellet in 100 ul of ddH₂O

76. Nanodrop the samples

qPCR

77. Dilute the input DNA to 2 ng/ul and use 1 ul per qPCR control reaction

78. Use 1 ul of immune-precipitated DNA per qPCR ChIP reaction

79. All qPCR reactions are run in triplicate using the Maxima SYBR green reagent.

80. Data is analyzed using the percent input method.

Buffers

ChIP Buffers

The following buffers can be made in advance in 50 ml falcon tubes and filtered through sterile 50 ml disposable vacuum filtration system 0.22 (Millipore, SCGP00525). They can be kept at 4°C for several weeks. The 1x protease inhibitors, and the PMSF need to be added freshly on the day of the experiment.

0.2 M PMSF

0.174g PMSF

5 ml Isopropanol

ChIP wash buffer A

10 mM Hepes pH 7.6

10 mM EDTA pH 8

0.5 mM EGTA pH 8

0.25 % Triton X-100

Filter sterilize

1x protease inhibitor

0.2 mM PMSF

ChIP wash buffer B

10 mM Hepes pH 7.6

100 mM NaCl

1 mM EDTA pH 8

0.5 mM EGTA pH 8

0.01% Triton X-100

Filter sterilize

1x protease inhibitor

0.2 mM PMSF

ChIP lysis buffer

0.1% SDS

10 mM EDTA pH 8

50 mM Tris-HCl pH 8

Filter sterilize

1x protease inhibitor

0.2 mM PMSF

ChIP dilution buffer

0.01% SDS

1.2 mM EDTA pH 8

16.7 mM Tris-HCl pH 8

1.1% Triton X-100

167 mM NaCl

Filter sterilize

1x protease inhibitor

0.2 mM PMSF

ChIP TE wash buffer

10 mM Tris-HCl pH 8

1 mM EDTA

0.01% SDS

Filter sterilize

1x protease inhibitor

0.2 mM PMSF

RIPA 300 buffer

50 mM Tris-HCl pH 8

1% NP-40

2 mM EDTA

0.1% Sodium deoxycholate

0.1% SDS

300 mM NaCl

Filter sterilize

RIPA 150 buffer

50 mM Tris-HCl pH 8

1% NP-40

2 mM EDTA

0.1% Sodium deoxycholate

0.1% SDS

150 mM NaCl

Filter sterilize

1x protease inhibitor

0.2 mM PMSF

LiCl/TE wash buffer

0.25 M LiCl

1% NP-40

1% Sodium deoxycholate

10 mM Tris-HCl pH 8

1 mM EDTA pH 8

Sodium bicarbonate elution buffer

1% SDS

0.1 M Sodium bicarbonate

3M NaOAc (Sodium acetate)

20.4g Sodium acetate

to 50 ml ddH₂O

Adjust pH to 5.2 with glacial Acetic acid

Other Buffers

Hyalin Extraction Media (for 1 liter)

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

Calcium-free Seawater (for 1 liter)

NaCl 26.5 gm

KCl 0.7 gm

MgSO₄ · 7H₂O 11.9 gm

NaHCO₃ 0.5 gm

pH to 8.0, salinity should be ~34ppt

Conclusions:

Remarkably sensitive and creative new technologies allow us to see more focused and detailed aspects of gene regulation. As new and rapidly changing as these approaches are, the technologies are robust and reproducible. The three approaches we document

here in concert allow greater confidence and insight into how a gene regulatory network functions and how it may have evolved. We do end though with the thought these protocols and experimental approaches are only as good as what they tell us about developmental mechanisms. Left to our creativity is how we can incorporate these results into a dynamic, living, and ever changing embryo.

References

- Davidson, EH 1976. Gene activity in early development, Academic press.
- Fresques, T.M. and G.M. Wessel, *Nodal induces sequential restriction of germ cell factors during primordial germ cell specification*. Development, 2018. **145**(2).
- Haque, A., Engel, J., Teichmann, S.A., Lonnberg, T., 2017. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med 9, 75.
- Juliano, C.E. and G.M. Wessel, An evolutionary transition of Vasa regulation in echinoderms. Evol Dev, 2009. 11(5): p. 560-73.
- Juliano, C.E., Swartz, S.Z., Wessel, G.M. Isolating specific embryonic cells of the sea urchin by FACS. Methods Mol Biol. 2014; 1128: 187–196.

websites:

<https://www.dolomite-bio.com/support/downloads/>
<https://satijalab.org/>

Appendix B: Regulation of dynamic pigment cell states at single-cell resolution

Perillo M., Oulhen N., **Foster S.**, Spurrell M., Wessel G.M. Regulation of dynamic pigment cell states at single cell level. *Elife* 2020, 9.

This study closely followed our single-cell RNAseq analysis of early sea urchin embryos (Chapter 3). The datasets I produced and analyzed for Chapter 3 were further analyzed and joined with new datasets of later stage embryos to analyze pigment cell development specifically. For this study, I contributed the additional single cell RNA-seq experiments and analysis with support from Nathalie Oulhen.

Abstract

Cells bearing pigment have diverse roles and are often under strict evolutionary selection. Here, we explore the regulation of pigmented cells in the purple sea urchin *Strongylocentrotus purpuratus*, an emerging model for diverse pigment function. We took advantage of single cell RNA-seq (scRNAseq) technology and discovered that pigment cells in the embryo segregated into two distinct populations, a mitotic cluster and a post-mitotic cluster. *Gcm* is essential for expression of several genes important for pigment function, but is only transiently expressed in these cells. We discovered unique genes expressed by pigment cells and test their expression with double fluorescence in situ hybridization. These genes include new members of the *fmo* family that are expressed selectively in pigment cells of the embryonic and in the coelomic cells of the adult - both cell-types having immune functions. Overall, this study identifies nodes of molecular intersection ripe for change by selective evolutionary pressures.

Introduction

Biologically crafted pigments are nearly ubiquitous in nature. Pigmentation can reveal identity, protection, and even have antimicrobial functions. Pigmentation in an individual or in a species can change annually, or even instantaneously for camouflage for example arctic hares or cephalopod mollusks and chameleons respectively, making pigmentation even more dynamic in response to the environment. Echinoderms, a phylum of marine organisms, display diverse pigmentation schemes in embryos, larvae, and adults. From an intense red sea star (*Fromia milleporella*) to a black and white striped brittle star (*Ophiactis savignyi*) to the variegated sea urchin (*Lytechinus variegatus*), pigment is a pervasive feature of this phylum. Because of the molecular techniques now available for echinoderms, mechanisms controlling pigmentation in these animals are being revealed (Hira et al., 2020; Liu et al., 2019; Wessel et al., 2020; Yaguchi et al., 2020). The purple sea urchin, *Strongylocentrotus purpuratus*, provides a molecularly tractable model organism to dissect the developmental importance of pigmented cells, and the biosynthesis of their pigment. The adult of this echinoderm is intensely dark purple as a result of a family of polyketides, including echinochrome and various related spinochromes (Anderson et al., 1969). Furthermore, like most sea urchins, *S. purpuratus* larvae are pigmented due to the accumulation of a red/orange pigment in single cells embedded in, and scattered throughout, the aboral ectodermal layer (Gibson and Burke, 1985; Griffiths, 1965; Gustafson and Wolpert, 1967; Kominami et al., 2001; McClendon, 1912). This pigment is a naphthoquinone called echinochrome A, which accumulates in the pigment cell precursors during gastrulation in *S. purpuratus* (Calestani et al., 2003; Griffiths, 1965; Kuhn and Wallenfels, 1940; Oulhen and Wessel,

2016). Mutations that affect the pigmentation pathway lead to albinism (Calestani et al., 2003; Oulhen and Wessel, 2016; Wessel et al., 2020), and adult sea urchins that lack pigments are less resistant to environmental challenges (Wessel et al., 2020). The functional relationship between these larval and adult pigments and associated cells, and whether their biosynthetic pathways are similar, are open questions.

A known function of sea urchin larval pigment cells includes an essential role in the innate immune defense system (Buckley and Rast, 2017; Ch Ho et al., 2016; Hibino et al., 2006; Hira et al., 2020; Kiselev et al., 2013; Ransick and Davidson, 2006; Schrankel et al., 2016; Solek et al., 2013). When larvae are exposed to bacteria, pigment cells migrate from the ectoderm to the gut, a site for invading microbes, and interact with other immune cells (Ch Ho et al., 2016). This cell-cell interaction is at least in part regulated by IL17 cytokine (Buckley and Rast, 2017). Echinochrome A is also present in eggs of certain sea urchin species, in immune cells of the coelomic fluid of the adult (the red spherule cells, RSC), in spines, gonads, the digestive system, and in tube feet (Brasseur et al., 2018; Coates et al., 2018; Johnson, 1969; Perry and Epel, 1981; Service and Wardlaw, 1984; Smith et al., 2018; Smith et al., 2010). It is also thought that pigment is released by the pigmented cell that directly harms microbes. The antimicrobial mechanism of echinochrome has not been completely resolved, but evidence suggests its production of hydrogen peroxide and/ or iron chelation, abates microbial proliferation, (Coates et al., 2018; Lebedev et al., 2005; Perry and Epel, 1981). All of these studies agree that sea urchin pigments have anti-microbial activity, and that these small molecules may also contribute to states of cell physiology and gene expression (Jeong et al., 2014; Kim et al., 2018).

The developmental origins of pigment cells in the purple sea urchin have been traced to a group of mesodermal cells, the non-skeletogenic mesoderm (NSM) (Cameron et al., 1991; Croce and McClay, 2010; Materna and Davidson, 2012; McClay et al., 2000; Oliveri et al., 2002; Ruffins and Ettensohn, 1996; Sherwood and McClay, 1999; Sweet et al., 1999). Among the NSM cell types, pigment cells are specified first by Delta/Notch (D/N) signaling from the micromeres (Calestani et al., 2003; Calestani and Rogers, 2010; Croce and McClay, 2010; Davidson et al., 2002a; Foster et al., 2019; Materna and Davidson, 2012; McClay et al., 2000; Oliveri et al., 2002; Ransick et al., 2002; Rast et al., 2002; Sherwood and McClay, 1999; Sweet et al., 2002; Sweet et al., 1999). The D/N signaling directly activates the transcription factor glial cells missing (*gcm*), which in turn activates several pigment cell genes, including: polyketide synthase1 (*pks1*), three flavin monooxygenases (*fmo1*, 2 and 3), a sulfotransferase (*sult*) and a dopachrome tautomerase (*dopt*) (Calestani et al., 2003; Calestani and Rogers, 2010; Davidson et al., 2002b; Ransick and Davidson, 2006; Ransick et al., 2002; Rast et al., 2002). The enzymes *Pks1* and *Fmo1* (reannotated as *Fmo3* in this study) are required for the biosynthesis of the echinochrome in pigment cells (Calestani et al., 2003). *Pks1* is required for any pigment biosynthesis in the adult while *Fmo3* is required for accumulation of Spinochrome B in spines, but not for echinochrome synthesis in RSC and tube feet (Wessel et al., 2020). Genes expressed selectively in pigment cells, such as *gcm* and *pks1*, are first detected at blastula stage in a ring of NSM precursors surrounding the skeletogenic mesenchyme (SM) and later during blastula stage they become restricted to the aboral NSM (Calestani et al., 2003; Duboc et al., 2010; Gibson and Burke, 1987; Gustafson and Wolpert, 1967; Kominami

et al., 2001; Materna et al., 2013; Ransick and Davidson, 2012; Ransick et al., 2002). At the gastrula stage, *gcm*, as well as the pigment biosynthetic genes are expressed in cells migrating into the blastocoel, and their expression is maintained throughout larval development (Calestani et al., 2003; Gibson and Burke, 1987; Gustafson and Wolpert, 1967; Kominami et al., 2001; Ransick and Davidson, 2006; Ransick et al., 2002). Some evidence suggests *gcm* may be required for the development of pigment cells, since in *gcm* knockdown embryos, the cells do not migrate to the ectoderm at the mesenchyme blastula stage, and pigments are lost (Ransick and Davidson, 2006). Moreover, ectopic expression of *gcm* in skeletogenic mesoderm (SM) cells is sufficient to induce a similar single cell migration to the ectoderm and the accumulation of pigment in these cells, but not the acquisition of the typical morphology of pigment cells (Damle and Davidson, 2012). *Gcm* is also expressed at higher levels in adult hematopoietic tissues, the pharynx and the axial organ, which produce new coelomocytes upon immune challenge (Golconda et al., 2019). Thus, while *gcm* is an important transcriptional regulator, its mechanistic role in larval and adult pigmented cells is still unclear.

Here we tested the role of *gcm* in pigment cell development through single cell transcriptomic approaches, and define polytypic pigment cell subpopulations. Our analyses revealed two distinct groups of pigment cells, a mitotic population and a differentiated population. We further analyzed the diversity of *fmo* genes, essential players in the pigmentation pathway, and found a group of *fmo* genes specifically expressed in the embryonic pigment cells and in the coelomocytes of the adult immune system. We also performed a single cell mRNA analysis of *gcm* morphants and revealed that pigment cell specification is dependent on *gcm*. Our study leverages our

understanding of the Gcm transcription factor, an essential factor in the development of pigmentation in the larvae (Calestani and Rogers, 2010; Oulhen and Wessel, 2016; Ransick and Davidson, 2006; Ransick and Davidson, 2012; Wessel et al., 2020) and adult (Wessel et al., 2020) and provides deep datasets for exploring evolutionary selection of the biosynthesis of pigmentation.

Results

The ectodermal transcriptional state at a single cell level

Pigment cells of the sea urchin larva appear as a homogeneous population based on their morphology, expression of the transcription factor *gcm*, the enzyme *pks1*, and the presence of red pigment. Yet the many immune functions ascribed to pigmented cells in larvae and in adult tissues implies a heterogeneous population of cells. We tested this hypothesis with single cell sequencing technology. To enrich for pigment cells, we dissociated embryos/larvae from 48hpf and 72hpf, enriching for the ectodermal layer that includes the pigment cells (*Calestani et al., 2003; McClay and Marchase, 1979; Ransick and Davidson, 2012*), captured cells by drop-seq technology (*Figure 1A*), and sequenced the resulting cDNAs at a single cell level. The ectoderm-enriched single cell transcriptome revealed multiple cell-types, mostly of ectodermal origin as expected, and the overall cell cluster organization was highly reproducible between 48hpf and 72hpf samples (*Figure 1B*). The cells that formed the main clusters of the tSNE plots represented ectodermal cell types, including ciliary band cells, apical, oral, aboral, and lateral ectodermal cells (*Figure 1C*) based on expression of known cell-type marker genes. The pigment cell markers *gcm* and *pks1* were found in clusters 2 and 13, which showed a similar transcriptome profile, and cluster 12, which had a transcriptomic profile that was distinct from clusters 2 and 13. In addition to *gcm* and *pks1*, cluster 12 also expressed the transcription factor *six1* and its co-factor *eya*, which are expressed in aboral secondary mesoderm at mesenchyme blastula (*Poustka et al., 2007; Ransick and Davidson, 2012*). Cluster 2 and 13 expressed *gcm*, *pks1* and well-known pigment cell markers (*fmo3, fmo5-1, sult1c2* [*Calestani et al., 2003; Ransick and Davidson,*

2006; Ransick and Davidson, 2012; Ransick et al., 2002]). Other key clusters we found are the proneural apical plate (cluster 10), ciliary band neurons (cluster 16), serotonergic neurons (cluster 15 and 17), skeletal cells (cluster 11), and mid-gut cells (cluster 14). Overall our single cell transcriptomes showed consistent cell types between 48hpf gastrulae and 72hpf early larvae.

Gcm is enriched in three clusters at Gastrula and early larva stages

Single cell transcriptomes revealed cryptic cell-state distinctions by assessing their individual gene expression profiles, even though the limiting mRNA captured in each drop likely over-represents the more abundant transcripts in a cell. With sufficient depth in sequencing, these datasets allowed us to identify cell-state specific marker genes and to predict the nature of each cluster (details in Figure 1 legend). We use the term cell state here instead of cell type for the compelling reason that distinctions in these cell groupings by this technology does not always mean a distinction in cell morphology, physiology, function, or cellular location, key characteristics of the use of cell type (Clevers et al., 2017). Transcript profiles outnumber the classic definition of cell types, which is necessarily limited when one considers lineage variations and boundaries between lineages. Instead, cell states refer to populations of cells that show distinctions in transcript accumulation, which may or may not be reflected in the end-point function of the cell.

First, we analyzed the cell state of *gcm* expression during embryonic development, from the early zygote to the larva stage. To this aim we integrated the three *gcm*-enriched clusters from the ectodermal dataset (2, 12 and 13) to the *gcm*-

enriched cluster of a time-course single cell analysis of sea urchin embryos encompassing eight developmental time points, from eight-cell stage to late gastrulae (Foster et al., 2020). The gcm enriched cluster from this dataset was integrated to that of the ectodermal datasets and analyzed together as in Seurat V3 (Stuart et al., 1821). This analysis revealed 7 cell states of gcm-expressing cells, across the developmental time points (Figure 2A, Figure 2—figure supplement 1), showing that gcm is dynamically expressed between many different cell states through development. This is an unusual feature for gene expression in that gcm is not just accumulating in more cell states as they appeared, but that gcm is transiently expressed in many different cell states. This suggests that gcm in this embryo may initiate various gene activities, but that maintenance of that gene activity is dependent on other transcriptional mechanisms. This may reflect the many different activator and repressor roles that gcm has in this embryo.

We then analyzed the expression of gcm in two time points key in pigment cell differentiation (48 and 72hpf). Of these cell clusters expressing gcm, we found enrichment for other pigment cell terminal differentiation genes in three clusters in late gastrulae (48hpf; clusters 2, 12, 13) and two clusters in larvae (72hpf; clusters 2, 13 shown in Figure 2B,C). We next examined pigment gene expression in these three clusters. Pks1 was highly expressed in cells of clusters 2 and 13, and at low levels in cells of clusters 12 at gastrula stage (see violin plot, Figure 2—figure supplement 2). Feature plots showed that clusters 2 and 13 are spatially close, but separate, meaning that cell states 2 and 13 had similarities in their transcript profiles, while cluster 12 diverged significantly from cluster 2 and 13 (Figure 2B). To define the differences in

each of these three gcm+ clusters we performed differential gene expression analysis (clusters 2 vs 12, 2 vs 13 and 12 vs 13) and analyzed the average logFC for pks1 in each cluster. We found that pks1 was highly expressed in clusters 2 and 13 with respect to cluster 12, where pks1 levels were low. To determine if the pks1 low levels are stage dependent we normalized the data and compared cluster 12 to cluster 13 at 48hpf and 72hpf separately and found that at both stages pks1 levels in cluster 12 were lower than in clusters 2 and 13 (Supplementary file 1). Similarly, other known pigment cell markers, such as sult1c2 or fmo5-1 (previously called sult and fmo2, (Calestani et al., 2003; Ransick and Davidson, 2012) were highly expressed in clusters 2 and 13 but not in cluster 12 (Supplementary file 1). This finding shows that gcm is dynamically expressed during embryogenesis, and that it transiently marks two distinct pigment cell clusters (2 and 13). Another cluster that included mesodermal markers (12) expressed gcm in gastrulae, and low pks1 expression, suggesting that this is a population of pigment cell mesodermal precursors.

Gcm marks mitotic and post mitotic pigment cell populations

Having defined that clusters 2 and 13 are related based on their transcriptomic profile, and represent pigment cells, we analyzed these clusters in additional detail. The number of cells in these two clusters changed from 48hpf to 72hpf. Over a period of 24 hours, the proportion of cells in cluster two increased, while the proportion of cells in cluster 13 decreased (Figure 2D). To investigate the differences between the clusters, we performed differential gene expression analysis of clusters 2 and 13 at 48hpf and 72hpf. Both clusters expressed pigment cell markers at high levels, and the only

difference was that cluster 13 (the cluster with fewer cells at both time points) also expressed transcripts encoding S-phase and mitotic activities at high levels. The elevated markers in cluster 13 included *cdk1*, *pcna*, DNA polymerases, DNA ligases, condensins, and centromere proteins (Supplementary file 2), all genes that were absent in cluster 2. This result suggested that there were two pigmented cell clusters with a different cell cycle potential, cluster 13 being pigment cells still undergoing mitosis and cluster two being post-mitotic pigment cells. To test whether there are pigment cells in a mitotic state, or whether migrating pigment cells are all mitotic, we performed an EdU pulse for 30 min, followed by Sp1 staining (a conserved pigment cell marker [Gibson and Burke, 1985]). We found that from gastrula to early larva, cells that were in S-phase represented 34% of total pigment cells, and were broadly distributed in the embryos (Figure 2E,F,G). These cells were synthesizing DNA and dividing (see anaphase cells, inset Figure 2E, and Figure 2H, white arrowhead). This finding of a mitotic pigment cell cluster suggests that pigment cells have the ability to divide as migratory cells. Unique about this tSNE cluster of pigment cells though is the significant increase in mRNAs involved in cell cycle progression, a feature not normally seen in other dividing cells within the embryo.

Expression of unique pigment genes identified by scRNA-seq

To test for polytypic pigment cells, we defined the expression of newly identified pigment cell markers found in the scRNA-seq by in situ hybridization. scRNA-seq data showed that *gcm* and *pkc1* are expressed in the same group of cells at these timepoints. The co-expression of these two genes was always assumed based on a

similar pattern of expression and functional studies (Calestani et al., 2003; Calestani and Rogers, 2010; Materna et al., 2013; Ransick and Davidson, 2006; Ransick and Davidson, 2012), but never tested by RNA co-localization. We independently tested the scRNA-seq result by performing double fluorescent in situ hybridization (FISH) of *gcm* and *pks1*. In blastulae and gastrulae, *pks1* and *gcm* were expressed in the same cells that have pigment cell morphology (Figure 3A–B’). In larvae, however, *pks1* expression was maintained in pigment cells, while *gcm* expression transitioned into the left coelomic pouch, the site of the future adult rudiment (Figure 3C–C’). Previously, one single *gcm*-expressing cell was found close to the coelomic pouch with a *gcm::gfp* recombinant BAC (Ransick and Davidson, 2012). We therefore used *gcm* in gastrulae and *pks1* in larvae as markers to assess whether the putative pigment cell genes were expressed exclusively in this cell population.

We then selected a few genes whose expression were relevant to pigment cell function and analyzed their expression pattern together with *gcm* (at gastrula, 48hpf) and *pks1* (at early larva, 72hpf) by double in situ hybridization. Among these genes, ABCG11 is an ATP-binding cassette transporter homolog of the white half-transporter that transports pigment precursors and intermediates into the *Drosophila* eye (Mackenzie et al., 2000; Shipp and Hamdoun, 2012). *Abcg11* was selectively expressed in the pigment cell clusters 2 and 13, but not in cluster 12 (violin plots in Figure 2—figure supplement 2). Double FISH showed that *Abcg11* was exclusively expressed in *gcm* and *pks1* + cells (Figure 3D,E–E’, F–F’).

MIF (macrophage inhibitory factor) is a member of a gene family of cytokines that regulate innate immunity (Hibino et al., 2006; Nishihira, 2000). Of these genes, *mif5*

was expressed in the same *gcm*⁺ pigment cell clusters at 48hpf (2, 12, and 13, violin plots in *Figure 2—figure supplement 2*), and by 72hpf it became enriched in the pigment cell clusters 2 and 13 exclusively. Double FISH confirmed expression of *mif5* in pigment cells at both stages (*Figure 3G–H'*). A known pigment cell marker is the enzyme sulfotransferase (*Calestani et al., 2003*). Single cell data showed that *sult1c2* was highly expressed in pigment cell clusters at 48hpf and 72hpf and violin plots showed that *sult1c2* expression remained low in a few cells of all the other cell types (violin plots in *Figure 2—figure supplement 2*). Double FISH confirmed a strong expression of *sult1c2* in the *gcm/pks1*⁺ cells (*Figure 3I–I'*). Opioid receptor, kappa 1-like (*oprk*) is a Cholecystinin Receptor Type A, a G-protein coupled receptor that binds to cholecystinin peptide hormones to modulate feeding and dopamine-induced behavior (*Crawley, 1991*). scRNA-seq data showed that *oprk* was enriched in clusters 2, 12 and 13 at 48hpf, but by 72hpf it was enriched also in the ciliary band neurons (cluster 16), and we detected a spotty expression in this region by FISH (arrowhead in *Figure 3L'* and violin plots in *Figure 2—figure supplement 2*). Lastly, we analyzed the expression profile of the gene glutamate receptor 6 (*gluR6*). *GluR6* was not detectable by single cell sequencing at 48hpf, and double FISH analysis showed only faint signals in *gcm*⁺ cells (violin plots in *Figure 2—figure supplement 2*; *Figure 3M–M'*). *GluR6* was detectable at 72hpf though at low levels in cluster 2 (differentiated pigment cells) and cluster 16, a cluster of neurons. Double FISH with *pks1* at 72hpf showed that *gluR6* was expressed in *pks1*⁺ cells in or underneath the ectoderm (*Figure 3M–M'* and inset one in N') and in cells that are not pigment cells also located in the blastocoel beneath (basal to) the ectoderm (*Figure 3N'* insets 1 and 2). *GluR6* was also highly

expressed in both coelomic pouches (*Figure 3N– N'* and inset), a tissue not included in the scRNA-seq enriched for the ectoderm. Altogether the single cell transcriptome data and double FISH on the transcripts found in the scRNA-seq dataset indicated that genes for pigment cell function (besides *gcm* and *pks1* in gastrulae) are never expressed in the mesodermal cluster (cluster 12), but rather in the mitotic (cluster 13) and post-mitotic (cluster2) pigment cell clusters.

A diverse family of fmos is present in distinct pigment cell states

Having defined that there are two distinct sub-populations of pigment cells in embryos that expressed *gcm* and *pks1*, we investigated the expression of a class of enzymes that at least one representative modifies the polyketide pigment scaffold, the Flavin-dependent monooxygenases (Fmo). Fmos are present throughout phylogeny, and although their conserved protein domain is typical of proteins with oxidizing activity, it is still unclear the exact role in the sea urchin pigment biosynthetic pathway/s. To determine the expression of fmos in echinoderms, we sought all Fmo domains in the *S. purpuratus* genome (echinobase.org, V4.2; RRID:SCR_013732), and we found that not all of the identified fmos were specifically expressed in pigmented cells. We performed a phylogenetic analysis to look for protein subfamilies, and found 4 Fmo clades: one is in the same clade with vertebrate Fmo proteins (clade 2), two are not closely related to Fmo proteins in other organisms (clades 1 and 3), and one sub-family is in the same clade with fly orthologs (Figure 4A). Among these 4 clades, only the fmo genes in clade one were found significantly enriched in both of the pigment cell states. Fmo5-1, fmo3 and fmo2-2 (previously named 1,2,3) were all expressed in the mitotic

and differentiated pigment cell clusters, but not in the mesodermal cluster 12 (violin plots in Figure 4—figure supplement 1; Figure 4L–M). We next investigated the co-expression of clade one fmos with gcm and pks1 by double in situ hybridization (Figure 4B–I'). We found that fmo5-1 and 3 were exclusively expressed in pigment cells, while fmo2-2 is expressed also in cells that are gcm-negative (Figure 4D–D'). Fmo2 was not detected in the ectoderm single cell transcriptomes, but we analyzed its spatial expression since it was at the base of clade 1 (Figure 4A). In addition to the pigment cells, fmo2 was expressed at low level in a few cells of the blastocoel that are gcm/pks1 –. They are in tight apposition to the ectoderm (Figure 4E,E' and I,I'). Overall, by scRNA-seq and FISH we found that all fmos from clade one were enriched in pigment cells. In a different sea urchin species (*Hemicentrotus pulcherrimus*), knocking out the gene function of fmo3 resulted in a change in pigmentation of the adult, distinct from the albino phenotype of the pks1 gene knockout (Wessel et al., 2020). Perhaps these Fmo family members uniquely have enzymatic activities for polyketide metabolites, so that in the pigment cell the polyketide derived from Pks1 is converted to the intensely deep purple pigment.

We next determined the expression of fmos that were not found in the scRNAseq transcriptome. We used the available data of *S. purpuratus* quantitative RNA expression through development (<http://www.spbase.org:3838/quantdev/>) as an additional tool to determine the temporal expression of representative genes from the 4 clades of fmos. Clades 2, 3 and 4 fmos were expressed at low levels during embryonic development, with less than 1000 transcripts per embryo (Figure 5A,B). Only fmo2-3 from clade three was expressed between 1000 and 2000 transcripts/embryo. All transcripts of clade one

fmos were instead above 10,000 transcripts/embryo, reflecting their high abundance in the embryos. Only fmo2, located at the base of clade one in the phylogenetic tree, was not abundant in the embryo and had less than 250 transcripts/embryo (Figure 5A,B). We investigated further the spatial expression of representative genes for the other three fmo clades that are poorly expressed at embryonic stages. Fmo1, fmo2-3 and fmo5 are not exclusively expressed in the pigment cells, and FISH analysis showed low expression levels in the whole embryo (Figure 5—figure supplement 1). Our experiments show that only clade one fmos are highly expressed in embryos (in agreement with our FISH and transcriptomic analyses), while the other clades of fmos are poorly or not detectable at all in embryos.

Since our results showed that many fmos were not expressed during development, we analyzed their temporal expression in adult pigmented structures (spines and tube feet) and in the adult immune system (coelomocytes) by performing quantitative PCR in these tissues (Figure 5C–E). We first documented that spines and tube feet have pigmented cells by the criteria of deep red color and immunoreactivity to a conserved pigment cell marker recognized by the antibody Sp1 (Gibson and Burke, 1985; Figure 5C,D). We found that at least one gene of clades 2, 3 and 4 was expressed in the spines, while fmos from clade one were lacking (Figure 5C). Tube feet were not a site of significant fmo expression, only two of the fmos were detected (Figure 5D). Coelomocytes, however, expressed fmos from clade 1, but not the other fmos (Figure 5E). These cells are diverse in morphology and putative function, of which the red spherule cells are known to produce echinochrome and spinochrome (Coates et al., 2018; Hira et al., 2020). In the pigment biosynthesis pathway, the enzyme pks1 is

upstream of the fmos in the biosynthetic pathway of echinochrome. We found that pks1 was highly expressed in the adult immune cells, but expressed only at low levels in spines and tube feet. We interpret these results as fmos might function independently from pks1, and that low levels of polyketide synthase gene activity are sufficient for maintenance levels of polyketide synthesis and pigmentation, especially in immune-quiescent animals (not challenged by bacteria). Taken together, these results suggested that of the four fmo families, only clade 1 was found specifically in pigment cells of the embryo and in the adult red spherule cells, two cell-types with known immune functions. Within this clade, fmo2-2 and fmo2 are expressed also in embryonic cells that do not contain pigments, pointing these to a distinct oxidizing function. Other fmos are expressed in the spines (clades 2, 3 and 4) where pks1 is expressed at low levels, suggesting a role for non-clade 1 fmos independent of pigment biosynthesis and from gcm and pks1. Overall, we conclude that the many pigmented cell populations in the larvae and adults of *S. purpuratus* have distinct gene profiles.

Pigment cell dependence on gcm function

Having defined that gcm expression is associated with the expression of pigment cell genes, we tested whether pigment cell gene expression was dependent on gcm. Gastrulae in which gcm is knocked down have a unique phenotype with no pigmentation and low levels of pks1, sult1c2 and fmo3 (Ransick and Davidson, 2006). Even if some indirect evidence exists (Ransick and Davidson, 2006) a direct test of pigment cells presence/absence in gcm KD or null embryos is still missing. To understand whether in gcm KD embryos pigment cells are 'pigmentless' because of a

lack of *pks1* and *fmo3*, or whether there are fewer pigment cells, we tested *gcm* role in pigment cells early specification and pigment production. We performed a single cell differential gene analysis of 48hpf embryos compared to control embryos to determine the fate of the pigment cell clusters when *gcm* is knocked down. We dissociated control and *gcm* KD gastrulae without enriching for the ectoderm to ensure that all cells were captured for the analysis. Therefore, these tSNE plots (Figure 6A) revealed information different than the scRNAseq of ectoderm-enriched wild-type embryos (Figure 1B). The tSNE plots of the single cell transcriptome of control gastrula show a main group of ectodermal clusters (ciliary band neurons, serotonergic neurons, apical and aboral ectoderm), and clusters for skeleton, coelomic pouches, foregut and mid-gut (Figure 6A). Cells from *gcm*-KD embryos were very sensitive to dissociation and more susceptible to lysis than normal embryos. Because of the low number of sequenced cells, and the consequent lack of many ectodermal cell types, pigment cells were represented by a single cluster (cluster 6). The *gcm* KD experiment resulted in the same tSNE cluster profile as with the control MO. Since fewer cells were sequenced in the *gcm* MO experiment as a result of greater sensitivity to dissociation, we normalized the cells in each cluster for the total sequenced cells in each experimental condition. While 40% of the clusters did not change size (Figure 6—figure supplement 1), the pigment cell cluster population dramatically decreased by a factor of 7 with *gcm* knocked down by the morpholino (Figure 6B). We interpret this result as *gcm* is responsible to specify pigment cells, and not just for driving expression of pigment cell genes.

We next examined the overall expression level of pigment cell genes following *gcm* knockdown, and found that the expression of most of the tested genes decreased

in cell state cluster 6 (Figure 6C shows fold change; feature plots and violin plots in Figure 6—figure supplement 2 and Figure 6—figure supplement 3), while *gcm* expression remained unchanged. In *gcm* morphants, the expression of pigment genes in the embryo decreased because there were significantly fewer pigment cells (Figure 6D shows gene expression in controls versus morphants). To focus the analysis on *gcm* targets, we tested by qPCR expression level changes of other *gcm* targets (Figure 6E). The expression of *gluR6* and *fmo2*, which were expressed in more cell-states than the *pks1*⁺ cells, was unchanged. *Fmo2-2* from clade1, which was not detectable in the single cell experiment, was down-regulated in the *gcm* morphant. Genes like *fmo2-3* and *fmo5*, not found in the pigment cell cluster, did not change expression in *gcm* morphant embryos (Figure 6E). In contrast, larvae of cas9-mediated knockout of *pks1* were indistinguishable from control larvae when stained for Sp1 antibody (Figure 6F). The *pks1* - KO is used here to test homeostasis in the pigment cells and whether Pks1 and/or pigment may have any function, directly or indirectly, in regulating gene expression. It has been previously shown that embryos lacking Pks1 still have pigment cells (Calestani et al., 2003). Pigment cells stained with the Sp1 antibody show that not only the number, but also the distribution of pigment cells does not change in Pks1 depleted larvae. Gene expression of candidate genes in *pks1*- larvae was also indistinguishable from controls (Figure 6—figure supplement 4). We conclude that pigmentation does not have a role in positioning these cells within the larvae, in regulating numbers or density of the pigment cells, or change overtly the physiology of the cells as this would have been revealed by the cellular phenotype or steady state levels of relevant candidate mRNAs.

Having shown that *gcm* defines the number and gene expression of pigment cells, we tested in vivo the localization of surviving pigment cells in the absence of *gcm*. To this aim, we stained surviving pigment cells with Sp1 in early larvae, when pigment cells reach their final position. In agreement with the single cell data, we find fewer pigment cells in the *gcm* MO larvae compared to controls, and their localization appears to be randomly dispersed throughout the aboral ectoderm (Figure 6G). Taken together our experiments show that *gcm* knock down compromises the number of pigment cells.

Discussion

Gcm as a regulator of pigment

Pigments are compounds that selectively absorb a certain wavelength, and their production relies on a complex biosynthetic pathway that, for most plants and animals, involves ill-defined pathways. The cells that make the pigment are diverse and distinct in their biology, and the echinoderm pigments themselves also likely serve diverse functions. For instance, spinochromes and other pigments have been identified in adult of *Echinometra mathaei*, a species of sea urchins with many color variations between adults of the species, and pigmentation in this animal is correlated with fitness (Brasseur et al., 2018). It has also been proposed that the water-soluble pigment Spinochrome E can penetrate into the egg cytoplasm of the sand dollar *Scaphechinus mirabilis* and be used as a source of hydrogen peroxide in the embryo (Brasseur et al., 2018; Drozdov et al., 2017). Here we use a combination of RNA single cell sequencing and fluorescent in situ hybridization to take a deep look at the genes responsible for this incredible variety of colors, and we show that pigment cells within an organism are highly diverse (Figure 7).

Gcm has been shown to be expressed first in the Veg2 lineage during early blastula and to be later restricted to pigment cells (Ransick and Davidson, 2006; Ransick and Davidson, 2012; Ransick et al., 2002). There was some preliminary evidence that gcm might be expressed in other cell types, that is the coelomic pouches, later in larval development (Ransick and Davidson, 2012). Here, by integrating the gcm positive cells from the ectodermal scRNAseq with a time course over eight embryonic stages, we found that gcm was expressed broadly and that its expression was dynamic.

Since Notch signaling is activated not at the same time in the ring of mesodermal cells surrounding primary mesenchyme cells (Sherwood and McClay, 1997; Sweet et al., 1999), it is possible that *gcm* and downstream genes are activated at slightly different times, one cell after the other. Our conclusion is that *gcm* is not essential for all these clusters, but it has only a transient expression regulated at the transcriptional and/or post-transcriptional level. This suggests either a housekeeping role for *gcm* that was underestimated, or a role in initiating transcription of certain genes, whose expression will be maintained by other regulatory factors following possibly an additional environmental cue or cell signaling. Last, of the three *gcm*⁺ clusters, cluster 12 defines mesodermal pigment cell progenitors (*six1* and *eya* markers), determined but not differentiated yet. This dynamic *gcm* expression would have been more difficult to establish without single cell analysis, since it was not detectable so far using other techniques.

It is known that the downstream effects of *gcm* loss of function is the lack of pigment cell precursors migration during gastrula and lack of pigment synthesis (Ransick and Davidson, 2006). Here we show that *gcm* is responsible for specifying pigment cells, not just driving expression of pigment cell genes. ScRNA-seq of *gcm* morphants shows that in the absence of *gcm* the number of pigment cells decreases, and the expression of pigment gene markers is reduced. Moreover, immunostaining with the Sp1 antibody found that a small population of pigment cells remain when *Gcm* is depleted, but we can't exclude the existence of *gcm*-independent compensatory mechanisms. Our results suggest that despite *gcm* being necessary to specify most pigment cells, *gcm*-independent mechanisms appear to activate pigment cell identity

when gcm activity is depleted. We cannot, however, exclude the possibility that residual gcm expression somehow fully activates a small population of precursor pigment cells.

A previous transcriptome study identified pigment cell genes by enriching the sample for late gastrula gcm-GFP expressing cells (Barsi et al., 2015). To evaluate our single cell data in this comparative light, we searched for the pigment cell markers described in our study in the list of genes previously reported, and found that all of the candidates we identified were enriched in the above gcm⁺ transcriptome. Some of the pigment markers we found were also expressed in other cell states. For example, oprk, that in other organisms modulates feeding and dopamine-induced behavior (Crawley, 1991) and the glutamate receptor glur6, were expressed in pigment cells and also in the ciliary band neurons (cluster 16). In particular, glur6 is expressed in a cluster of neurons located in the blastocoel basal to the ciliary band that resemble the localization of the lateral ganglion cells (Bisgrove and Burke, 1987; Perillo et al., 2018). We interpret the neuronal expression as new gene expression in those cells, and not a cell fate transition of the pigment cells since pigment cells are differentiated in larvae. The fact that pigment cells were expressing these neuronal receptors and others (this study and [Barsi et al., 2015]) suggests that they might also have a sensory function possibly coordinated with their immune response function. Our study provides new insights into the expression and function of gcm, a transcriptional factor that we found dynamically expressed in several embryonic clusters, and that controls the different states of pigment cells and possibly their different functions.

Do pigment cells maintain a stem state?

How the pigment cells are regulated through larval development and into adulthood is not clear. Do all the various pigment cells of the adult, including in the spines, tube feet, and coelomocytes, come from a core stem cell population? Or do they each have their own developmental lineage from a unipotential precursor? Each of these populations, as shown here, appear to have different transcript profiles and functions. What is clear is that the numbers of pigment cells expand if larvae are immune challenged (Kiselev et al., 2013; Ageenko et al., 2014; Wessel et al., 2020). In our scRNA-seq of ectodermal cells we found two distinct populations of pigment cells, one of which is enriched in mitotic markers, suggesting the possibility of a stem cohort of dividing cells. We favor the concept of a stem population for this lineage because the transcript profile is distinct in the mitotic cohort and those cells are resistant to the *gcm* knockdown. Perhaps *Gcm* is essential for the lineage expansion, but not for the stem precursors.

The molecular analysis does suggest distinct populations, or at least different cell states, of pigment cells. By labeling mitotic cells with a pulse of EdU and pigment cells with the marker *Sp1*, we visualized the mitotic pigment cells in vivo. These cells do not seem to belong to a specific embryonic stem niche, but are distributed throughout the embryo, suggesting that a portion of all pigment cells undergo mitosis to increase the pool. The presence of mitotic pigment cells at the early larva stage also suggests that these cells retain a very active cell cycle while migrating. With such essential functionality likely in the innate immune defense, it may be expected that a stem cell-type population may be responsible for homeostatic functions in pigment cell populations. Pigment cells from embryos of the sand dollar *S. mirabilis* and in the sea urchin *S. intermedius* can be induced to differentiate when cultured in the adult coelomic

fluid in vitro, suggesting that genes involved in pigment synthesis in the adult also regulates pigment cells of the larva (Ageenko et al., 2016). Our finding that adult coelomocytes from *S. purpuratus* express *gcm*, *pkx1* and group1 *fmos* suggests that pigmented cells from the adult cells share biosynthetic and cell fate pathways.

We pose two hypotheses to explain the finding of a mitotic pigment cell cluster: either there is a unique stem cell population of pigment cells, or all pigment cells have the ability to divide. In both cases, a follow up question is whether they divide upon a signal, and what this signal might be (i.e. coelom infection, arm growth, ectodermal injury). It has been suggested that in adult sea urchin a population of pigment stem cells is set aside in the pre-metamorphosis larva to serve as a reservoir for the adult cells (Wessel et al., 2020). This is more than initially activating the program for these cells to develop, but also to maintain and replace these cells in the adult. We can rule out pigment presence as an indicator of homeostasis since the *pkx1* knock out larvae had similar pigment cells distribution as wild-type larvae. Thus, a stem population for these cells may be key for adult survival and the single cell results herein may point to this population. We propose that the hereby-identified embryonic mitotic cluster might be the cells that are selected in early embryos to replenish the larvae and/or adult stages. However, as we learned here, pigmented cells in the diverse tissues of an adult are pigmented cells in distinct states. Either the presumed stem cluster is capable of diversifying, or new pigment lineages form later in development. An alternative hypothesis is that the pigment cell mitotic cluster serves in the embryo as a reservoir to replenish blastocoelar cell-types. We propose that the mitotic pigment cell cluster might represent a group of cells that retain the developmental potential that allows these cells

to convert to other blastocoelar cells, or to the highly migratory pigment cells of the larvae. Overall our interpretation of mitotic and post-mitotic pigment cell clusters is compatible with cells in the latter cluster having immune function, whereas the mitotic pigment cells population may be a stem/dividing population. Indelible lineage labeling will be important to test these hypotheses.

A unique family of fmos is expressed in cells with immune functions

Spines, tube feet, cells in coelomic fluid and larvae of sea urchins express different types of flavin containing monooxygenases (fmo) enzymes that contribute to the pigment biosynthesis. In adult spines and tube feet under control conditions (not immune challenged) *pk1* is not transcribed, or is so at very low levels. Two scenarios are possible to explain this observation: either cells could have a level of enzyme (protein) sufficient to synthesize some pigment or in the absence of degranulation of pigment in response to infections or injuries, the synthesis of pigment is halted. In the case of cells with predicted similar functions, for example coelomic-pigmented cells (the red spherule cells), versus larval-pigmented cells, the composition of pigment synthesizing enzymes and other gene expression is instead similar. Cells with immune functions (adult coelomocytes and larval pigment cells) always express the same group of fmos, together with the enzyme *pk1* and the transcriptional factor *gcm*. Both cells produce the pigment echinochrome A, a molecule with antimicrobial properties (Brasseur et al., 2018; Coates et al., 2018; Gerardi et al., 1990; Johnson, 1969; Lebedev et al., 2005; Perry and Epel, 1981), suggesting that *gcm*, *pk1*, and group1 fmos work together to make this unique pigment, but might not interact to make other

echinoderm pigments. These results may reveal that distinct biochemical pathways lead to the same chemical outcome, or that pigmentation is just one of multiple different functions in what we routinely refer to as a single pigment cell.

The revelation of the fmo family in sea urchins is striking. Fmos are broadly expressed in yeast, plants and animals where their main role is detoxification (Dolphin et al., 1996; Janmohamed et al., 2001; Phillips et al., 1995; Phillips and Shephard, 2017). Seeing diverse expressions throughout the sea urchin tissues suggests more than ‘just’ a metabolic or detoxifying function for these family members. Fmo enzymes catalyze the oxidation of different types of substrates that are involved in different types of biological processes across phyla: detoxification, metabolite activation, and in the case of the sea urchin, pigmentation and maybe more. In the sea urchin, *S. purpuratus*, one cohort of fmos (Group 4, that includes fmo1, 2–5 and 6) groups together with the fly orthologs. Flies have two fmos, one with housekeeping functions and another involved in development (Scharf et al., 2004). Fmos from clade two are instead grouped with the human Fmos. These fmos were not detectable in embryogenesis, but rather in the spines. Human fmos are expressed in liver, lung, kidney and skin (Peters, 1985; Schlaich, 2007), cannot be induced or inhibited, and although being involved with drug metabolism, their physiological role is poorly understood (Scharf et al., 2004). The most studied FMO is the human Fmo3, an enzyme that is constitutively active in the adult human liver where it participates in the oxidation of drugs and xenobiotics, and mutations in Fmo3 contribute to the disease trimethylaminuria (Cashman and Zhang, 2002; Dolphin et al., 1996; Krueger and Williams, 2005). We find no expression of the fmos in clade 2 (the sea urchin group closest to vertebrates) in embryos, nor in the

coelomocytes, as it is only expressed in spines. Sea urchin spines are covered by a thin epithelium (Peters, 1985) that is populated by pigmented cells that we show here are marked with the embryonic pigment cell marker Sp1, most likely being red spherule cells. We suggest that the fmo clade two role might be connected to human Fmo function in the skin, where it might detoxify xenobiotics to which the spines are exposed in the seawater. The other fmo clades are also expressed in the spines, and further studies will need to determine the expression and function of these genes in other adult structures.

It is still not understood why the fmo family is so expanded in this sea urchin. Only 1 fmo gene is present in yeast, whereas many flowering plants have 26 (Schlaich, 2007), two in flies (Scharf et al., 2004), six in humans (Dolphin et al., 1996) and at least 12 in sea urchin that are not pseudogenes, since each are expressed at some time and place in the animal. A hypothesis to explain this expansion is that such a variety of enzymes is necessary for the detoxification of a wide range of xenobiotics present in diverse ecosystems, especially benthic niches. We predict that with Cas9 targeting mechanisms of fmos followed by pigment MS/MS analysis will enable definitive Fmo functionality to expose this enigmatic family of enzymes. Furthermore, single cell RNA-seq of adult tissues will likely reveal the cell-types that express these genes.

In conclusion, by leveraging two distinctly derived RNA-seq datasets, we were able to create confident profiles of genes relevant to pigment cell function. We discovered that pigment cells conserve a mitotic cell population that might represent a stem population for embryonic and adult immune cells. We also realize how important in situ FISH and single-cell RNA-seq is for confidence in conclusions. Bulk sequencing,

even isolated populations of cells, brings many caveats of interpretation which likely shift upon testing with scRNA-seq and double-FISH.

Materials and Methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional Information
Gene (Strongylocentrotus purpuratus)	<i>gcm</i>	Echinobase	SPU_006462	
Gene (Strongylocentrotus purpuratus)	<i>pks1</i>	Echinobase	SPU_002895	
Gene (Strongylocentrotus purpuratus)	<i>fmo2</i>	Echinobase	SPU_002963	
Gene (Strongylocentrotus purpuratus)	<i>fmo3</i>	Echinobase	SPU_017374	
Gene (Strongylocentrotus purpuratus)	<i>fmo2-2</i>	Echinobase	SPU_014947	
Gene (Strongylocentrotus purpuratus)	<i>fmo5-1</i>	Echinobase	SPU_012348	
Gene (Strongylocentrotus purpuratus)	<i>sult1c2</i>	Echinobase	SPU_006187	
Gene (Strongylocentrotus purpuratus)	<i>oprk1c</i>	Echinobase	SPU_000719	
Gene (Strongylocentrotus purpuratus)	<i>mif5</i>	Echinobase	SPU_020036	
Gene (Strongylocentrotus purpuratus)	<i>glur6</i>	Echinobase	SPU_018876	
Gene (Strongylocentrotus purpuratus)	<i>abcg11</i>	Echinobase	SPU_018876	
Antibody	anti-Sp1 (mouse monoclonal)	DSHB	DSHB Cat# Sp1, RRID:SCR_013527	IF(1:50)
Sequenced- based reagent	control MO; control morpholino	Gene Tools		5'-GCTTTGGAGTAACCTTCTGCACCAT-3' (0.5 mM)
Sequenced- based reagent	<i>gcm</i> MO; <i>gcm</i> morpholino	Gene Tools		5'-GCTTTGGAGTAACCTTCTGCACCAT-3' (0.5 mM)
Software,algorithm	Image J	Image J	RRID:SCR_003070	
Software,algorithm	Seurat v 3.1.4	SEURAT	SEURAT, RRID:SCR_007322	
Other	DAPI stain	Invitrogen	D1306	(1 mg/mL)

Animals

Adult *Strongylocentrotus purpuratus* were obtained from Josh Ross (info@scbiomarine.com) off the California coast and kept in artificial seawater at 16°C. Gametes were obtained by shaking adult sea urchins or by intracoelomic injection of 0.5M KCl. Eggs were fertilized in the presence of 1 mM 3- amino-triazol (3-AT) (Sigma, Cat. #:A8056). Embryos were cultured at 16°C in filtered seawater from the Marine Biological Laboratory (MBL).

RNA whole mount in situ hybridization

For fluorescent whole mount in situ hybridization (FISH), we followed the protocol outlined in Cole and Arnone, 2009; Perillo et al., 2016. Signal was developed with fluorophore-conjugated tyramide (Perkin Elmer, Cat. #:NEL752001KT; RRID:AB_2572409). Labeled probes were transcribed from linearized DNA using digoxigenin-11-UTP (Sigma Aldrich, Cat. #:11277073910), or labeled with DNP (Mirus Bio, Cat. #:MIR3825) following kit instructions. Probes were synthesized using primers in Supplementary file 3. 20 to 40 samples were imaged with a Zeiss 800 confocal microscope from the Brown University Leduc Bioimaging Core Facility (RRID:SCR_017861).

Immunocytochemistry and EdU pulse labeling

Larvae and adult tissues were fixed in 4% paraformaldehyde (PFA) in filtered seawater (FSW) for 15 min at room temperature, washed multiple times in phosphate-buffered saline with 0.1% Tween-20 (PBST), and incubated overnight at 4°C with the Sp1

antibody 1:50 (DSHB; RRID:SCR_013527) in 1 mg/ml Bovine Serum Albumin (BSA) and 4% sheep serum in PBST. Samples were then washed three times with PBST and incubated for 2 hr at room temperature with the secondary anti-mouse anti-body conjugated to Alexa 488 (Life Technologies; Cat#:A-21121, RRID:AB_2535764) diluted 1/500 in 1 mg/ml BSA in PBST. 5-ethynyl-2-deoxyuridine (EdU) pulse labeling of the pigment cells was performed with the Click-IT EdU imaging kit (Life Technologies; Cat#:C10340). Embryos and larvae were soaked in 10 mM EdU in FSW for 30 min (considering a 20 min cell cycle), washed five times with FSW and fixed in 4% PFA/FSW. Following EdU detection with fluorescent azide (according to the manufacturer's instructions), samples were stained with the Sp1 antibody as described above. 20 to 40 larvae were stained for imaging. Samples were mounted for imaging with an Olympus FV3000 Confocal Microscope (RRID:SCR_017015) equipped with high sensitivity GaAsP detectors managed by cellSens software (RRID:SCR_016238) from the Brown University Leduc Bioimaging Core Facility (RRID:SCR_017861). Raw files were analyzed and figures were made using the software Image J (RRID:SCR_003070).

Perturbation experiments with MO injection and CRISPR Cas9

Translation-blocking antisense morpholino (MO) against Gcm (SPU_006462; MO synthesized by Gene-Tools; sequence 5'-GCTTTGGAGTAACCTTCTGCACCAT-3') was used at a concentration of 0.5 mM. MOs were injected with 20% glycerol and 10,000 MW fluorescent dextran (injection solution). Eggs were dejelled by washing in pH4.0 seawater. For each experiment, around 600 zygotes were injected with ~2–4 pl of

oligonucleotide injection solution by constant pressure injection in with 1 mM 3-AT (Sigma). For each condition, only 300 of these injected embryos with the expected phenotypes were dissociated at 48hpf and used for single cell RNA seq. The gcm morphant has a unique phenotype with a straight gut and no pigmentation. As a negative control, fertilized eggs were injected with a MASO sequence irrelevant to *S. purpuratus* (the MO to Foxy3 of *Patiria miniata*, and is referred to as the control morpholino (5'-TGCGATTAGAATCAAAACGGAGTGA-3'). It is used to compare to the gcm morphants. We chose to use a gcm morpholino to knockdown gcm instead of CRISPR/Cas9 as used previously (Oulhen and Wessel, 2016). CRISPR/Cas9 randomly mutates the targeted genomic sequence and in the wildtype animals we decided it best to not have the variability of different mutations (between experiments, and between each injected embryos) that could have different effects in the gene expression and the phenotypes obtained.

Three Cas9 guide RNAs (gRNAs; 200 ng/ul of each gRNA) previously used in our lab (Oulhen and Wessel, 2016) were used to target *pks1* DNA. gRNAs were mixed with 500 ng/ul of Cas9 mRNA, injected into freshly fertilized eggs as described previously in Oulhen and Wessel, 2016.

Phylogenetic analysis

Phylogenetic analyses were conducted using MEGA version 6 (MEGA Software, RRID:SCR_000667) (Tamura et al., 2011). Protein sequences were obtained from NCBI and echinobase. Phylogenetic reconstruction was carried out using the maximum likelihood method, and tested also by Neighbor-joining methodology. Both tests

resulted in the same outcome. Bootstrap values were determined by 500 replicates. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The analysis involved 25 amino acid sequences. There were a total of 671 positions in the final dataset. SPU or NCBI accession numbers are as follows: AAA52457 HsFMO; P31513 HsFMO3; Q99518 HsFMO2; CAA77797 HsFMO4; P49326 HsFMO5; SPU_002963 Sp-Fmo2; SPU_007044 Sp-FMO6L; SPU_009114 Sp-Fmo5; SPU_012348 Sp-Fmo5-1; SPU_014947 Sp-Fmo2-2; SPU_017252 Sp-Fmo5-2; SPU_017374 Sp-Fmo3; SPU_017639 Sp-Fmo; SPU_022596 Sp-Fmo5-3; SPU_022597 Sp-Fmo3-1; SPU_022765 Sp-Fmo2-3; SPU_023681 Sp-Fmo2-5; SPU_024227 Sp-Fmo5-6; SPU_024261 Sp-Fmo5-4; SPU_025958 Sp-Fmo2- 4; EDL39293.1 Mm-Fmo4; NP_001171509 Mm-Fmo6; BAA03745 Mm-Fmo; AAQ94601 Dr-Fmo1; NP_989910 Gg-Fmo3; AAK94940 Dm-Fmo1; AAL27708 Dm-Fmo2; Q47PU3 Gg-Fmo5; NP_001087387 XI-Fmo5-1; CAD10798 Comamonas_testosteroni.

Adult tissues RNA extraction and qPCR

RNA from 100 embryos was isolated with the RNeasy Micro kit (Qiagen, Cat#:74004), while RNA from adult tissues was isolated from Trizol (Thermo Fisher Scientific, Cat#:10296010). cDNA synthesis was performed using Maxima kit (Life Technologies, Cat#:K1641). qPCR was performed using ABI 7900 real time instrument with Maxima SYBR master mix (Life Technologies, Cat#:FERK0222) and normalized to ubiquitin transcripts. Experiments were run in three biological replicates, and every biological

replicate was run on the qPCR machine with three technical replicates. In the graphs, the mean of the three technical replicates is shown, and error bars represent biological replicates. List of primers used for qPCR are in Supplementary file 4.

Embryo dissociations

Embryos were collected and washed twice with calcium-free seawater, and then suspended in hyalin-extraction media (HEM) for 15 min (George and McClay, 2019). For enrichment of ectodermal cells, embryos were transferred to 0.5M NaCl as soon as the squamous ectodermal epithelium cells became rounded and loosened from each other. The embryos were then gently sheared with a pipette to remove the ectoderm from the basal lamina and the remainder of the embryo was removed from the enriched ectoderm population with a 40 micron Nitex mesh filter. When the entire embryo was to be dissociated, the embryos were subjected to more prolonged HEM treatment, then transferred into 0.5M NaCl, gently sheared with a pipette to complete dissociation, and residual cellular clumps were removed with a 40 micron Nitex mesh. Dissociated cells were counted on a hemocytometer, and diluted with 0.5M NaCl to reach the appropriate concentration for the scRNA-seq protocol. Embryos were collected only by settling at 1xg, and at no time were cells or embryos pelleted in a centrifuge. Throughout the procedure, specimens were kept at 4°C.

Single cell RNA sequencing

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 (RRID:SCR_016387) (2 x 150 bp paired-end runs). The non-injected 48hpf and 72hpf samples were sequenced in separate lanes (350M reads per lane). The morpholino injected samples were sequenced together in the same lane (350M reads total). Single cell unique molecular identifier (UMI) counting was performed using Cell Ranger Single Cell Software Suite 3.0.2 from 10X Genomics (Cell Ranger, RRID:SCR_017344). The custom transcriptome reference was generated from assembly Spur_4.2 using Cell Ranger. Cell Ranger matrices were further analyzed using the R package Seurat v 3.1.4 (SEURAT, RRID:SCR_007322) (Stuart et al., 1821). Cells with at least 400 and at most 2500 different represented genes were included in downstream analysis. The top 2000 highly variable gene representatives (features) across the datasets were used to integrate datasets. TSNE projection and clustering analysis for all datasets were conducted using 15 dimensions and a resolution of 0.5. Cluster markers were found using Find Conserved Markers and Find Markers functions. The integration of gcm-enriched clusters from multiple time points was performed using Seurat following the standard integration pipeline (dims:10 res:0.1). The time course dataset included samples collected from eight-cell stage, 64 cell stage, morula, early blastula, hatched blastula, mesenchyme blastula, early gastrula and late gastrula (4-8hpf) stage (Foster et al., 2020). The sequencing data generated here have been made publicly available at Gene Expression Omnibus [<https://www.ncbi.nlm.nih.gov/geo/>] (GSE155427).

Acknowledgements

We are grateful for, and acknowledge the support for this work from the National Institutes of Health grants 9RO1GM125071 and 1R01GM132222 to GMW and 1P20GM119943 to NO. The Sp1 antibody developed by Robert Burke was obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH and maintained at The University of Iowa, Department of Biology, Iowa City, IA 52242; <https://dshb.biology.uiowa.edu/> We are also grateful for use of the resources and services at the Computational Biology Core and Center for Computation and Visualization, Brown University.

Additional Information

Funding

	Grant reference number	Author
National Institutes of Health	9RO1GM125071	Gary Wessel
National Institutes of Health	1R01GM132222	Gary Wessel
National Institutes of Health	1P20GM119943	Nathalie Oulhen

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Margherita Perillo, Conceptualization, Data curation, Formal analysis, Validation, Investigation, Visu- alization, Methodology, Writing - original draft, Writing - review and editing; Nathalie Oulhen, Con- ceptualization, Software, Formal analysis, Validation, Investigation, Methodology, Writing - review and editing; Stephany Foster, Data

curation, Software, Formal analysis, Investigation, Methodology, Writing - review and editing; Maxwell Spurrell, Formal analysis, Investigation, Methodology, Writing - review and editing; Cristina Calestani, Conceptualization, Formal analysis, Writing - review and editing; Gary Wessel, Conceptualization, Data curation, Formal analysis, Supervision, Funding acquisition, Investigation, Visualization, Methodology, Writing - original draft, Project administration, Writing - review and editing

Author ORCIDs

Margherita Perillo <https://orcid.org/0000-0003-0845-507X>

Stephany Foster <https://orcid.org/0000-0002-9404-1692>

Maxwell Spurrell <https://orcid.org/0000-0001-7314-7482>

Cristina Calestani <https://orcid.org/0000-0001-8122-9349>

Gary Wessel <https://orcid.org/0000-0002-1210-9279>

Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.60388.sa1>

Author response <https://doi.org/10.7554/eLife.60388.sa2>

Data availability

The sequencing data have been made publicly available at Gene Expression Omnibus under GSE155427.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Foster S, Oulhen N, Wessel GM	2020	Sp single cell RNA-seq	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155427	NCBI Gene Expression Omnibus, GSE155427

References

- Ageenko NV, Kiselev KV, Dmitrenok PS, Odintsova NA. 2014. Pigment cell differentiation in sea urchin blastula- derived primary cell cultures. *Marine Drugs* 12:3874–3891. DOI: <https://doi.org/10.3390/md12073874>, PMID: 24979272
- Ageenko NV, Kiselev KV, Odintsova NA. 2016. Freezing tolerance of sea urchin embryo pigment cells. *Russian Journal of Marine Biology* 42:437–441. DOI: <https://doi.org/10.1134/S1063074016050023>
- Anderson HA, Mathieson JW, Thomson RH. 1969. Distribution of spinochrome pigments in echinoids. *Comparative Biochemistry and Physiology* 28:333–345. DOI: [https://doi.org/10.1016/0010-406X\(69\)91347-4](https://doi.org/10.1016/0010-406X(69)91347-4), PMID: 5777380
- Barsi JC, Tu Q, Calestani C, Davidson EH. 2015. Genome-wide assessment of differential effector gene use in embryogenesis. *Development* 142:3892–3901. DOI: <https://doi.org/10.1242/dev.127746>, PMID: 26417044
- Bisgrove B, Burke R. 1987. Development of the nervous system of the *Pluteus* larva of *Strongylocentrotus droebachiensis*. *Cell and Tissue Research* 248:335–343. DOI: <https://doi.org/10.1007/BF00218200>
- Brasseur L, Demeyer M, Decroo C, Caulier G, Flammang P, Gerbaux P, Eeckhaut I. 2018. Identification and quantification of spinochromes in body compartments of *Echinometra mathaei*'s coloured types. *Royal Society Open Science* 5:171213. DOI: <https://doi.org/10.1098/rsos.171213>, PMID: 30224975
- Buckley KM, Rast JP. 2017. An organismal model for gene regulatory networks in the Gut-Associated immune response. *Frontiers in Immunology* 8:1297. DOI: <https://doi.org/10.3389/fimmu.2017.01297>, PMID: 29109720
- Calestani C, Rast JP, Davidson EH. 2003. Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening. *Development* 130:4587–4596. DOI: <https://doi.org/10.1242/dev.00647>, PMID: 12925586
- Calestani C, Rogers DJ. 2010. Cis-regulatory analysis of the sea urchin pigment cell gene polyketide synthase. *Developmental Biology* 340:249–255. DOI: <https://doi.org/10.1016/j.ydbio.2010.01.026>, PMID: 20122918 Cameron RA, Fraser SE,
- Britten RJ, Davidson EH. 1991. Macromere cell fates during sea urchin development. *Development* 113:1085–1091. PMID: 1811928
- Cashman JR, Zhang J. 2002. Interindividual differences of human flavin-containing monooxygenase 3: genetic polymorphisms and functional variation. *Drug Metabolism and Disposition* 30:1043–1052. DOI: <https://doi.org/10.1124/dmd.30.10.1043>, PMID: 12228178

Ch Ho E, Buckley KM, Schrankel CS, Schuh NW, Hibino T, Solek CM, Bae K, Wang G, Rast JP. 2016. Perturbation of gut Bacteria induces a coordinated cellular immune response in the purple sea urchin larva. *Immunology & Cell Biology* 94:861–874. DOI: <https://doi.org/10.1038/icb.2016.51>, PMID: 27192936

Clevers H, Rafelski S, Elowitz M, Klein M, Trapnell C, Shendure J, Love JC. 2017. What is your conceptual definition of "Cell Type" in the Context of a Mature Organism? *Cell Systems* 4:225–229. DOI: <https://doi.org/10.1016/j.cels.2017.03.006>

Coates CJ, McCulloch C, Betts J, Whalley T. 2018. Echinochrome A release by red spherule cells is an Iron- Withholding strategy of sea urchin innate immunity. *Journal of Innate Immunity* 10:119–130. DOI: <https://doi.org/10.1159/000484722>, PMID: 29212075

Cole AG, Arnone MI. 2009. Fluorescent in situ hybridization reveals multiple expression domains for SpBrn1/2/4 and identifies a unique ectodermal cell type that co-expresses the ParaHox gene SpLox. *Gene Expression Patterns* 9:324–328. DOI: <https://doi.org/10.1016/j.gep.2009.02.005>, PMID: 19250980

Crawley JN. 1991. Cholecystokinin-dopamine interactions. *Trends in Pharmacological Sciences* 12:232–236. DOI: [https://doi.org/10.1016/0165-6147\(91\)90558-A](https://doi.org/10.1016/0165-6147(91)90558-A), PMID: 2048219

Croce JC, McClay DR. 2010. Dynamics of Delta/Notch signaling on endomesoderm segregation in the sea urchin embryo. *Development* 137:83–91. DOI: <https://doi.org/10.1242/dev.044149>, PMID: 20023163

Damle SS, Davidson EH. 2012. Synthetic in vivo validation of gene network circuitry. *PNAS* 109:1548–1553. DOI: <https://doi.org/10.1073/pnas.1119905109>, PMID: 22238426

Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY, Revilla R, Rust AG, Pan Z, Schilstra MJ, Clarke PJ, Arnone MI, et al. 2002a. A genomic regulatory network for development. *Science* 295:1669–1678. DOI: <https://doi.org/10.1126/science.1069883>, PMID: 11872831

Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh C-H, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY, Revilla R, Schilstra MJ, Clarke PJC, Rust AG, Pan Z, Arnone MI, et al. 2002b. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Developmental Biology* 246:162–190. DOI: <https://doi.org/10.1006/dbio.2002.0635>

Dolphin CT, Cullingford TE, Shephard EA, Smith RL, Phillips IR. 1996. Differential developmental and tissue- specific regulation of expression of the genes encoding three

members of the flavin-containing monooxygenase family of man, FMO1, FMO3 and FMO4. *European Journal of Biochemistry* 235:683–689. DOI: <https://doi.org/10.1111/j.1432-1033.1996.00683.x>, PMID: 8654418

Drozдов AL, Artyukov AA, Elkin YN. 2017. Pigments in egg cells and epidermis of sand dollar *Scaphechinus mirabilis*. *Russian Journal of Developmental Biology* 48:257–262. DOI: <https://doi.org/10.1134/S106236041704004X>

Duboc V, Lapraz F, Saudemont A, Bessodes N, Mekpoh F, Haillot E, Quirin M, Lepage T. 2010. Nodal and BMP2/4 pattern the mesoderm and endoderm during development of the sea urchin embryo. *Development* 137:223–235. DOI: <https://doi.org/10.1242/dev.042531>, PMID: 20040489

Foster S, Oulhen N, Wessel G: A single cell RNA sequencing resource for early sea urchin development. *Development* 2020, 147(17).

George AN, McClay DR. 2019. Methods for transplantation of sea urchin blastomeres. *Methods in Cell Biology* 150:223–233. DOI: <https://doi.org/10.1016/bs.mcb.2018.11.012>, PMID: 30777177

Gerardi P, Lassegues M, Canicatti C. 1990. Cellular distribution of sea urchin antibacterial activity. *Biology of the Cell* 70:153–157. DOI: [https://doi.org/10.1016/0248-4900\(90\)90372-A](https://doi.org/10.1016/0248-4900(90)90372-A)

Gibson AW, Burke RD. 1985. The origin of pigment cells in embryos of the sea urchin *Strongylocentrotus purpuratus*. *Developmental Biology* 107:414–419. DOI: [https://doi.org/10.1016/0012-1606\(85\)90323-9](https://doi.org/10.1016/0012-1606(85)90323-9), PMID: 3972163

Gibson AW, Burke RD. 1987. Migratory and invasive behavior of pigment cells in normal and animalized sea urchin embryos. *Experimental Cell Research* 173:546–557. DOI: [https://doi.org/10.1016/0014-4827\(87\)90294-1](https://doi.org/10.1016/0014-4827(87)90294-1), PMID: 3691675

Golconda P, Buckley KM, Reynolds CR, Romanello JP, Smith LC. 2019. The axial organ and the pharynx are sites of hematopoiesis in the sea urchin. *Frontiers in Immunology* 10:870. DOI: <https://doi.org/10.3389/fimmu.2019.00870>, PMID: 31105697

Griffiths M. 1965. A study of the synthesis of naphthaquinone pigments by the larvae of two species of sea urchins and their reciprocal hybrids. *Developmental Biology* 11:433–447. DOI: [https://doi.org/10.1016/0012-1606\(65\)90049-7](https://doi.org/10.1016/0012-1606(65)90049-7)

Gustafson T, Wolpert L. 1967. Cellular movement and contact in sea urchin morphogenesis. *Biological Reviews* 42:442–498. DOI: <https://doi.org/10.1111/j.1469-185X.1967.tb01482.x>, PMID: 4864367

Hibino T, Loza-Coll M, Messier C, Majeske AJ, Cohen AH, Terwilliger DP, Buckley KM, Brockton V, Nair SV, Berney K, Fugmann SD, Anderson MK, Pancer Z, Cameron RA,

Smith LC, Rast JP. 2006. The immune gene repertoire encoded in the purple sea urchin genome. *Developmental Biology* 300:349–365. DOI: <https://doi.org/10.1016/j.ydbio.2006.08.065>, PMID: 17027739

Hira J, Wolfson D, Andersen AJC, Haug T, Stensvåg K. 2020. Autofluorescence mediated red spherulocyte sorting provides insights into the source of spinochromes in sea urchins. *Scientific Reports* 10:1149. DOI: <https://doi.org/10.1038/s41598-019-57387-7>, PMID: 31980652

Janmohamed A, Dolphin CT, Phillips IR, Shephard EA. 2001. Quantification and cellular localization of expression in human skin of genes encoding flavin-containing monooxygenases and cytochromes P450. *Biochemical Pharmacology* 62:777–786. DOI: [https://doi.org/10.1016/s0006-2952\(01\)00718-3](https://doi.org/10.1016/s0006-2952(01)00718-3), PMID: 11551524

Jeong SH, Kim HK, Song IS, Noh SJ, Marquez J, Ko KS, Rhee BD, Kim N, Mishchenko NP, Fedoreyev SA, Stonik VA, Han J. 2014. Echinochrome A increases mitochondrial mass and function by modulating mitochondrial biogenesis regulatory genes. *Marine Drugs* 12:4602–4615. DOI: <https://doi.org/10.3390/md12084602>, PMID: 25196935

Johnson PT. 1969. The coelomic elements of sea urchins (*Strongylocentrotus*) III. in vitro reaction to Bacteria. *Journal of Invertebrate Pathology* 13:42–62. DOI: [https://doi.org/10.1016/0022-2011\(69\)90237-7](https://doi.org/10.1016/0022-2011(69)90237-7)

Kim H, Cho S, Heo H, Jeong S, Kim M, Ko K, Rhee B, Mishchenko N, Vasileva E, Fedoreyev S, Stonik V, Han J. 2018. A novel atypical PKC- ι inhibitor, echinochrome A, enhances cardiomyocyte differentiation from mouse embryonic stem cells. *Marine Drugs* 16:192. DOI: <https://doi.org/10.3390/md16060192>

Kiselev KV, Ageenko NV, Kurilenko VV. 2013. Involvement of the cell-specific pigment genes pks and sult in bacterial defense response of sea urchins *Strongylocentrotus intermedius*. *Diseases of Aquatic Organisms* 103: 121–132. DOI: <https://doi.org/10.3354/dao02570>, PMID: 23548362

Kominami T, Takata H, Takaichi M. 2001. Behavior of pigment cells in gastrula-stage embryos of *Hemicentrotus pulcherrimus* and *Scaphechinus mirabilis*. *Development, Growth and Differentiation* 43:699–707. DOI: <https://doi.org/10.1046/j.1440-169X.2001.00605.x>

Krueger SK, Williams DE. 2005. Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacology & Therapeutics* 106:357–387. DOI: <https://doi.org/10.1016/j.pharmthera.2005.01.001>

Kuhn R, Wallenfels K. 1940. Echinochrome als prosthetische gruppen hochmolekularer symplexe in den eiern von *Arbacia pustulosa*. *Berichte Der Deutschen Chemischen Gesellschaft* 73:458–464. DOI: <https://doi.org/10.1002/cber.19400730511>

Lebedev AV, Ivanova MV, Levitsky DO. 2005. Echinochrome, a naturally occurring iron chelator and free radical scavenger in artificial and natural membrane systems. *Life Sciences* 76:863–875. DOI: <https://doi.org/10.1016/j.lfs.2004.10.007>

Liu D, Awazu A, Sakuma T, Yamamoto T, Sakamoto N. 2019. Establishment of knockout adult sea urchins by using a CRISPR-Cas9 system. *Development, Growth & Differentiation* 61:378–388. DOI: <https://doi.org/10.1111/dgd.12624>, PMID: 31359433

Mackenzie SM, Howells AJ, Cox GB, Ewart GD. 2000. Sub-cellular localisation of the white/scarlet ABC transporter to pigment granule membranes within the compound eye of *Drosophila melanogaster*. *Genetica* 108:239–252. DOI: <https://doi.org/10.1023/a:1004115718597>, PMID: 11294610

Materna SC, Ransick A, Li E, Davidson EH. 2013. Diversification of oral and aboral mesodermal regulatory states in Pregastrular Sea Urchin embryos. *Developmental Biology* 375:92–104. DOI: <https://doi.org/10.1016/j.ydbio.2012.11.033>, PMID: 23261933

Materna SC, Davidson EH. 2012. A comprehensive analysis of Delta signaling in pre-gastrular sea urchin embryos. *Developmental Biology* 364:77–87. DOI: <https://doi.org/10.1016/j.ydbio.2012.01.017>

McClay DR, Peterson RE, Range RC, Winter-Vann AM, Ferkowicz MJ. 2000. A micromere induction signal is activated by beta-catenin and acts through notch to initiate specification of secondary mesenchyme cells in the sea urchin embryo. *Development* 127:5113–5122. PMID: 11060237

McClay DR, Marchase RB. 1979. Separation of ectoderm and endoderm from sea urchin *Pluteus* larvae and demonstration of germ layer-specific antigens. *Developmental Biology* 71:289–296. DOI: [https://doi.org/10.1016/0012-1606\(79\)90170-2](https://doi.org/10.1016/0012-1606(79)90170-2), PMID: 387500

McClendon JF. 1912. A red substance in sea urchins. *Journal of Biological Chemistry* 11:435–441.

Nishihira J. 2000. Macrophage migration inhibitory factor (MIF): Its essential role in the immune system and cell growth. *Journal of Interferon & Cytokine Research* 20:751–762. DOI: <https://doi.org/10.1089/10799900050151012>

Oliveri P, Carrick DM, Davidson EH. 2002. A regulatory gene network that directs micromere specification in the sea urchin embryo. *Developmental Biology* 246:209–228. DOI: <https://doi.org/10.1006/dbio.2002.0627>

Oulhen N, Wessel GM. 2016. Albinism as a visual, in vivo guide for CRISPR/Cas9 functionality in the sea urchin embryo. *Molecular Reproduction and Development* 83:1046–1047. DOI: <https://doi.org/10.1002/mrd.22757>

Perillo M, Wang YJ, Leach SD, Arnone MI. 2016. A pancreatic exocrine-like cell regulatory circuit operating in the upper stomach of the sea urchin *Strongylocentrotus purpuratus* larva. *BMC Evolutionary Biology* 16:117. DOI: <https://doi.org/10.1186/s12862-016-0686-0>

Perillo M, Paganos P, Mattiello T, Cocurullo M, Oliveri P, Arnone MI. 2018. New Neuronal Subtypes With a “Pre- Pancreatic” Signature in the Sea Urchin *Strongylocentrotus purpuratus*. *Frontiers in Endocrinology* 9:650. DOI: <https://doi.org/10.3389/fendo.2018.00650>

Perry G, Epel D. 1981. Ca²⁺-stimulated production of H₂O₂ from naphthoquinone oxidation in *Arbacia* eggs. *Experimental Cell Research* 134:65–72. DOI: [https://doi.org/10.1016/0014-4827\(81\)90463-8](https://doi.org/10.1016/0014-4827(81)90463-8)

Peters BH. 1985. The innervation of spines in the sea-urchin *Echinus esculentus* L. *Cell and Tissue Research* 239: 219–228. DOI: <https://doi.org/10.1007/BF00214922>

Phillips IR, Dolphin CT, Clair P, Hadley MR, Hutt AJ, McCombie RR, Smith RL, Shephard EA. 1995. The molecular biology of the flavin-containing monooxygenases of man. *Chemico-Biological Interactions* 96:17–32. DOI: [https://doi.org/10.1016/0009-2797\(94\)03580-2](https://doi.org/10.1016/0009-2797(94)03580-2), PMID: 7720101

Phillips IR, Shephard EA. 2017. Drug metabolism by flavin-containing monooxygenases of human and mouse. *Expert Opinion on Drug Metabolism & Toxicology* 13:167–181. DOI: <https://doi.org/10.1080/17425255.2017.1239718>, PMID: 27678284

Poustka AJ, Ku“hn A, Groth D, Weise V, Yaguchi S, Burke RD, Herwig R, Lehrach H, Panopoulou G. 2007. A global view of gene expression in lithium and zinc treated sea urchin embryos: new components of gene regulatory networks. *Genome Biology* 8:R85. DOI: <https://doi.org/10.1186/gb-2007-8-5-r85>

Ransick A, Rast JP, Minokawa T, Calestani C, Davidson EH. 2002. New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. *Developmental Biology* 246:132–147. DOI: <https://doi.org/10.1006/dbio.2002.0607>, PMID: 12027439

Ransick A, Davidson EH. 2006. cis-regulatory processing of notch signaling input to the sea urchin glial cells missing gene during mesoderm specification. *Developmental Biology* 297:587–602. DOI: <https://doi.org/10.1016/j.ydbio.2006.05.037>, PMID: 16925988

Ransick A, Davidson EH. 2012. Cis-regulatory logic driving glial cells missing: self-sustaining circuitry in later embryogenesis. *Developmental Biology* 364:259–267. DOI: <https://doi.org/10.1016/j.ydbio.2012.02.003>, PMID: 22509525

Rast JP, Cameron RA, Poustka AJ, Davidson EH. 2002. Brachyury target genes in the early sea urchin embryo isolated by differential macroarray screening. *Developmental Biology* 246:191–208. DOI: <https://doi.org/10.1006/dbio.2002.0654>, PMID: 12027442

Ruffins SW, Ettensohn CA. 1996. A fate map of the vegetal plate of the sea urchin (*Lytechinus variegatus*) mesenchyme blastula. *Development* 122:253–263. PMID: 8565837

Scharf ME, Scharf DW, Bennett GW, Pittendrigh BR. 2004. Catalytic activity and expression of two flavin- containing monooxygenases from *Drosophila melanogaster*. *Archives of Insect Biochemistry and Physiology* 57: 28–39. DOI: <https://doi.org/10.1002/arch.20012>, PMID: 15352153

Schlaich NL. 2007. Flavin-containing monooxygenases in plants: looking beyond detox. *Trends in Plant Science* 12:412–418. DOI: <https://doi.org/10.1016/j.tplants.2007.08.009>, PMID: 17765596

Schrankel CS, Solek CM, Buckley KM, Anderson MK, Rast JP. 2016. A conserved alternative form of the Purple Sea Urchin HEB/E2-2/E2A transcription factor mediates a switch in E-protein regulatory state in differentiating immune cells. *Developmental Biology* 416:149–161. DOI: <https://doi.org/10.1016/j.ydbio.2016.05.034>, PMID: 27265865

Service M, Wardlaw AC. 1984. Echinochrome-A as a bactericidal substance in the coelomic fluid of *Echinus esculentus* (L.). *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* 79:161–165. DOI: [https://doi.org/10.1016/0305-0491\(84\)90008-7](https://doi.org/10.1016/0305-0491(84)90008-7)

Sherwood DR, McClay DR. 1997. Identification and localization of a sea urchin notch homologue: insights into vegetal plate regionalization and notch receptor regulation. *Development* 124:3363–3374. PMID: 9310331

Sherwood DR, McClay DR. 1999. LvNotch signaling mediates secondary mesenchyme specification in the sea urchin embryo. *Development* 126:1703–1713. PMID: 10079232

Shipp LE, Hamdoun A. 2012. ATP-binding cassette (ABC) transporter expression and localization in sea urchin development. *Developmental Dynamics* 241:1111–1124. DOI: <https://doi.org/10.1002/dvdy.23786>

Smith LC, Ghosh J, Buckley KM, Clow LA, Dheilly NM, Haug T, Henson JH, Li C, Lun CM, Majeske AJ, Matranga V, Nair SV, Rast JP, Raftos DA, Roth M, Sacchi S, Schrankel CS, Stensvåg K. 2010. Echinoderm immunity. *Advances in Experimental Medicine and Biology* 708:260–301. DOI: https://doi.org/10.1007/978-1-4419-8059-5_14, PMID: 21528703

- Smith AV, Barela Hudgell MA, Barone G, Bodnar AG, Buckley KM. 2018. Echinodermata: the complex immune system in echinoderms. In: Cooper E (Ed). *Advances in Comparative Immunology*. Springer. p. 409–501. DOI: https://doi.org/10.1007/978-3-319-76768-0_13
- Solek CM, Oliveri P, Loza-Coll M, Schrankel CS, Ho EC, Wang G, Rast JP. 2013. An ancient role for Gata-1/2/3 and scl transcription factor homologs in the development of immunocytes. *Developmental Biology* 382:280– 292. DOI: <https://doi.org/10.1016/j.ydbio.2013.06.019>, PMID: 23792116
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM. 1821. Comprehensive integration of Single-Cell data. *Cell* 177:31. DOI: <https://doi.org/10.1016/j.cell.2019.05.031>
- Sweet HC, Hodor PG, Ettensohn CA. 1999. The role of micromere signaling in notch activation and mesoderm specification during sea urchin embryogenesis. *Development* 126:5255–5265. PMID: 10556051
- Sweet HC, Gehring M, Ettensohn CA. 2002. LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. *Development* 129:1945–1955. PMID: 11934860
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28:2731–2739. DOI: <https://doi.org/10.1093/molbev/msr121>, PMID: 21546353
- Wessel GM, Kiyomoto M, Shen TL, Yajima M. 2020. Genetic manipulation of the pigment pathway in a sea urchin reveals distinct lineage commitment prior to metamorphosis in the bilateral to radial body plan transition. *Scientific Reports* 10:5. DOI: <https://doi.org/10.1038/s41598-020-58584-5>, PMID: 32029769
- Yaguchi S, Yaguchi J, Suzuki H, Kinjo S, Kiyomoto M, Ikeo K, Yamamoto T. 2020. Establishment of homozygous knock-out sea urchins. *Current Biology* 30:R427–R429. DOI: <https://doi.org/10.1016/j.cub.2020.03.057>, PMID: 32428469

Figures

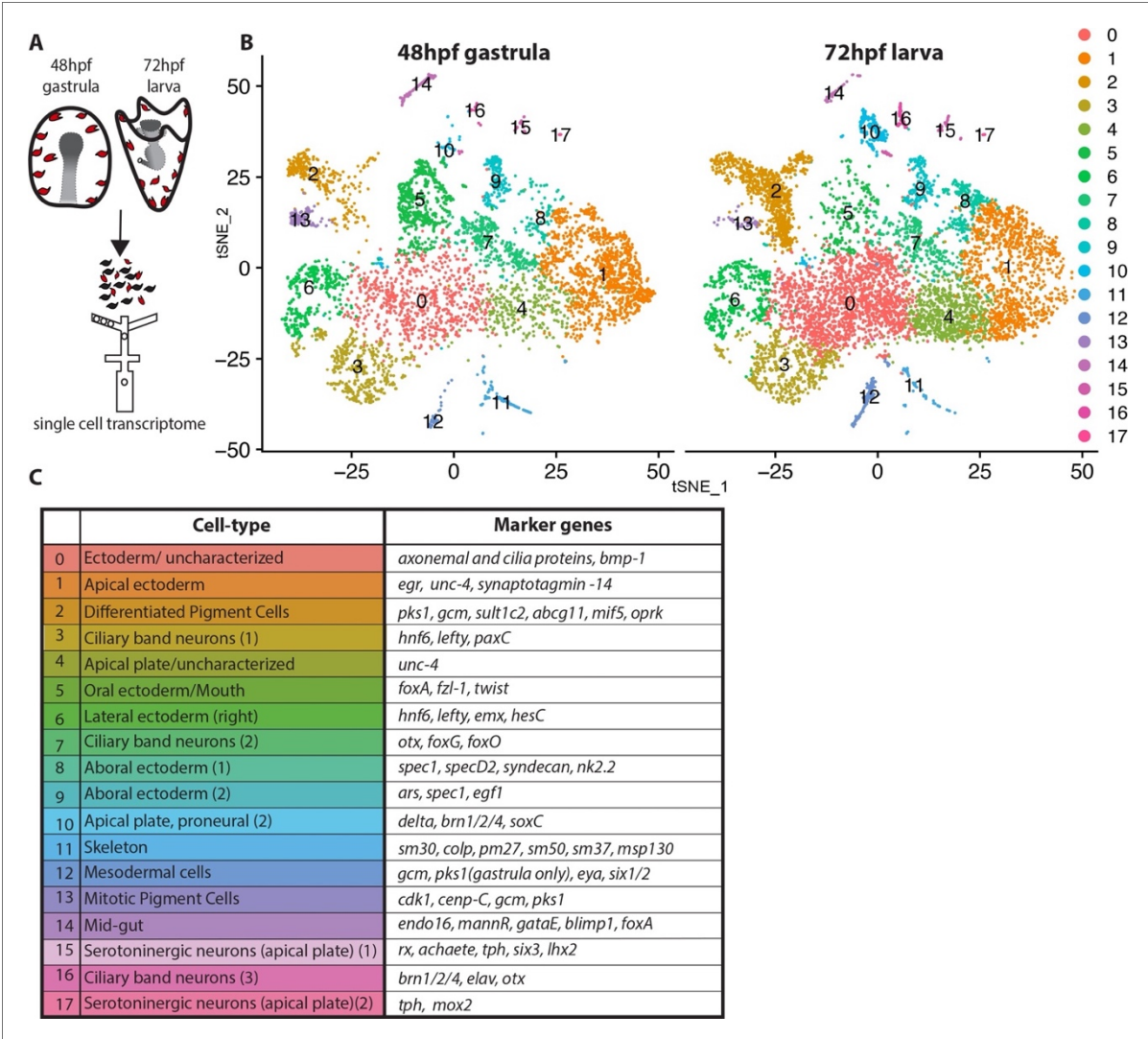


Figure 1. Dissection of *S. purpuratus* ectodermal cell-types by scRNA-seq. (A) Cartoon summarizing the dissociation of gastrulae and larvae (pigment cells in red, other cell-types in black), cell separation through drop-seq technology and RNA sequencing. Our single cell sequence datasets detected 20,489 reads/genes and a median 2,909 UMI counts per cell at 48hpf as well as 19,716 reads/genes and a median 1,136 UMI counts per cell at 72hpf. (B) tSNE plots of gastrula (48hpf) and larvae (72hpf) enriched for the ectodermal cell-types. Colors indicate major cell-types grouped by gene expression similarity. In the 48hpf sample, 5688 single cells were captured for downstream analysis, sequenced at a level of 81,121 reads per cell. The 72hpf sample includes 8178 single cells with an average of 54,788 reads per cell. The two samples were integrated to identify conserved cell types and cluster marker genes using Seurat. (C) Table summarizing the cell-type for each cluster (48hpf +72hpf) with the most representative marker genes (colors in the table reflect colors of the tSNE plot clusters).

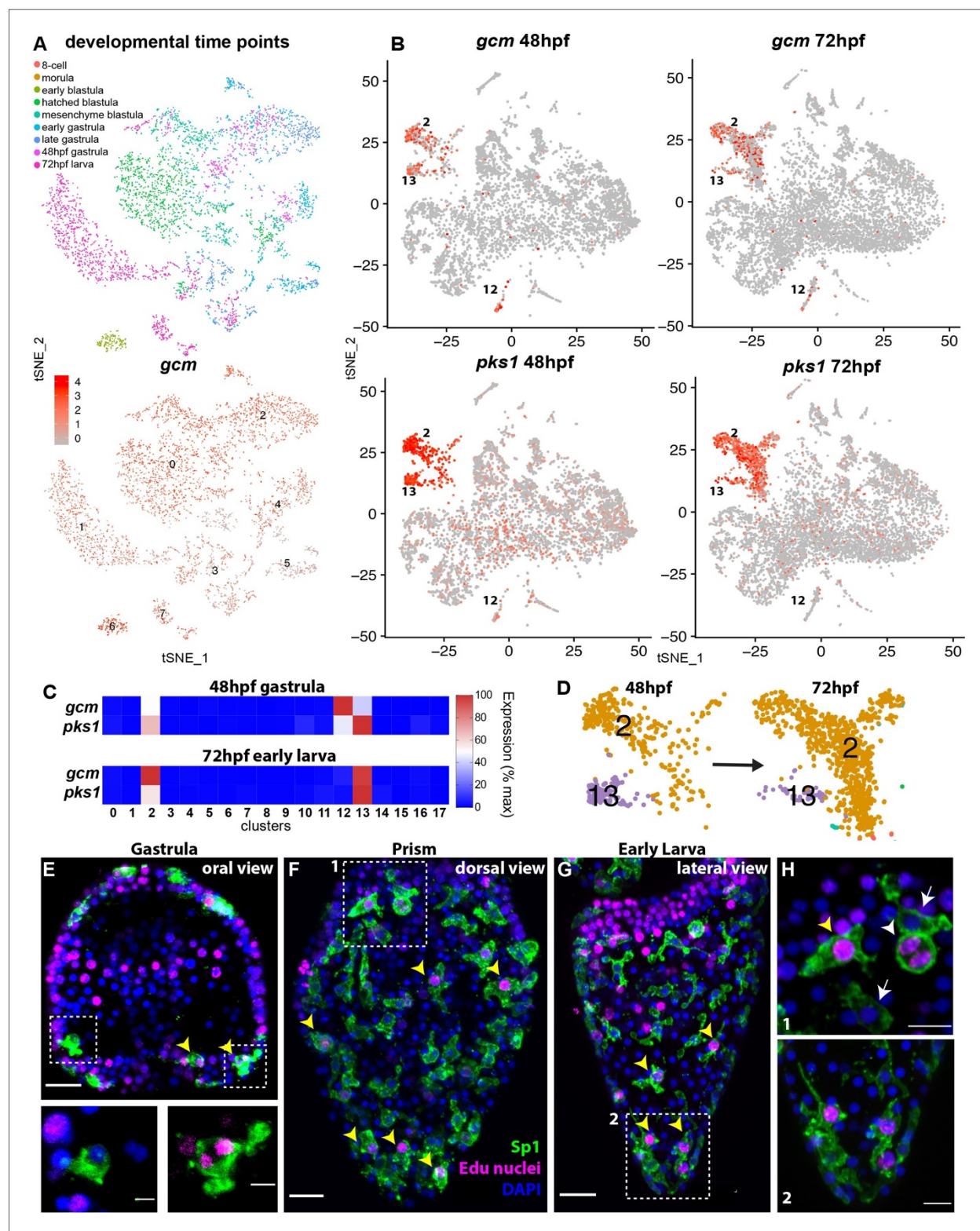


Figure 2. *gcm* marks pigment cell mitotic and post-mitotic populations. (A) tSNE plot showing integration of *gcm* enriched clusters from datasets encompassing nine

time points across sea urchin development: 8 cell stage, morula, early blastula, hatched blastula, mesenchyme blastula, early gastrula, late gastrula, 48hpf, and 72hpf. Cells are colored by time point. Feature plots for selected genes are provided in Figure 2—figure supplement 1. Using the dataset of Foster et al., 2020 we found that Gcm was first detected at 64 cell stage, when there are nine gcm+ cells, 1% of total cells. At morula stage 125 cells express gcm, 3.48%. Early blastula: 13.9% Hatched: 9.5% Mesenchyme blastula: 7.6% Early gastrula: 6.5% and late gastrula 4.7%. (B) Feature plots of gastrulae and larvae colored for gcm or pks1 showing gene expression in clusters 2 and 13 at 48hpf and 72hpf, and cluster 12 at 48hpf. Violin plots showing expression of gcm and pks1 are in Figure 2—figure supplement 2. (C) Heatmap from scRNA-seq data represents expression of gcm and pks1 in the three clusters mitotic (13) and differentiated (2) pigment cells, and mesodermal cells (12). gcm expression is high in the mesodermal cluster at 48hpf, but decreases at 72hpf. (D) Magnification of tSNE plot from Figure 1B shows that the number of cells in the differentiated pigment cells cluster (2) increases over time, while the mitotic cluster (13) decreases. (E,F,G) Double staining for pigment cells (Sp1 antibody) and Edu labeling to mark cells that have recently synthesized DNA (yellow arrowheads). Note that mitotic pigment cells are broadly distributed within the embryos and larvae. (H) Magnifications of boxes 1 and 2 from figure F and G. White arrowheads show a pigment cell undergoing mitosis, yellow arrowhead shows another mitotic pigment cell, arrows show pigment cells that are non-mitotic. Scale bars are 100 μ m (F, G), 20 μ m (E) and 5 μ m (H). All images are stacks of merged confocal Z sections.

The online version of this article includes the following figure supplement(s) for figure 2:

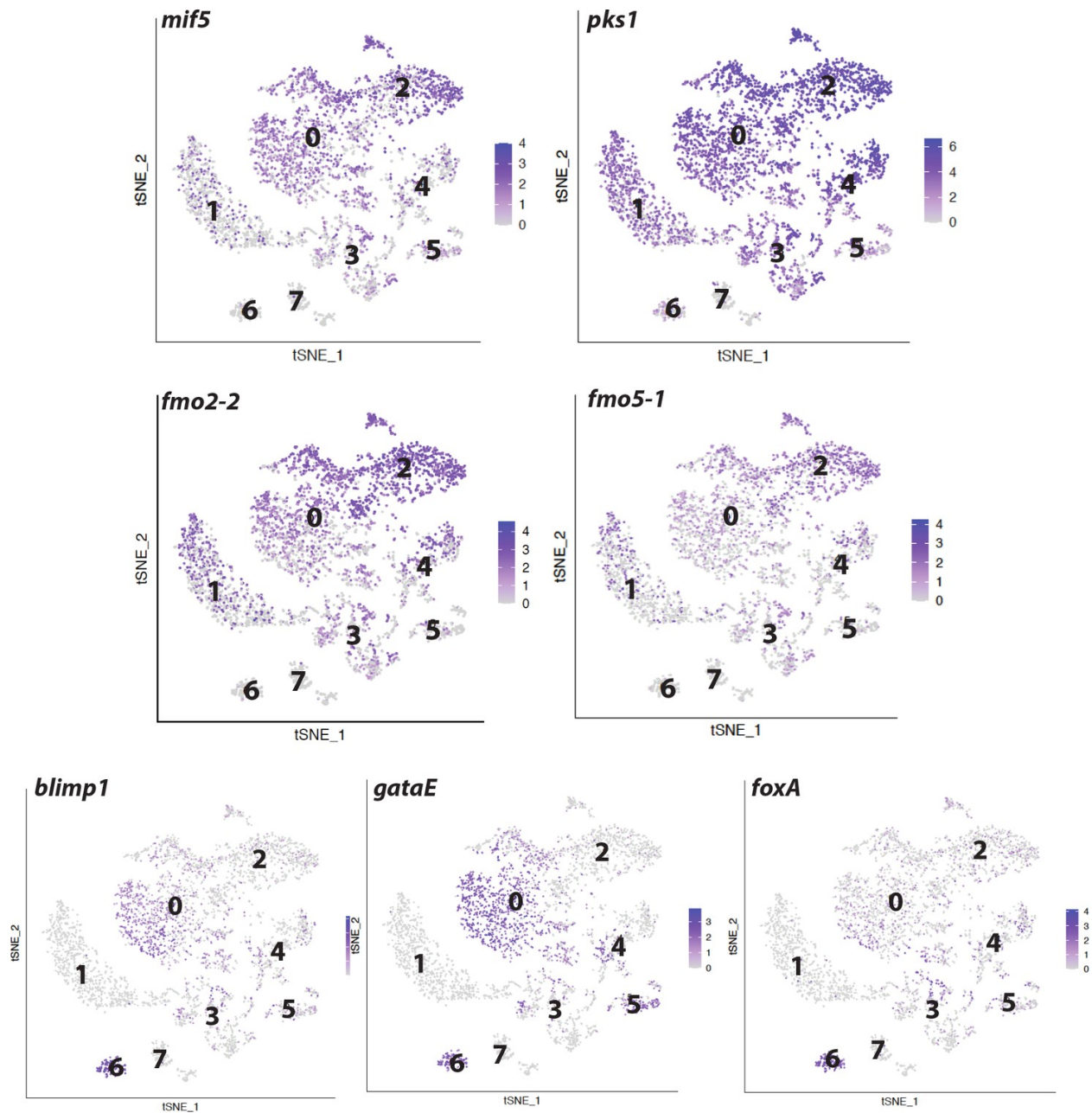


Figure supplement 1. Examples of cell states across developmental stages (A) Feature plots showing expression of pigment cell markers across developmental stages. Co-expression of these markers identifies pigment cells in the data. (B) Feature plots showing co-expression of stomach cell genes, *blimp1*/*krox*, *foxA*, and *gataE* in cluster 6.

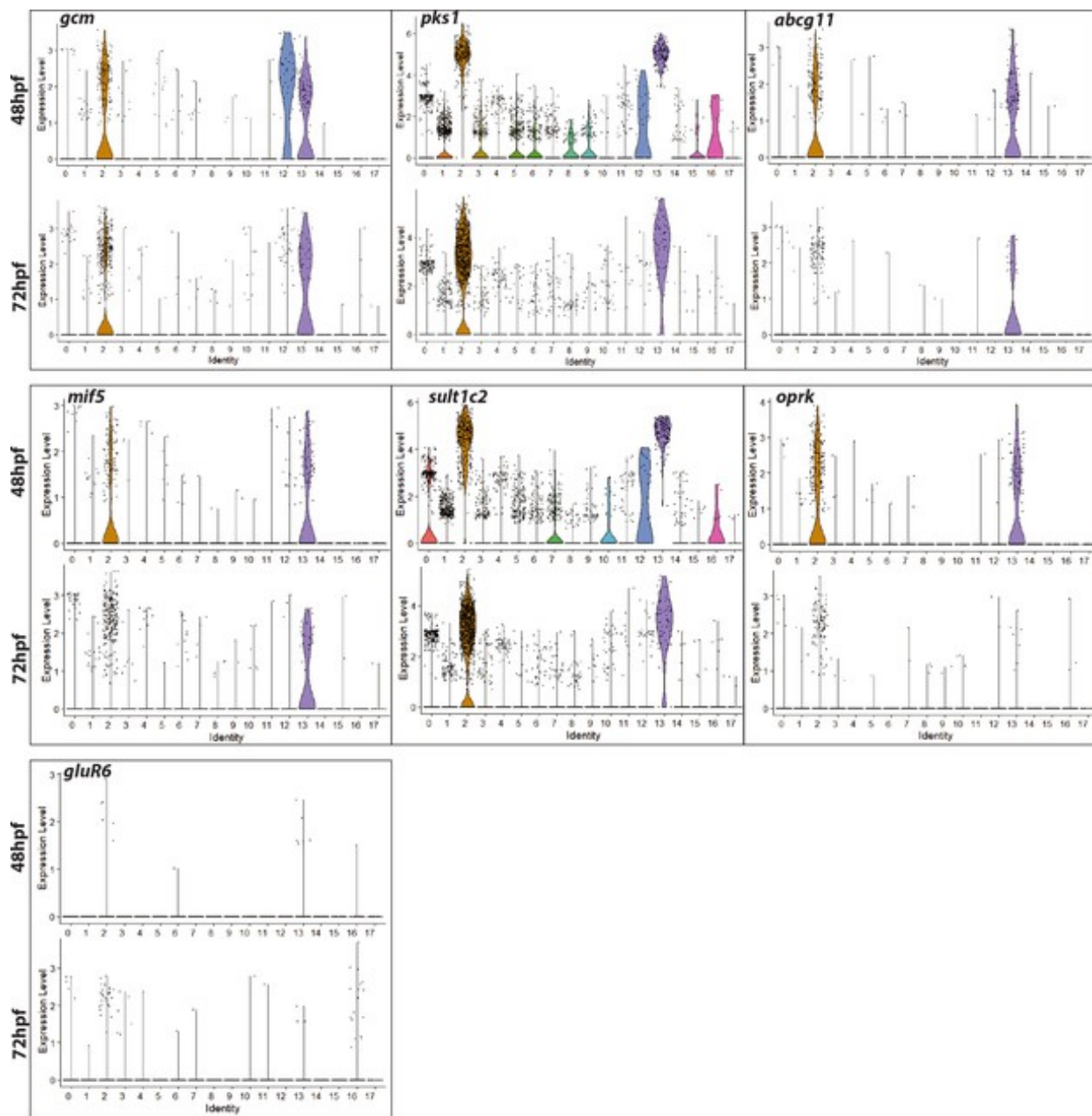
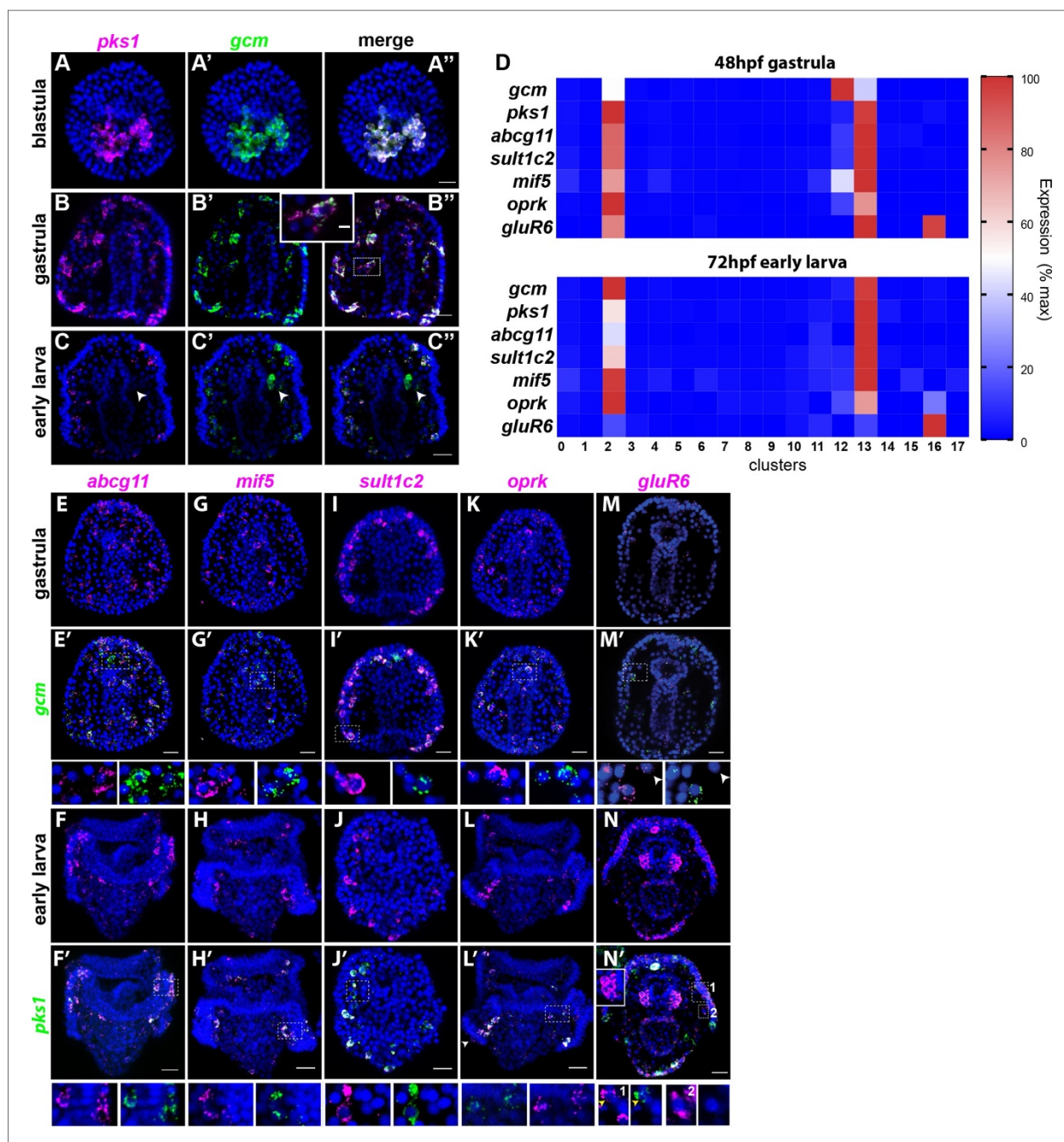


Figure supplement 2. Violin plots showing expression of pigment cell marker genes across different clusters at gastrula and early larval stage.



expression in cluster 12 decreased and also the expression of the other pigment genes decreased. GluR6 and oprk are also expressed in cluster 16, ciliary band neurons. (E–N') Double FISH of identified pigment genes with either gcm or pks1 to mark pigment cells in gastrulae and larvae, respectively. Dashed line boxes indicate cells magnified below each picture. Violin and feature plots for the pigment cell genes are shown in Figure 3—figure supplement 2. (M–M'). GluR6 expression in gastrulae is low. In larvae, glur6 is expressed in pigment cells (yellow arrowheads in J' inset 1) and in neurons (cluster 16) that appear to be lateral ganglions (inset 1 and 2 white arrowheads in N'). GluR6 is also expressed in both coelomic pouches (inset in N'). Nuclei are depicted in blue (DAPI). Scale bars 50 mm, inset in B'-B'' scale bar 20 mm. All images are stacks of merged confocal Z sections.

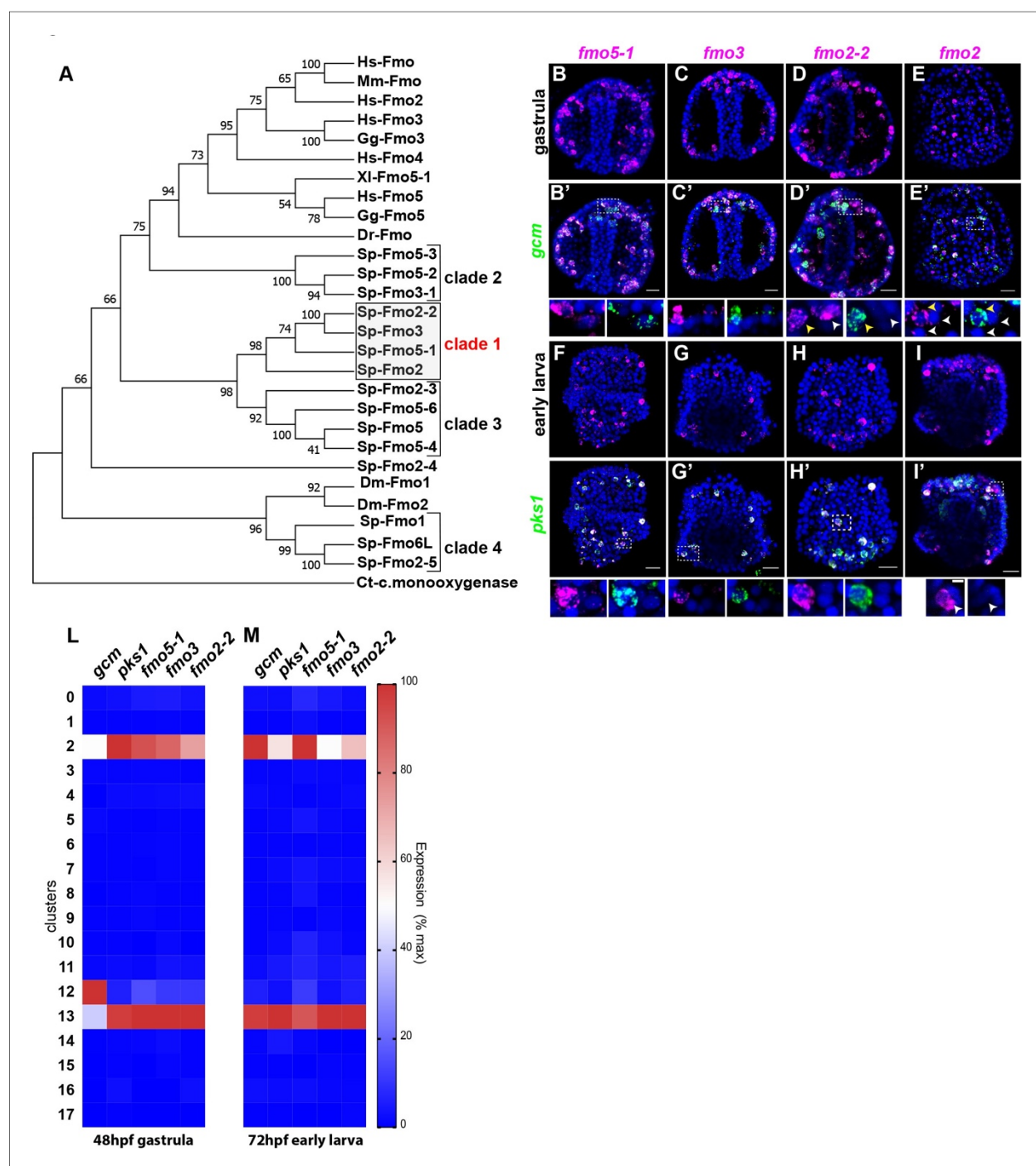


Figure 4. A unique group of *fmo* genes is specifically expressed in pigment cells. (A) Maximum Likelihood tree showing the relationship of sea urchin flavin-dependent monooxygenases (FMOs) to other FMO proteins. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates (500 replicates) are collapsed. Baeyer–Villiger monooxygenase was used as an outgroup. The same phylogeny was also supported by a maximum parsimony tree. (B–I') Double FISH of *fmos* from clade

one with *gcm* or *pks1*. In the insets, yellow arrowheads indicate co-expression in the pigment cells, white arrowheads indicate cells that are not pigment cells. Dashed line boxes indicate cells magnified below each picture. Nuclei are depicted in blue (DAPI). Scale bars 50 μ m in whole embryo images and 20 μ m in magnifications. All images are stacks of merged confocal Z sections. (L–M) single cell data show that the fmos of clade one are found in clusters 2 and 13.

The online version of this article includes the following figure supplement(s) for figure 4:

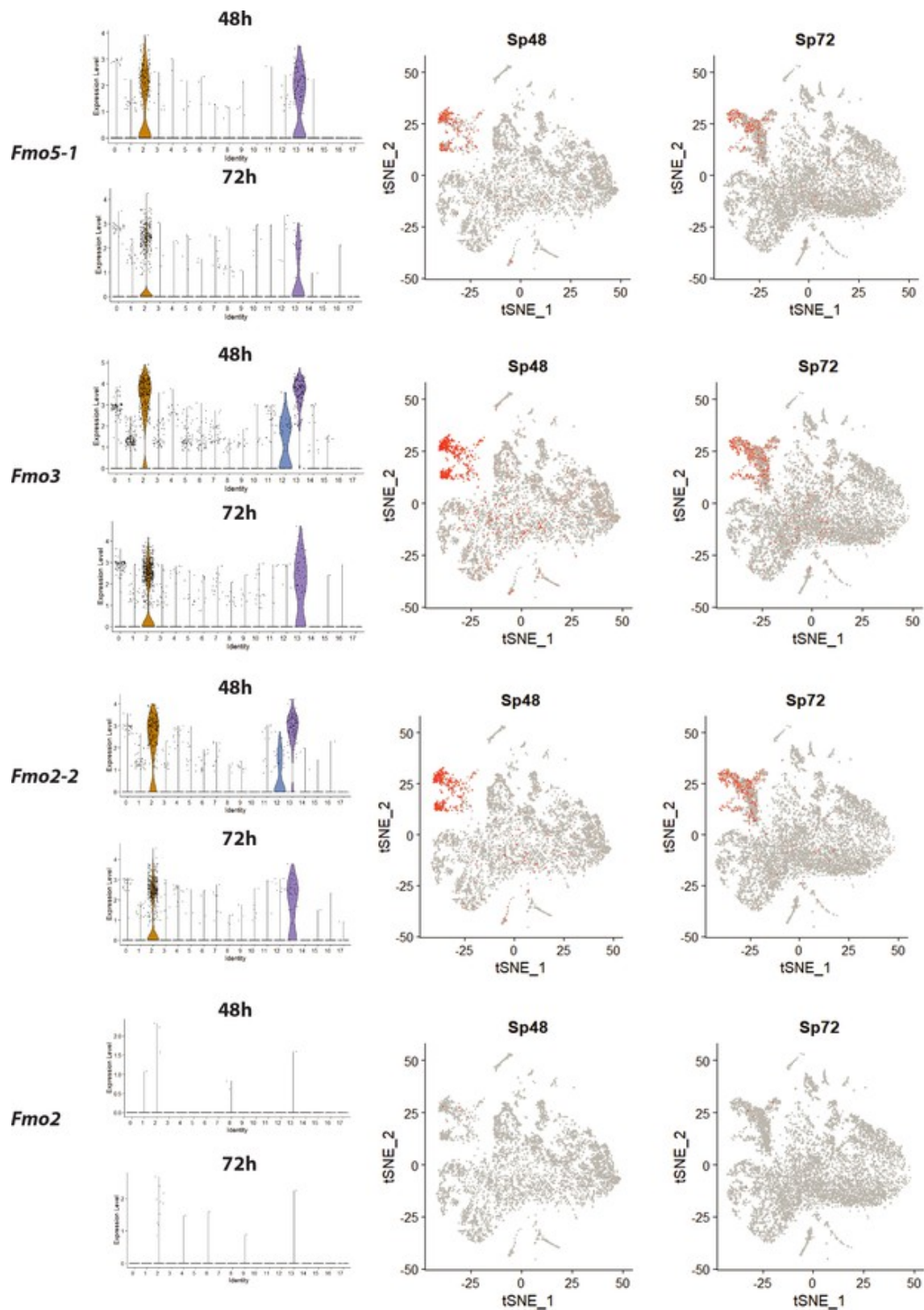


Figure supplement 1. Distribution of fmos in the single cell transcriptome shows high fmo gene expression in cluster 2, 12 and 13.

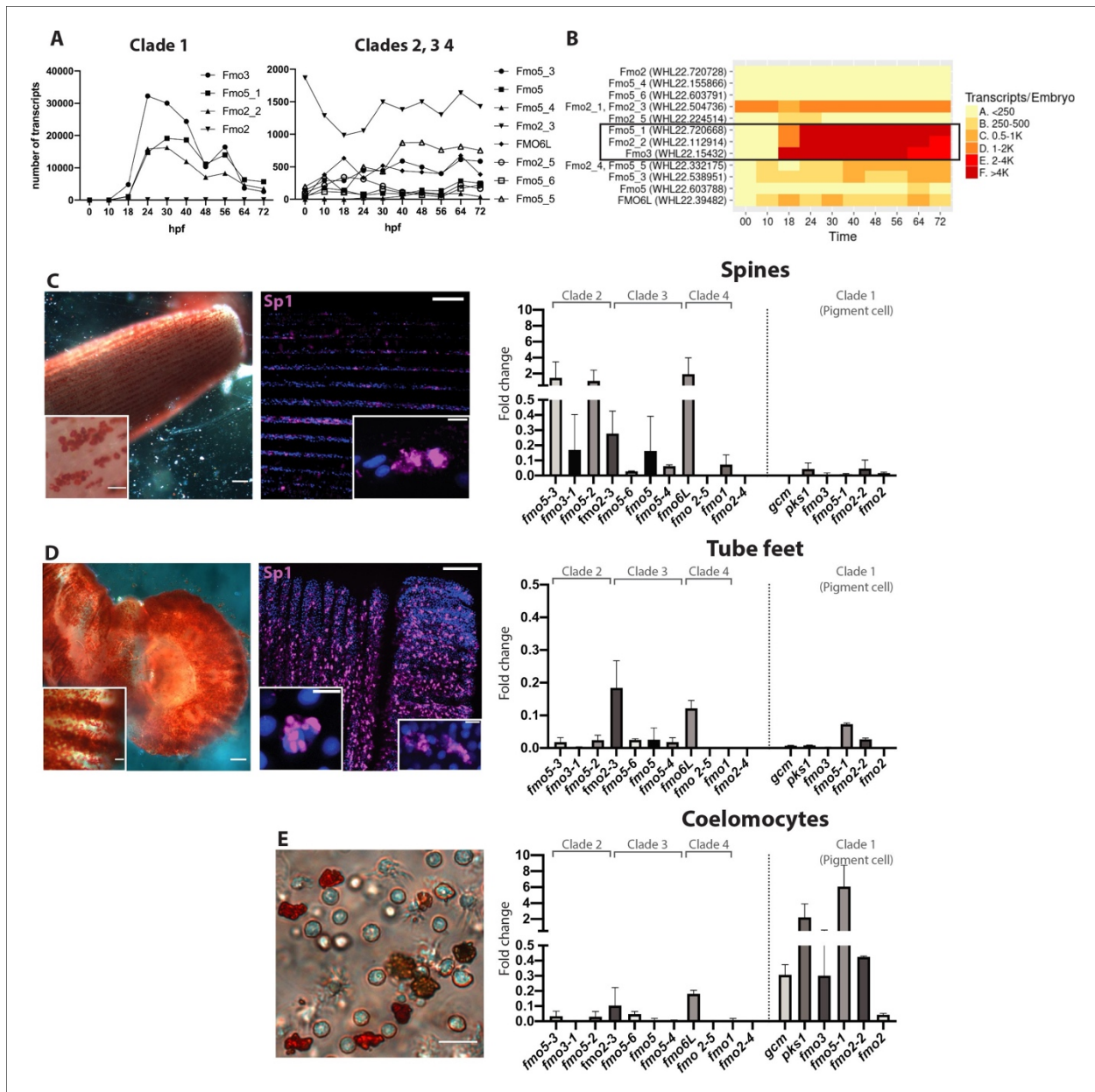


Figure 5. Fmos from clade one are expressed in embryonic pigment cells and in adult coelomocytes. (A, B) Transcripts abundance of the four fmo clades in the embryo, 0 to 72 hr post fertilization (hpf). This graph was made using the online resource <http://www.spbase.org:3838/quantdev/>. All fmos in clade 1 (fmo5-1, fmo2-2, fmo3, squared box) and fmo2-3 from clade three are expressed in the embryo. A and B represent the same data. (C, D, E) Spine, tube feet and coelomocytes bright field images show red pigmented cells. Scale bars are 100 mm, magnification scale bars 20 mm. Confocal images of spines and tube feet labeled with the pigment cell marker Sp1, showing that adult structures have common features with embryonic pigment cells. Nuclei are in blue (DAPI); scale bar 100 mm and 5 mm for insets. qPCR of fmos in the adult spines (C), tube feet (D) and coelomocytes (E) show that clade one is expressed in coelomocytes. Gcm and pks1 are expressed in the coelomocytes, but not in spines or

tube feet. All experiments were repeated in three biological replicates. Fold change = $2^{(-Dct)}$. FISH of genes not in clade one is in Figure 5—figure supplement 1.

The online version of this article includes the following figure supplement(s) for figure 5:

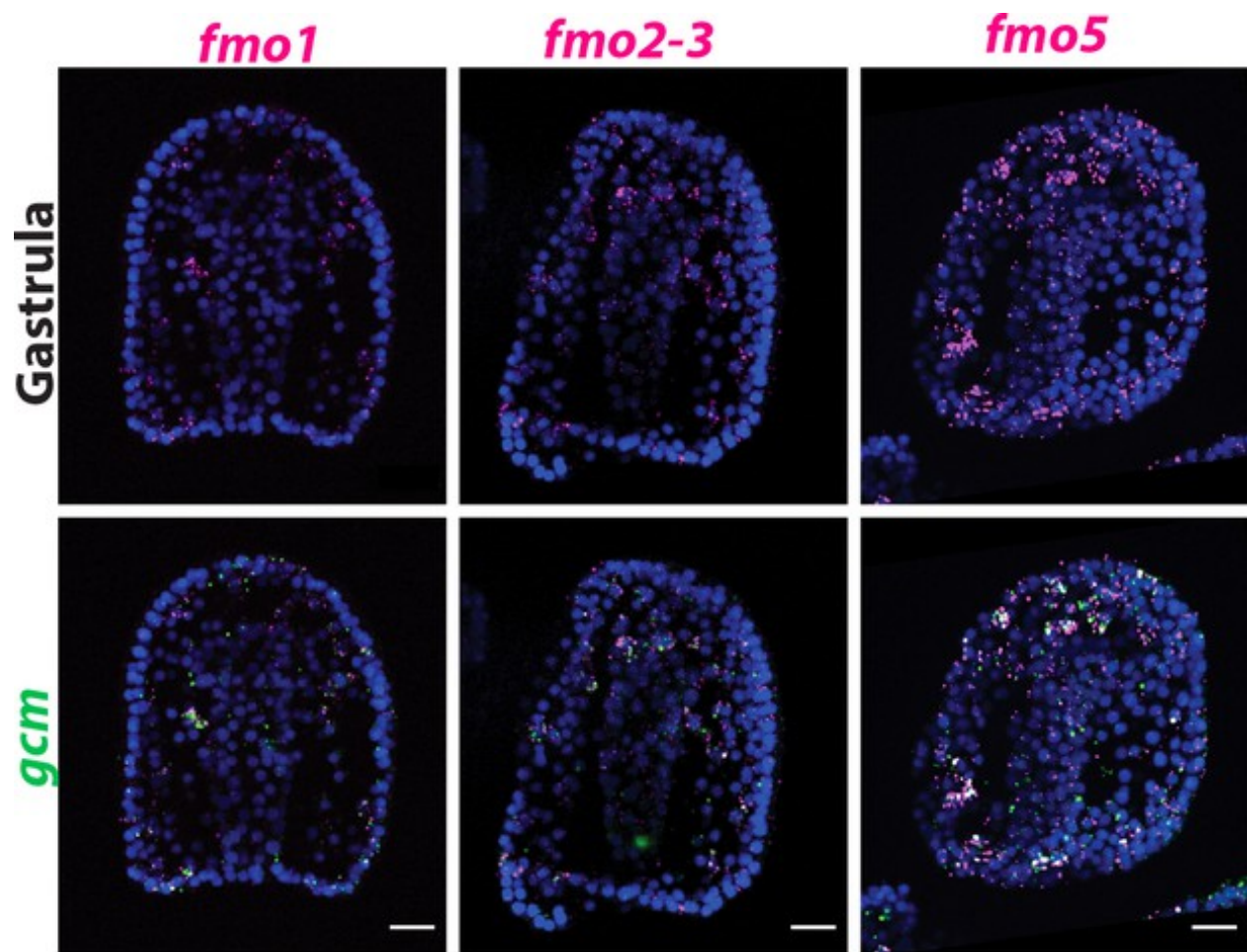


Figure supplement 1. Spatial expression of other *fmo* genes that are not present in clade 1.

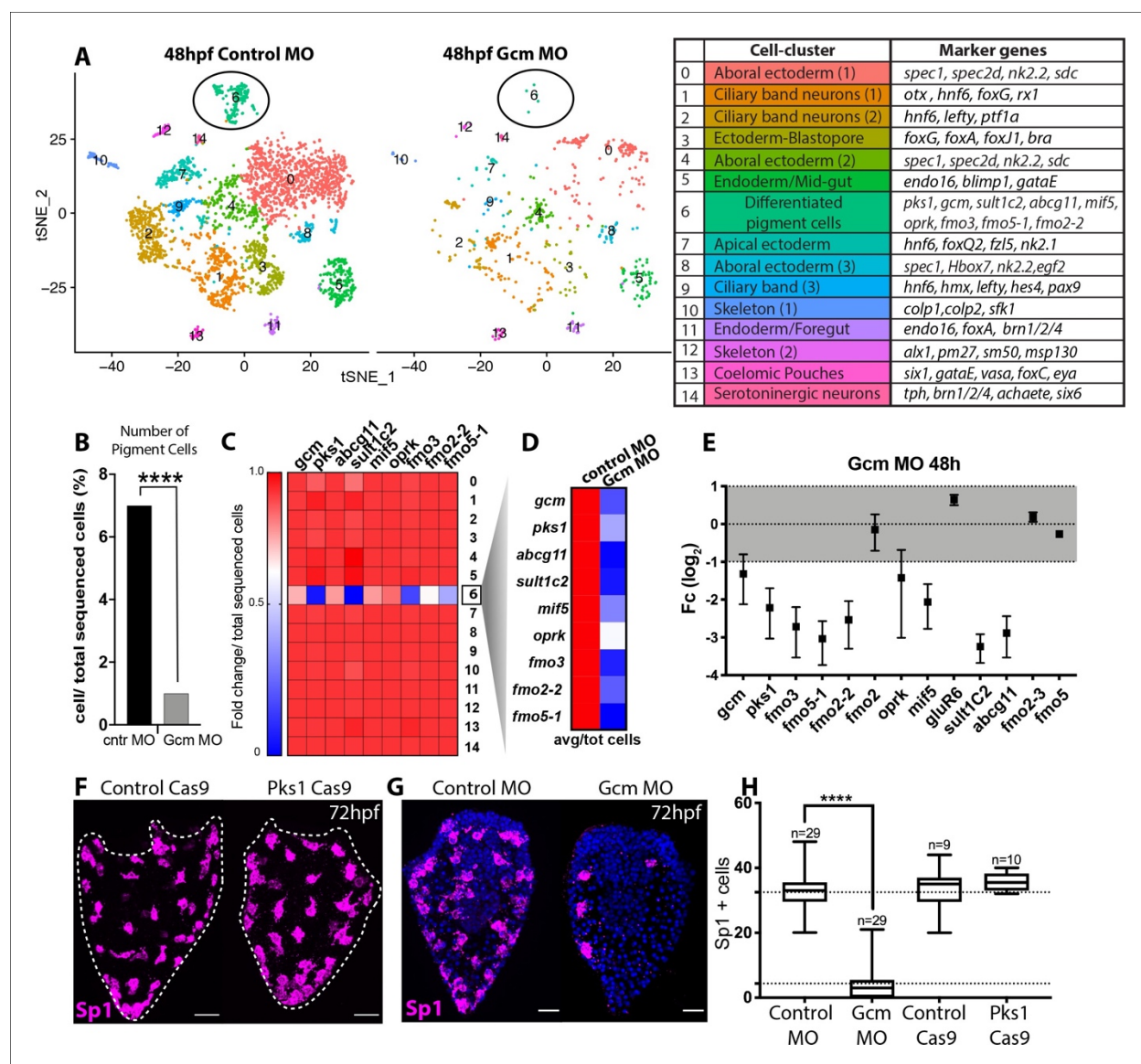
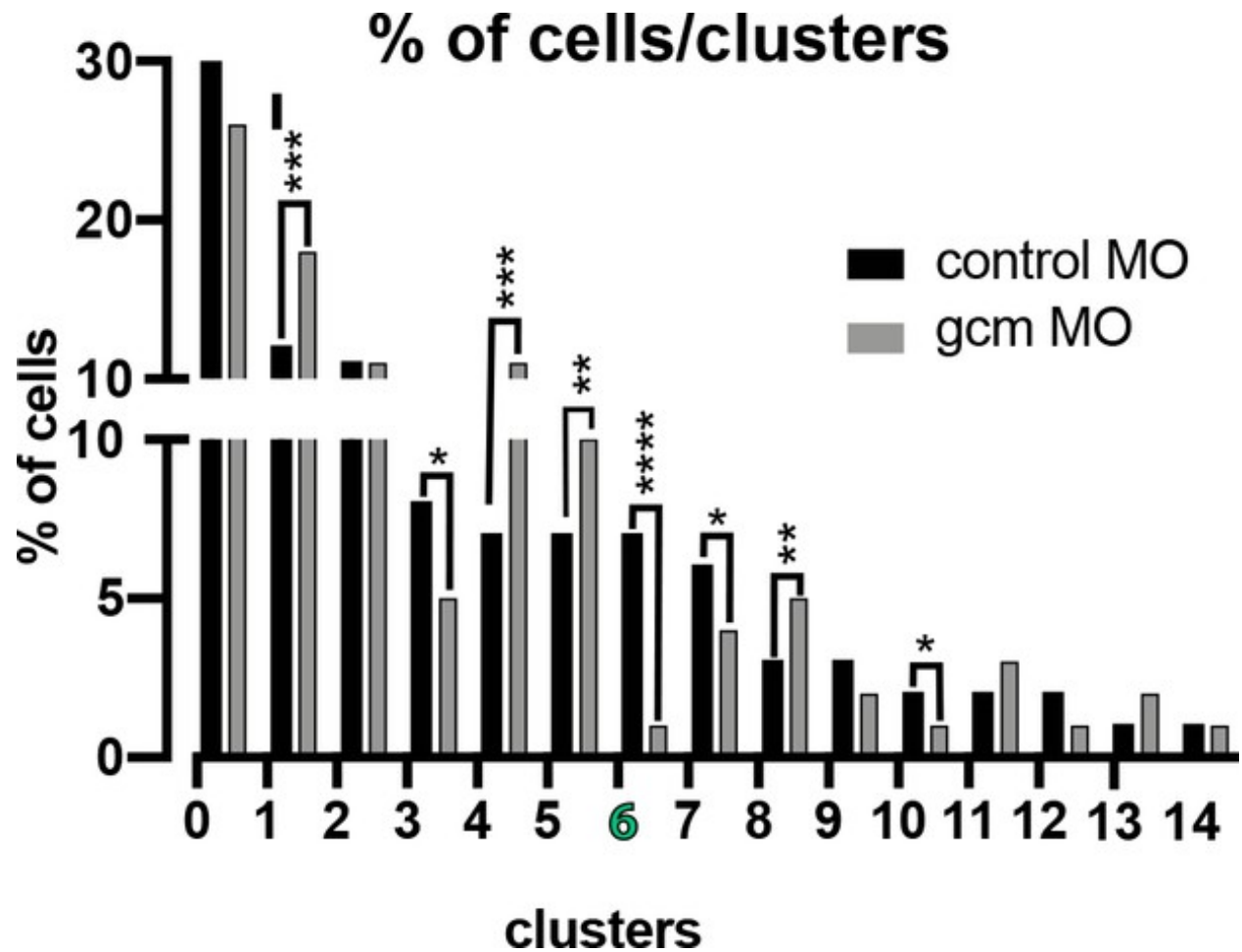


Figure 6. Gcm controls pigment cell specification. (A) Global visualization of single cell profiled in control MO and gcm MO gastrulae. tSNE plots show that in controls and morphants there is no change in general cluster organization. Pigment cell cluster is number 6, green. The table summarizes the main cell-types and relative marker genes, colors represent clusters in the tSNE plots. In the 48hpf control morpholino injected sample, 3402 single cells were captured for downstream analysis, sequenced at a level of 56,550 reads per cell. The 48hpf gcm morpholino sample includes 643 single cells with an average of 379,895 reads per cell. The two samples were integrated to identify conserved cell types and cluster marker genes using Seurat. Feature and violin plot of pigment cell genes in controls and gcm depleted embryos are shown in Figure 6—figure supplement 2 and Figure 6—figure supplement 3. (B) Percentage of Pigment Cells relative to total sequenced cells in control MO versus gcm MO experiments, two-sided Chi-square and Fisher's exact test gave $p < 0.001$ (****); Chi-square df 31,35, 1, z 5.599. A graph with percentage of cells relative to total sequenced cells for each cluster is

shown in Figure 6—figure supplement 1. (C) Expression of pigment genes in gcm MO (Fold change, Fc) compared to control MO normalized by number of total sequenced cells in each experiment (average LogFc*number of cells in each cluster/total sequenced cells in each experiment). (D) Expression levels of pigment genes normalized for the total sequenced cells shows that there is an overall decrease in gene expression in the pigment gene clusters. (E) qPCR of pigment genes in gcm MO gastrulae. Experiments were performed in three biological replicates (error bars). (F) Sp1 Ab staining of 72hpf larvae depleted of pks1 by CRISPR Cas9 shows that distribution of pigment cells does not change. A qPCR of pigment genes in control and pks1 depleted embryos is shown in Figure 6—figure supplement 4. (G) Sp1 Ab staining in gcm MO early larva (72hpf). Nuclei are stained with DAPI. (H) Number of Sp1 positive cells decreases in gcm MO compared to controls (n = number of larvae, two-tailed student t-test $p < 0.001$ (****) corresponds to $p = 1.44704 \times 10^{-17}$; 95% confidence interval of -31,07 to -25,21), (mean for control MO is 35,51; mean for gcm MO is 4.37), and it is unchanged in pks1 Cas9 (two-tailed student t-test p value 0,4191; 95% confidence interval of -3,326 to 7,393; mean of control Cas9 is 33,67, mean of pks1 Cas9 is 35,70). Additional file in Figure 6—source data 1. Images are confocal Z-projections, scale bar 50 μ m.

The online version of this article includes the following source data and figure supplement(s) for figure 6:



Source data 1. Number of Sp1+ cells in gcm MO, control MO, pks1 Cas9 and Cas9 controls.

Figure supplement 1. Number of cells in each cluster in control embryos versus embryos depleted of Gcm.

Figure supplement 1—source data 1.

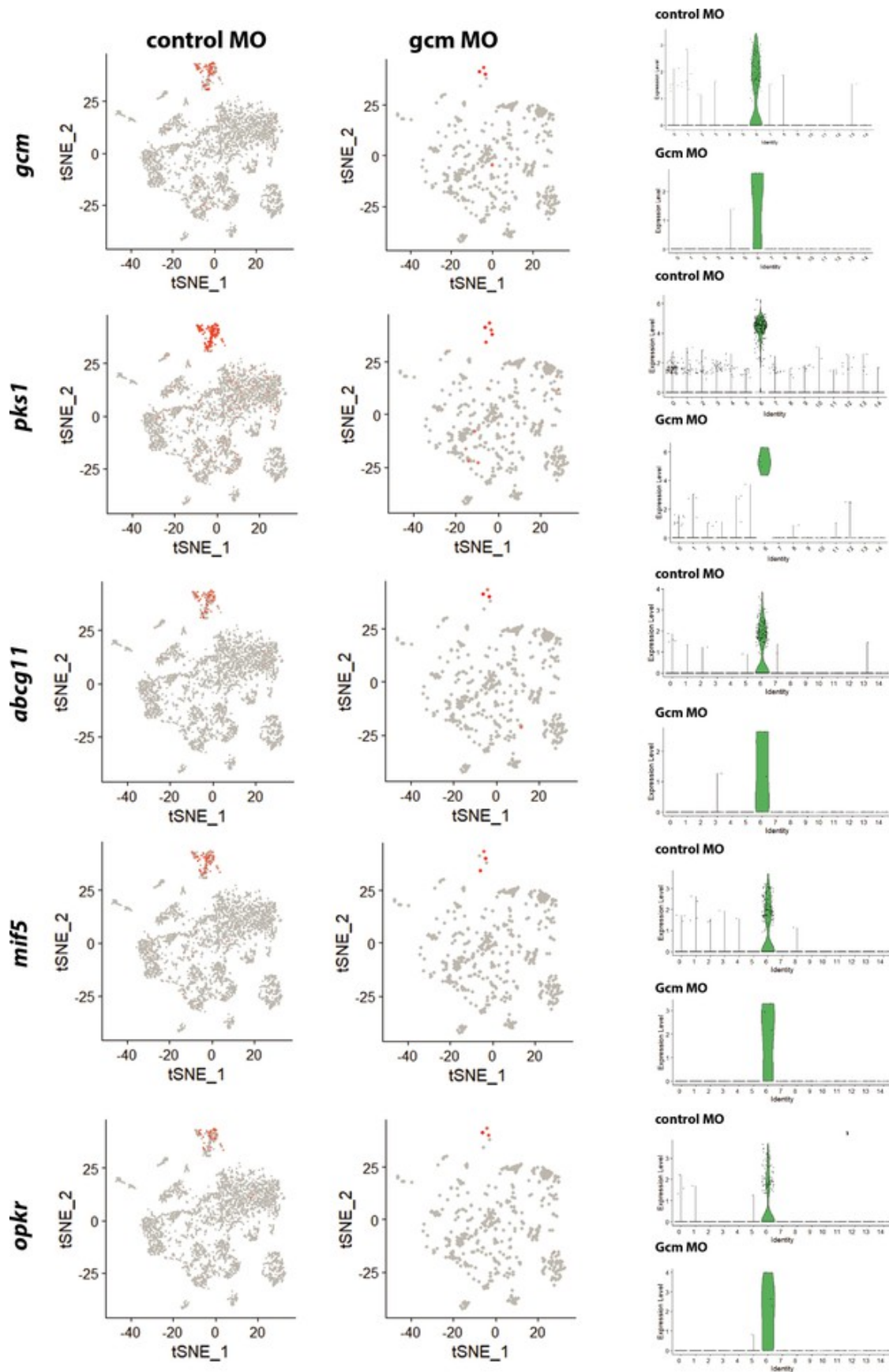


Figure supplement 2. Distribution of pigment genes in single cell transcriptomes of control and gcm depleted embryos. Data for each gene are shown with a feature plot that provides a clear overview of single cells expressing that unique transcript in the context of all other clusters (left), and violin plot that clearly shows the expression levels of that specific transcript for each cluster (right).

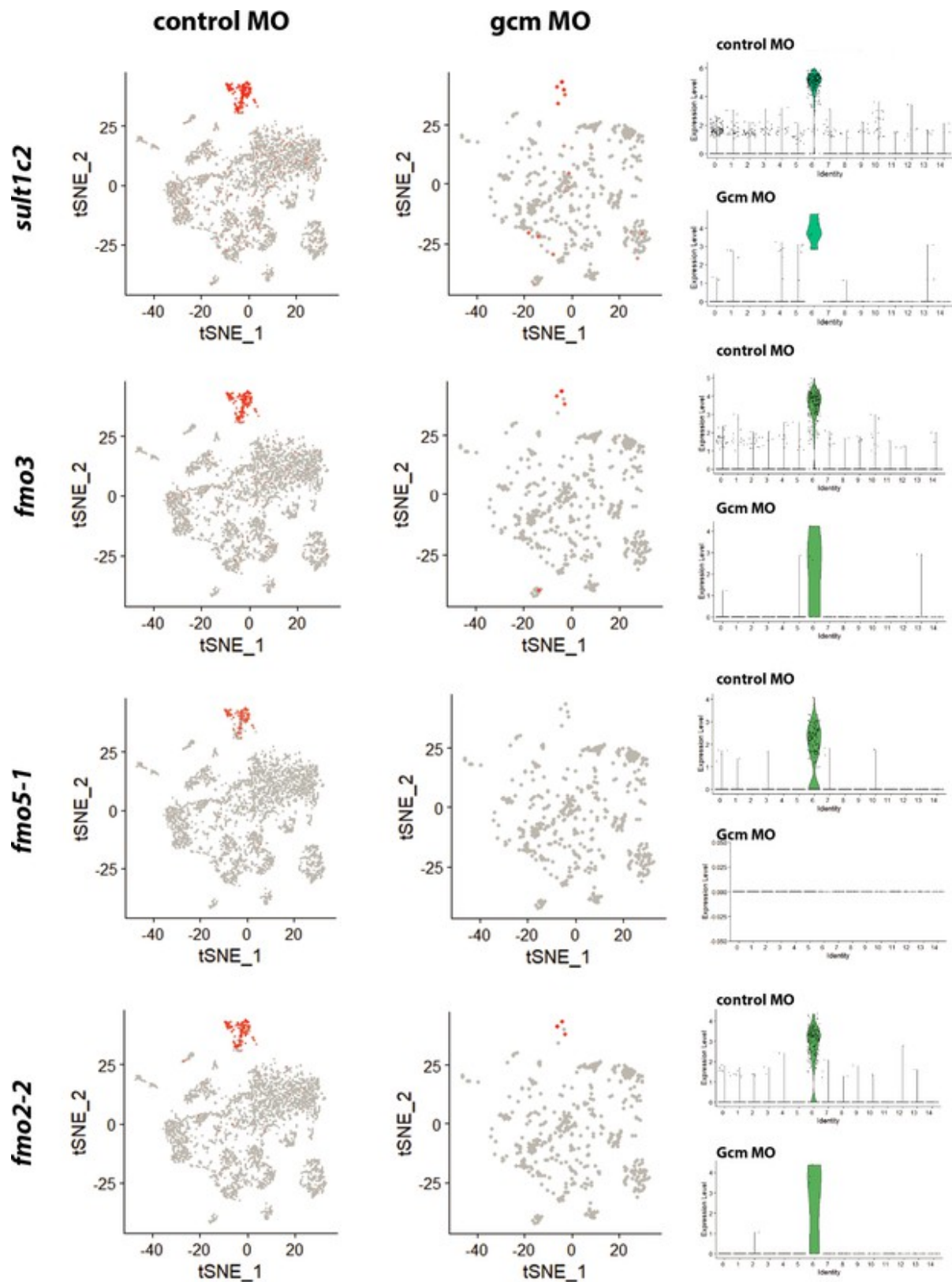


Figure supplement 3. Distribution of pigment genes in single cell transcriptomes of control and *gcm* depleted embryos.

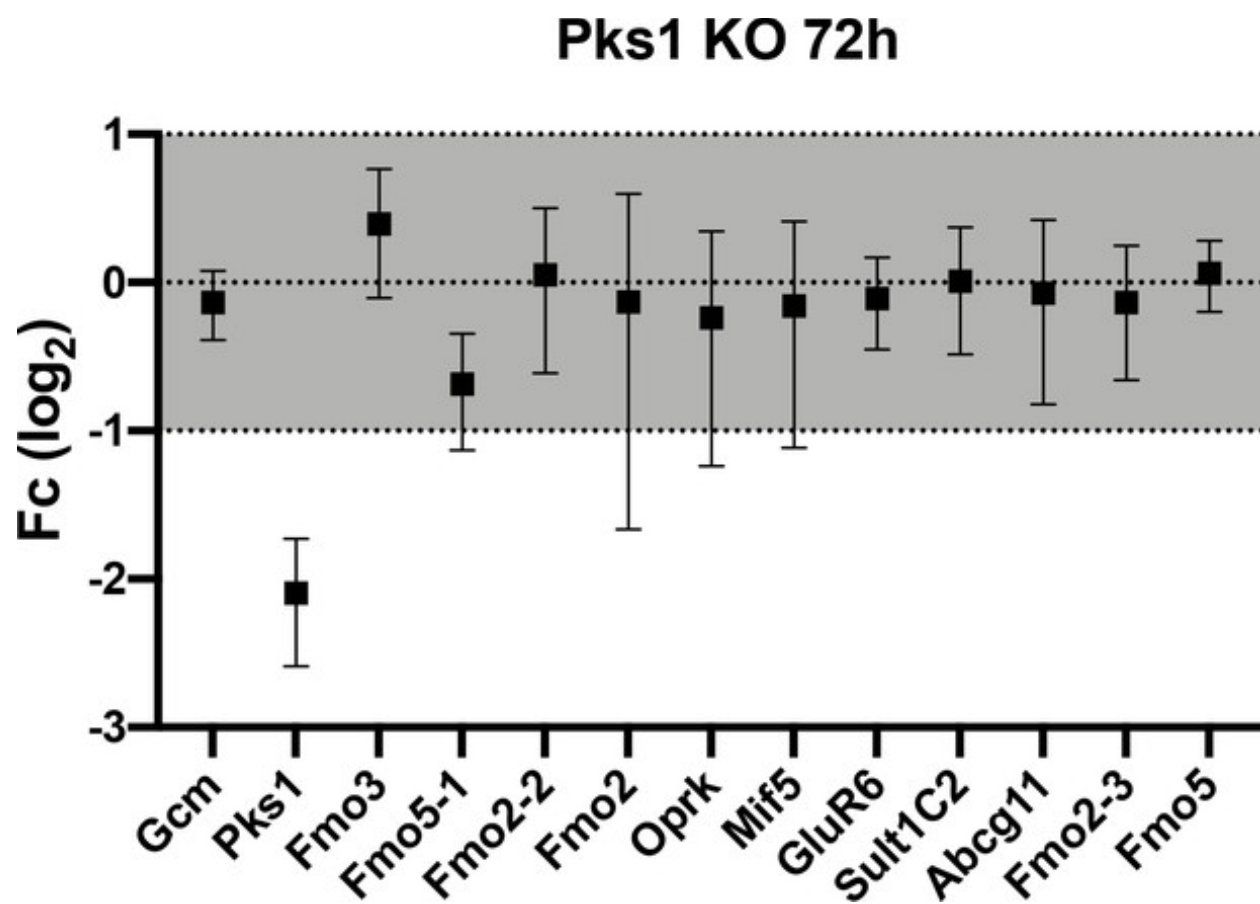


Figure supplement 4. Gene expression of candidate genes in *pks1*- larvae is indistinguishable from controls.

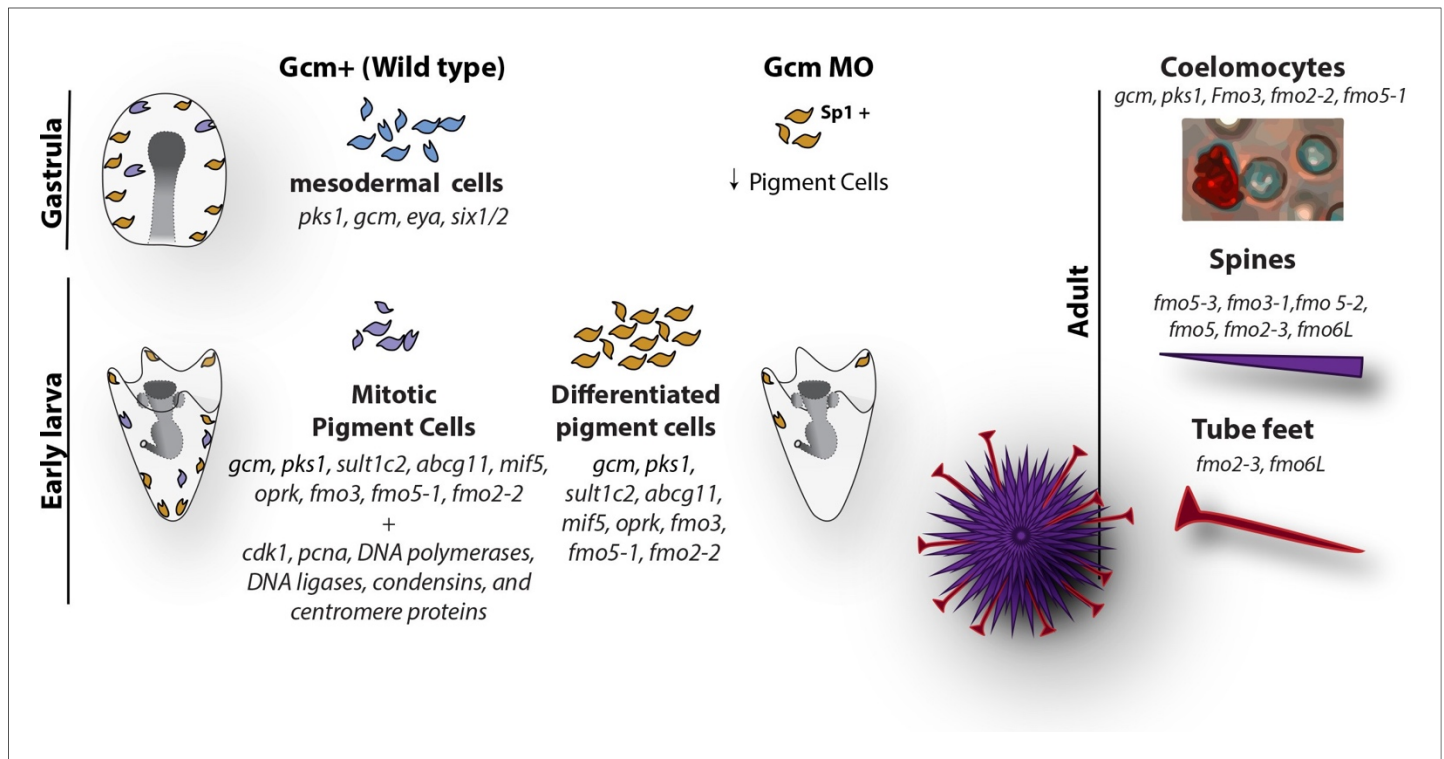


Figure 7. Diagram summarizing the fate of pigment cells in wild type and gcm MO embryos. In wild type embryos there are 3 subpopulations of cells expressing gcm: mesodermal cells not yet differentiated, differentiated pigment cells and mitotic pigment cells. Colors are derived from the clusters of Figure 1. Larvae have both differentiated and mitotic pigment cells. In embryos where gcm is depleted there are fewer pigment cells. Adult sea urchins express different levels of fmo genes in their pigmented structures. Fmos that are expressed in the embryonic pigment cells are also expressed in the adult red spherule cells, both groups of cells having immune functions.

Appendix C: A single-cell RNA-seq analysis of Brachyury-expressing cell clusters suggests a morphogenesis-associated signal center of oral ectoderm in sea urchin embryos

Noriyuki Satoh, Kanako Hisata, Stephany Foster, Shumpei Morita, Koki Nishitsuji, Nathalie Oulhen, Hitoshi Tominaga, Gary M. Wessel

This project was a collaboration between the Satoh and Wessel Labs. The dataset used in this manuscript was the one I produced for early sea urchin embryogenesis (Chapter 3). I contributed to this project through some reanalysis of that dataset, contributing to discussion regarding the analysis and writing.

Abstract

Brachyury is a T-box family transcription factor and plays pivotal roles in morphogenesis. In sea urchin embryos, Brachyury is expressed in the invaginating endoderm, and in the oral ectoderm of the invaginating mouth opening. The oral ectoderm is hypothesized to serve as a signaling center for oral (ventral)-aboral (dorsal) axis formation and to function as a ventral organizer. Our previous results of a single-cell RNA-seq (scRNA-seq) atlas of early *Strongylocentrotus purpuratus* embryos categorized the constituent cells into 22 clusters, in which the endoderm consists of three clusters and the oral ectoderm four clusters (Foster et al., 2020). Here we examined which clusters of cells expressed Brachyury in relation to the morphogenesis and the identity of the ventral organizer. Our results showed that cells of all three endoderm clusters expressed Brachyury in blastulae. Based on expression profiles of genes involved in the gene regulatory networks (GRNs) of sea urchin embryos, the three clusters are distinguishable, two likely derived from the Veg2 tier and one from the Veg1 tier. On the other hand, of the four oral-ectoderm clusters, cells of two clusters expressed Brachyury at the gastrula stage and genes that are responsible for the ventral organizer at the late blastula stage, but the other two clusters did not. At a single-cell level, most cells of the two oral-ectoderm clusters expressed organizer-related genes, nearly a half of which coincidentally expressed Brachyury. This suggests that the ventral organizer contains Brachyury-positive cells which invaginate to form the stomodeum. This scRNA-seq study therefore highlights significant roles of Brachyury-expressing cells in body-plan formation of early sea urchin embryos, though cellular and

molecular mechanisms for how Brachyury functions in these processes remain to be elucidated in future studies.

1. Introduction

Sea urchin embryos provide a model experimental system to explore gene regulatory networks (GRNs) responsible for specification and differentiation of early embryonic cells (Davidson, 2006; McClay, 2011). Detailed descriptions of embryonic cell lineages, experimental manipulations of embryos, decoded genomes, a list of well-characterized genes encoding transcription factors and signaling pathway molecules, observation of spatio-temporal expression of genes, and functional manipulation of genes, have created sophisticated approaches to explore details of interaction of GRN component genes in each embryonic territory (EchinoWiki; <https://wiki.echinobase.org>).

A recently developed technology of single cell RNA sequencing (scRNA-seq) provides a powerful tool to categorize genes that are expressed in constituent cells of embryos, tissues, or organs on a cell-by-cell basis (e.g., Stuart and Satija, 2019; Cao et al., 2019). In a previous study, we reported a scRNA-seq resource for early sea urchin development (Foster et al., 2020), in which early embryos of *Strongylocentrotus purpuratus* were interrogated at eight developmental stages, 8-cell, 64-cell, morula, early blastula, hatched blastula, mesenchyme blastula, early gastrula and late gastrula. This resource includes mRNA information in embryonic cells delineated into 22 clusters: two clusters for the neural lineage, one for germline, four for oral ectoderm, two for aboral ectoderm, two for the skeleton originating from the primary mesenchyme cells (PMCs), one for secondary mesenchyme cells (SMCs), one for pigment cells, two for ciliated cells, and three for endoderm (Fig. 1A and B) (Foster et al., 2020). Each cluster possesses marker gene expression: the ectoderm clusters were identified by

expression of SoxB2, Foxq2, NK2.1 and AnkAT-1; the endoderm clusters by FoxA and Endo16; the mesodermal PMCs by Alx1, SM50 and SM37; SMCs by Six1/2 and Eya; and the germ cell cluster by Nanos2, Vasa and Seawi (Foster et al., 2020). In addition, scRNA-seq resources for 48-hpf gastrula and 72-hpf larva are also available (Perillo et al., 2020).

Taking advantage of this scRNA-seq resource, the present study aimed to identify coordinate gene activity for transcription factors and signaling molecules that may be involved in early morphogenesis of sea urchin embryos. We sought to test coordinate expression of the transcription factor member of the T-box family, Brachyury, with key signaling molecules to posit functional organizer testing in the future. Brachyury (Bra) is expressed in the invaginating endoderm and in the oral ectoderm (Gross and McClay, 2001; Croce et al., 2001; Peter and Davidson, 2010) and suppression of Bra function resulted in a failure of gastrulation and gut formation (Gross and McClay, 2001; Rast et al., 2002).

Here we test the following: (1) The sc-RNAseq analysis identified three clusters of endoderm cells in late gastrulae (Foster et al., 2020). Do cells of all three clusters express SpBra (Brachyury of *S. purpuratus*)? The endoderm of sea urchin embryos consists of cells from two different origins, Veg1 and Veg2 (e.g., Ransick and Davidson, 1998; Howard-Ashby et al., 2006). What are these three clusters in relation to the Veg1 and Veg2 lineages? The larval digestive system is tripartite, composed of a foregut or muscular esophagus, a midgut or large spherical stomach, and a hindgut or short tubular intestine, and various sets of developmentally relevant genes are dynamically expressed in each region (reviewed by Annunziata et al., 2014; Annunziata et al.,

2019). How are the three endoderm clusters related to the tripartite structures of the larval digestive system? (2) The oral ectoderm of sea urchin embryos serves as a signaling center or ventral organizer for oral (ventral) - aboral (dorsal) axis formation (Duboc et al., 2004; Lapraz et al., 2009, 2015). The oral (ventral) ectoderm expresses nodal, BMP2/4, chordin and ADMP1 whereas ADMP2 is expressed in the aboral (dorsal) ectoderm (Duboc et al., 2004, 2008; Lapraz et al., 2009, 2015; Wei et al., 2012). The embryonic oral ectoderm consists of two parts with different properties, one remains at the outer surface to form the oral epidermis and the other invaginates to form the mouth opening (stomodeum) (Su et al., 2009; Li et al., 2013; Materna et al., 2013). The oral ectoderm was shown to consist of four cell clusters by the scRNA-seq analysis (Foster et al., 2020). Which of these clusters represent the oral epidermis and which ones represent the stomodeum? Is the ventral organizer a property of the surface ectoderm cells or of cells invaginating to form the stomodeum, or both? Results of our present analyses provide insights into molecular mechanisms involved in the body plan formation of early sea urchin embryos.

2. Results

2.1. SpBra is expressed in cells solely of the endoderm and oral ectoderm

Of 22 (0–21) clusters (Fig. 1A), Foster et al. (2020) have assigned clusters 6, 8 and 14 to be endoderm and those 1, 4, 12 and 13 to be oral ectoderm (Fig. 1B). We found that five out of the seven clusters contained cells with significant SpBra expression (Fig. 1C). These clusters are 1, 6, 8, 12 and 14, the latter four of which

showed high level of SpBra expression. Clusters 6, 8, and 14 are endoderm (Fig. 1B), indicating that all three endoderm clusters express SpBra in blastula stages (Figs. 1C and 2). On the other hand, two of the four oral-ectoderm clusters, 12 and 1, expressed SpBra in later blastula stage and gastrula stage (Figs. 3 and 4), whereas the two others, 4 and 13, did not express SpBra (Figs. 3 and 4). We examined whether other clusters of cells also express SpBra, especially PMCs, SMCs and aboral ectoderm. PMCs give rise to the larval skeleton and are represented by clusters 16 and 19 (Fig. 1B), based on the fact that cells of the clusters express several skeletogenic marker genes, including *Alx1*, *SM50* and *SM37* (Foster et al., 2020), and *msp130* and *SM30* (Perillo et al., 2020). The present analysis confirmed the expression of these markers in clusters 16 and 19 (*msp130* in Fig. 1C; Data in Brief Tables 1–3) but the expression of SpBra was at undetectable level in the two PMC clusters (Figs. 1C and 4). Pigment cells originate from SMCs and are represented by cluster 11 based on the expression of *gcm*, *six1/2* and *pkx1* in the cells (Foster et al., 2020; Perillo et al., 2020). The present independent analysis also found that cells of cluster 11 expressed *pkx1* (Fig. 1C), *six1* and *gcm* (Data in Brief Tables 1 and 2). *Six1* was reported to be expressed in mesodermal domain of the tip of the archenteron (Andrikou et al., 2013) and *gcm* in cells of pigment lineage (Calestani and Rogers, 2010), confirming that cluster 11 is of SMC lineage. The expression of SpBra was at undetectable level in this cluster (Figs. 1, 2 and 4). Whole mount in situ hybridization (WMISH) analyses in previous research showed SpBra expression in SMCs (Harada et al., 1995; Peterson et al., 1999). Therefore, results of the present scRNA-seq analysis were inconsistent with these earlier observations, but recent WMISH analyses showed that SpBra is not expressed in SMCs (Peter and

Davidson, 2010; S. Yaguchi, personal communication). Clusters 0 and 5 are assigned to be aboral ectoderm (Fig. 1B), since cells of these clusters express SPEC1, Hox1, Nk2.2, spec2c, IrxA, and EGF2 (Data in Brief Tables 1–5) (Perillo et al., 2020). The expression of SpBra was at undetectable level in clusters 0 and 5 (Figs. 1C and 3). Therefore, we concluded that SpBra is expressed in cells of only endoderm and oral ectoderm.

2.2. Property of the three endoderm clusters

Studies of GRN in sea urchin embryos show that (a) the Veg2 tier includes Veg2-Endo and Veg2-Meso (SMC), the latter of which occupies the vegetal pole (Fig. 5A), (b) the Veg1 tier forms Veg1-Endo and Veg1-Ecto, the latter of which is adjacent to the animal ectoderm (Fig. 5A), (c) genes involved in Veg2-Endo specification include hox11/13b, eve, blimp1/ krox, foxA, and gata-e (Fig. 5B; wnt genes that act up-stream of these genes are not discussed here) (e.g., Sethi et al., 2012; Cui et al., 2014; Li et al., 2014; Erkenbrach, 2016; Echinobase, https://www.echinobase.org/common/jsp/showWiki.jsp?Davidson_Lab_Gene_Regulatory_Networks; Ettensohn, 2020), (d) genes involved in Veg1-Endo specification include hox11/13b, eve, and hnf1 (Sethi et al., 2012; Cui et al., 2014). Since our scRNA-seq studies (Foster et al., 2020; this study) demonstrated that the endoderm is comprised of three clusters (6, 8 and 14), it is likely that two of them belong either to Veg2-Endo or Veg1-Endo. Therefore, taking these facts in mind, we examined which transcription factor genes are expressed in each of clusters 6, 8 and 14 (Figs. 1C and 2).

We found two discrete profiles of gene expression in the three clusters, namely, clusters 14 and 6 exhibited a similar profile, and cluster 8 a different profile (Fig. 2). Expression profiles of *SpBra*, *hox11/13b* and *eve* were similar among the three clusters except at early blastula stage, although cluster 8 retained expression of the three genes in later stages (Figs. 1C and 2; Figs. S1 and S2). Although Cui et al. (2014) showed that *hnf1* is expressed in Veg1-Endo but not Veg2-Endo, this study showed that *hnf1* is expressed in the three clusters (Fig. 2). On the other hand, *blimp1/krox*, *foxA*, and *gata-e* were highly expressed in cells of clusters 14 and 6 but not in cluster 8 (Fig. 2; Fig. S1). Since *blimp1/krox*, *foxA*, and *gata-e* are not included in Veg1-Endo GRN (see Fig. 5B) and since this pattern was seen only in cluster 8, we concluded that cluster 8 corresponds to Veg1-Endo. Accordingly, clusters 14 and 6 correspond to Veg2-Endo (Fig. 5B). The transcription factors of the Ets family are posited as activators of *SpBra* (Rast et al., 2002; Oliveri et al., 2006; Li et al., 2013). Since this scRNA-seq analysis showed that *Ets1* and *Ets4* expression in blastulae and gastrulae are quite low, it is not clear whether Ets could be an activator.

In addition to the transcription factor genes mentioned above, *Endo16*, which encodes a large, secreted protein of the embryonic and larval midgut (Nocente-McGrath et al., 1989; Soltysik-Espanola et al., 1994), has been used as an endoderm and archenteron marker (EchinoWiki) (Ransick et al., 1993; Yuh et al., 2001). We detected high levels of *Endo16* gene expression in clusters 14 and 6 at early and late gastrula stages, although mesenchyme blastula stage showed detectable level of expression (Fig. 1C2 and 2; Fig. S1). In contrast, *Endo16* expression was at lower levels in cluster 8 (Figs. 1C and 2), supporting the notion in which Veg2-derived clusters 14 and 6 are

major components of the gut. A point to be mentioned here is that Endo16 was expressed at high levels in cells of cluster 11 at blastula and early gastrula stages, earlier than that in clusters 14 and 6 (Figs. 1C and 2). Cells of cluster 11 also expressed *gata-e* at early and hatched blastula stages (Figs. 1C and 2). Foster et al. (2020) assigned cluster 11 to be pigment cell-containing SMCs, and another cluster named SMC gut subcluster (Fig. 1B). This suggests that cluster 11 contains cells of the SMCs and a part of them forms the anterior-most region of archenteron or gut (Fig. 5). However, cells of cluster 11 did not express *SpBra*, though low-level and transient expression of *foxA* was seen in early blastulae (Fig. 1C).

Transient expression is a property of *Brachyury*; that is, *Brachyury* is expressed in a circumferential region of presumptive endoderm cells neighboring the vegetal plate (but PMCs and SMCs in the vegetal plate do not express *SpBra*) and its expression becomes undetectable when endoderm cells move inside the gastrula to form the archenteron (Gross and McClay, 2001; Croce et al., 2001). Our scRNA-seq analyses showed that *SpBra* was first expressed at hatched blastula stage, and the high level of gene expression was evident in mesenchyme blastulae. However, the level of expression was gradually lower and lower by early gastrula stage in clusters 6 and 14, although cluster 8 retained the *SpBra* expression at gastrula stage (Figs. 1C and 2). That is, the *SpBra* expression in endoderm cells is transient.

Further survey of structural genes specifically or preferentially expressed in the three endodermal clusters found that *Cyl* and *SLBP* fit the criterion of specific expression in the clusters. *Cyl* encodes a member of various types of cytoplasmic actin (Collura and Katula, 1992; Fang and Brandhorst, 1994), and higher levels of *Cyl*

expression were detected in clusters 6, 14 and 16, and considerable level expression in cluster 8 and 11 (Fig. 2; Fig. S1). Since cells of clusters 14 and 6 are involved in archenteron invagination, these cells likely have the potential of more locomotive activity, and thereby they express high level of cytoplasmic type actin. Aforementioned, cluster 11 is SMC gut subcluster and also involved in gut invagination. SLBP encodes a stem-loop-binding protein (Robertson et al., 2004) and plays roles in histone pre-mRNA processing and cell cycle regulation but its spatial expression remains to be explored. SLBP was expressed in cells of cluster 14 but not 6 and 8 (Figs. 1 and 2; Fig. S1). Therefore, SLBP likely provides a clue to distinguish cluster 14 from clusters 6 and 8. We tentatively assigned cluster 14 to Veg2-Endo1 with SLBP expression and cluster 6 to Veg2-Endo2 without SLBP expression (Fig. 5B, lower). However, the genic property of Veg2-Endo1 and Veg2-Endo2 should be examined in future studies including WMISH analysis, taking two results in mind; firstly, SLBP expression was detected in cells of cluster 4 (oral ectoderm lineage without SpBra expression) (Fig. 1), and secondly, it is unlikely that clusters 14 and 4 neighbor each other based on their distinct profiles (Fig. 1A).

2.3. Co-expression of SpBra with hox11/13b, eve, foxA, gata-e, and Endo 16 at single cell level

SpBra is specifically expressed in cells of endoderm clusters 8, 6 and 14, coincident with several genes encoding transcription factors (Fig. 2). We examined the grade of co-expression of these genes at the single-cell level by three methods. First,

we superimposed t-SNE plots of the two genes, that is, combinations of SpBra and hox11/13b (Supplementary Fig. S3A), SpBra and eve (Fig. S3B), SpBra and blimp1/krox (Fig. S3C), SpBra and foxA (Fig. S3D; Fig. 6A), SpBra and gata-e (Fig. S3E) or SpBra and Endo16 (Fig. S3F; Fig. 6B). In Fig. 6, SpBra-positive cells of clusters 8, 6 and 14 are seen at left-upper part of feature plots, marked with red (left panels). Cells that express foxA or Endo16 are marked by green (middle panels), and therefore, if single cells express both genes simultaneously, they appeared yellow (right panels). These superimposed t-SNE plots showed that single cells co-expressed SpBra and hox11/13b (Fig. S3A), SpBra and eve (Fig. S3B), SpBra and blimp1/krox (Fig. S3C), SpBra and foxA (Fig. 6A) and SpBra and Endo16 (Fig. 6B). Most yellow cells appeared in clusters 14, 6 and 8 (left-upper part of the panels) except for co-expression of SpBra and foxA in clusters 12 and 1 (Fig. 6A; ectoderm clusters, left-lower part of the panels).

Second, the average expression of genes was measured as count data of mRNAs that were normalized to library size and log transformed. Heatmap and clustered matrix were constructed to determine co-expression of SpBra with other genes at single cell level (Supplementary Fig. S4). Results of co-expression of SpBra with foxA or Endo16 are shown in Fig. 7A. If single cells express these genes coincidentally, color-coded bars (yellow, brown and red) appear on the same vertical mass. Several regions with such feature were evident in the three clusters 14, 6, and 8 (Fig. 7A, boxes). On the other hand, the heatmap also showed the presence of cells that expressed high level of foxA but not SpBra (Fig. 7A). Third, we counted the numbers of cells that highly expressed SpBra, foxA or Endo16 with a criterion of $2^>$ or higher standardized expression level, as to each cluster of 14, 6 and 8 at four

embryonic stages (Supplementary Table S2). Then, we examined how many cells co-expressed these genes (Table S2). For example, cluster 14 at hatched blastula stage comprised of 326 cells, in which high levels of expression of SpBra, foxA and Endo16 were detected in 182 (56%), 288 (88%) and 94 (29%) cells, respectively. Of them, 164 (50%) cells simultaneously expressed SpBra and foxA, 32 (10%) cells SpBra and Endo16, and 31 cells (10%) SpBra, foxA and Endo16 (Table S2). All data are presented by a bar graph of Fig. 8. In the endoderm clusters 14 and 6 each, nearly 90% of constituent cells highly expressed foxA at all four stages. On the other hand, cells with SpBra expression occupied nearly a half of constituent cells at hatched blastula stage and the number gradually decreased to 10% at late gastrula stage (Fig. 8). Cells with Endo16 was 10% at hatched blastula stage and gradually increased to 70% at late gastrula stage (Fig. 8). Most of the SpBra-positive cells expressed foxA simultaneously (Fig. 8A). On the other hand, only 20% of SpBra-positive cells expressed Endo16 simultaneously (Fig. 8A). Cells of cluster 6 exhibited a similar expression profile of SpBra, foxA and Endo16 (Fig. 8B). Cells of cluster 8 also showed simultaneous expression of the three genes, although the quantitative changes in the gene expression profile was not so conspicuous in this cluster (Fig. 8C). Therefore, most of endoderm cells that express SpBra also expressed foxA simultaneously.

2.4. SpBra expression in the oral ectoderm clusters is associated with the ventral organizer

According to previous studies including Li et al. (2014) and Erkenbrach (2016), the oral ectoderm of the sea urchin embryo is derived and/or composed of four regions: near apical ectoderm, stomodeal, central, and Veg1 oral ectoderm (Fig. 9A). Consistent with this interpretation, the scRNA-seq analysis has assigned clusters 1, 4, 12 and 13 to the oral ectoderm (Fig. 1A and B) (Foster et al., 2020). We found high levels of SpBra expression in cells of cluster 12; SpBra transcripts became detectable at later blastula stages, and higher and highest levels at early and late gastrula stages, respectively (Figs. 3 and 4). In addition, lower levels of SpBra expression were found in cells of cluster 1 at gastrula stages (Figs. 3 and 4). On the other hand, clusters 4 and 13 contained only a few cells with SpBra expression (Figs. 3 and 4). This suggests that clusters 12 and 1, especially the former, are involved in stomodeum formation and clusters 4 and 13 in oral surface ectoderm formation. Therefore, we first examined whether genes involved in the ventral organizer activity are expressed in clusters 12 and 1 (stomodeum) or 4 and 13 (surface ectoderm) or both.

Previous studies have shown that key regulatory genes responsible for the ventral organizer of sea urchin embryos are chordin, bmp2/4, nodal and ADMP1 (Duboc et al., 2004, 2008; Yaguchi et al., 2008; Lapraz et al., 2009, 2015; Wei et al., 2012; Cavalieri and Spinelli, 2014). In addition, not and Lim1 act synergistically with these genes (Lapraz et al., 2015). Therefore, we examined whether the expression of these six genes is found in cells of SpBra-positive clusters 12 and 1 or SpBra-negative 4 and 13 or in cells being randomly distributed over the four clusters. We found that (1) cells of cluster 12 exhibited high expression levels of chordin, bmp2/4, nodal, not, Lim1, and goosecoid at blastula and gastrula stages (Fig. 4; Fig. S5), although the number of

nodal-positive cells was comparatively small, (2) the expression of the ventral-organized related genes commences earlier (blastula stage) than SpBra (gastrula stage), (3) cells of cluster 1 also expressed moderate-level expression of chordin, bmp2/4, not and Lim1, although nodal expression was detected only in hatched blastulae (Fig. 4; Fig. S5), and, in contrast, (4) cells of clusters 4 and 13 showed undetectable level of expression of the five ventral organizer genes (Fig. 4). Therefore, we concluded that the ventral organizer activity is a property of cells of clusters 12 and 1.

Antagonistic interactions between the ventral organizer ADMP1 and the dorsal signaling center ADMP2 has been proposed to play a key role in ventral-dorsal axis formation in sea urchin embryos (Lapraz et al., 2015). The expression of ADMP1 was undetectable in blastulae and early gastrulae, and a few cells with ADMP1 expression appeared at late gastrula stage in clusters 12 and 1 (Figs. 3 and 4; Supplementary Fig. S6). On the other hand, ADMP2 expression was detected in a few cells of clusters 0, 3, and 5 at early and late gastrula stages (Fig. 3; Fig. S6). The clusters 0 and 5 have been assigned to aboral ectoderm (Fig. 1B) (Foster et al., 2020). Therefore, although the number of cells with ADMPs expression was comparatively small, a combinatorial expression between oral ectoderm bmp2/4, chordin and ADMP1 and aboral ectoderm ADMP2 was supported by this analysis.

2.5. SpBra is co-expressed with genes important for an ectodermal fate

As in the case of SpBra-positive endoderm clusters, we used three methods to examine co-expression at a single-cell level of SpBra and other genes in the ectoderm

clusters. First, we constructed superimposed t-SNE plots in combinations of SpBra and BMP2/4 (Supplementary Fig. S7A; Fig. 6D), SpBra and chordin (Fig. S7B; Fig. 6C), SpBra and nodal (Fig. S7C), SpBra and not (Fig. S7D; Fig. 6E), and SpBra and Lim1 (Fig. S7E). Cells of clusters 12 and 1 are seen as dots located at left lower side of the panels (Fig. 6C–E). Superimposed t-SNE plots showed clear co-expression of SpBra and chordin (Fig. 6C), and SpBra and not (Fig. 6E) at a single cell level (dots appeared as yellow). Co-expression was found in combinations of SpBra and bmp2/4 (Fig. 6D), and SpBra and Lim1 (Fig. S7E). Because of the scarcity of cells with nodal expression (Fig. S7C, middle), the presence of cells with simultaneous expression of SpBra and nodal was not evident (Fig. S7C).

Second, a heatmap of gene expression in the clustered matrix also showed co-expression of SpBra and other genes at a single cell level (Supplementary Fig. S8; Fig. 7B). As in the case of the endoderm clusters, cells of several regions in clusters 12 and 1 showed coincidental expression of SpBra, chordin, BMP2/4 and not (Fig. 7B, boxes). Third, a direct count of cells with high levels of SpBra expression also supported the results (Fig. 10; Supplementary Table S3). In clusters 12 and 1, between 60 and 80% of constituent cells highly expressed chordin and Not (Fig. 10). The number of cells with high level of SpBra expression in cluster 12 increased from 10% at hatched blastula stage to 40% at late gastrula stage (Fig. 10A). Most of SpBra-positive cells expressed simultaneously chordin and/or not (Fig. 10A), although the number of cells with simultaneous SpBra and BMP2/4 expression was small (less than 5%) (Fig. 10A). Cells of cluster 1 also exhibited a similar expression profile (Fig. 10B).

These results indicate that most of the constituent cells of clusters 12 and 1 are involved in the ventral organizer, a portion of which expressed SpBra and invaginated to form the stomodeum. In other words, the ventral organizer property overlaps morphogenesis of the oral invagination at a single-cell level, i.e. the two are not independent.

2.6. GRN of oral ectoderm clusters with and without SpBra expression

GRN studies have demonstrated pathways of transcription factors and signaling molecules in the specification of the oral ectoderm (e.g., Li et al., 2014) (Fig. 9B). In addition to BMP2/4, chordin, nodal, not, and ADMP1 described above, we examined expression profiles of the developmentally relevant genes for the oral ectoderm, including *bmp1*, *gooseoid* (*gsc*), *emx*, *hnf6* (*onecut*), *foxG*, *otx*, *hes*, *dri*, *FGFR*, *DVR1*, *nk2.1*, *foxq2*, *univin* and *six3/6* in oral ectoderm clusters (Figs. 3 and 4; Figs. S9 and S10). Of those, *otx*, *hes* and *dri* showed high levels of expression in cells of oral ectoderm clusters 1, 4, 12, and 13 at gastrula stage, though the expression of these genes was found in other clusters as well (Fig. 3). *foxG* expression was detected in all the four clusters at gastrula stage, and therefore does not provide a cue to distinguish clusters 12 and 1 from clusters 4 and 13 (Figs. 3 and 9B). *nk2.1* and *foxq2* were highly expressed in cells of clusters 9 and 17 in blastulae and gastrulae (Fig. 3; Fig. S10), and both clusters are assigned to apical and neural lineages (Fig. 1B).

gsc was highly expressed in clusters 12 and 1 but not in clusters 4 and 13 (Figs. 4 and 9B), suggesting an involvement of *gsc* in specification of the two former clusters

(stomodeum-related ectoderm), although *gsc* was reportedly expressed in the apical ectoderm and the central ectoderm (Li et al., 2014). On the other hand, expression of *emx* and *hnf6* (onecut) was higher in cells of clusters 4 and 13 than in clusters 12 and 1 (Fig. 4). Li et al. (2014) showed the involvement of *emx* and *hnf6* (onecut) in animal ectoderm specification (not for Veg1-Ecto), and our data suggests that in clusters 4 and 13, maybe one is animal and the other vegetal, and are not distinguished by the expression profiles of *emx* and *hnf6* (onecut) (Fig. 9B). *Lim1* is proposed to be a marker of central oral ectoderm (including stomodeum) and Veg1-derived oral ectoderm (Li et al., 2014). This study showed high levels of *Lim* expression in cluster 12 and 1 and moderate levels in cluster 13 but undetectable levels in cluster 4 (Fig. 4). Altogether, if the lineage scheme is forced into an scRNA-seq scheme, it is likely that the stomodeal ectoderm corresponds to cluster 12, near apical to cluster 4, central to cluster 1, and oral veg1 ectoderm to cluster 13 (Fig. 9).

Of further note is higher expression of *FGFR* in cells of cluster 12 in gastrulae (Fig. 4), although moderate levels of *FGFR* expression was found in other clusters as well (Fig. 3). This *FGFR* corresponds to *FGFR1* reported by Roettinger et al. (2008). On the other hand, *FGF* is reportedly expressed rather specifically, first in cells of cluster 13 at hatched blastula stage, and 4, 13, 16 and 19 at mesenchyme blastula stage, and 1, 4, 13, 16 and 19 at early gastrula stage (McCoon et al., 1996; Roettinger et al., 2008) (Fig. S9). Clusters 13 and 4 are *SpBra*-negative oral ectoderm (Fig. 1B) (Foster et al., 2020), suggesting a possible scenario in which *FGF* signaling is transmitted from *SpBra*-negative cells of clusters 4 and 13 and received by *SpBra*-expressing cells of

clusters 12 and 1 clusters, although such details should be examined experimentally in future studies.

2.7. Genes proposed as possible downstream targets of SpBra

Rast et al. (2002) have succeeded in the identification and characterization of possible target genes of SpBra, by subtraction of mRNAs of normally SpBra expressed embryos minus SpBra morpholino-knockdown embryos. Eleven genes are listed in Supplemental Table S4, representatives of which include kakapo and gelsolin for blastopore formation, APOBEC (a cytidine deaminase) and OrCt for gut formation, and PI103, PKS (polyketide synthase), DopT and CAPK for pigment cell formation (Rast et al., 2002). If SpBra is involved in positive regulation of these genes, they might be expressed in SpBra-positive cells. Therefore, we examined here the specificity and expression level of these genes by the sc-RNAseq method. High levels, but rather broad gene expression was found for Hes, CAPK, kakapo, gelsolin, and ephx2 (Table S4; Supplementary Fig. S11). The expression level was very low and without specificity in hyalin-like, OrCT, EH1, and Nk1 (Table S4; Fig. S11). On the other hand, specific and high level of expression was detected in APOBEC in endoderm clusters 6 and 14 (Table S4; Fig. S11). Therefore, APOBEC provides an experimental system to explore molecular mechanisms of how SpBra activates this gene in endoderm.

3. Discussion

The present scRNA-seq deep-analysis unambiguously showed that SpBra is expressed in endoderm cells (clusters 14, 6, and 8) to form the archenteron, and in ectoderm cells (clusters 12 and 1) to form the stomodeum. Cells of cluster 12 and 1 exclusively express genes responsible for the ventral organizing activity, nearly a half of which coincidentally express SpBra. This dataset thereby is well situated to identify and to prioritize exactly which cells express SpBra, and what genes SpBra regulates.

3.1. Three endoderm lineages

The endoderm of a sea urchin embryo is divided into Veg2-Endo derived from Veg2 tier and Veg1-Endo from Veg1 tier of blastulae (Fig. 5A) (Ransick and Davidson, 1998). We propose here that clusters 14 and 6 correspond to Veg2-Endo and that cluster 8 corresponds to Veg1-Endo. Veg1-Endo and Veg2-Endo differ from each other by gene expression profiles, in which Veg2-Endo exclusively exhibits expression of *blimp1/krox*, *foxA*, and *gata-e* (Fig. 5B) (Cui et al., 2014). *hnf1* is reportedly expressed only in Veg1-Endo (Fig. 5B) (Cui et al., 2014). However, the present results showed that this gene is also expressed in Veg2-Endo as well (Fig. 2), being cautious to use *hnf1* as a marker of

Veg1-Endo. Veg2-Endo comprises of two clusters 14 and 6, and the two clusters are distinguished by specific expression of SLBP in cluster 14 but not in cluster 6 (Figs. 1C and 2). Based on these results, we proposed here cluster 14 as Veg2-Endo1 and

cluster 6 as Veg2-Endo2 (Fig. 5A). However, this scenario should be confirmed by future studies.

It should be mentioned here that the three clusters of endoderm cells commence the expression of SpBra, hox11/13b, eve, blimp1/krox, and foxA at blastula stages, before these progenitor cells form a gut at the gastrula stage. In other words, cell fate specification of the endoderm cells occurs earlier than actual morphogenesis of gastrulation, and cells precisely follow their fate to form the endoderm-derived larval structures (Davidson, 2006). This scRNA-seq study provides additional data to support this notion on the early and lineage-dependent specification mode of embryogenesis in sea urchins.

3.2. Relationship between three endoderm clusters and three gut regions

The larval digestive system is tripartite, composed of a foregut or muscular esophagus, a midgut or large spherical stomach, and a hindgut or short tubular intestine (reviewed by Annunziata et al., 2014; Annunziata et al., 2019). Therefore, a question raised here is a compositional relationship between the three embryonic cell clusters and the tripartite structure of the larval gut. Various developmentally relevant genes are dynamically expressed in each of the three regions of the larval digestive system (see Fig. 3 of Annunziata et al., 2014). These genes include Bra, foxA, Blimp1, gata-e, Endo16, Limp1, Xlos, and Cdx (Olivier et al., 2008; Cui et al., 2014; Annunziata et al., 2014). First, there is a difference in temporal SpBra expression profile among the three clusters. Although SpBra expression commences almost simultaneously in the three

clusters at hatched blastula, the gene expression was downregulated in cluster 14 and 6 by early gastrula stage, while SpBra expression was retained in cluster 8 by late gastrula stage. This transient expression profile of SpBra coincides with the result shown by Gross and McClay (2001), in which Bra protein is expressed in a circumferential region of presumptive endoderm cells neighboring the vegetal plate and the expression becomes undetectable when those same endoderm cells move inside the gastrula to form the archenteron. The last component of the tripartite gut to invaginate is the hindgut and SpBra expression remained there (Annunziata et al., 2014). The hindgut does not express gata-e, which coincides to our result that cluster 8 exhibits less expression of gata-e (Fig. 2). Lower levels of expression of foxA and Endo16 in cluster 8 also suggests that Veg1-Endo-derived cluster 8 gives rise to the hindgut.

Midgut morphogenesis is concluded by identifying higher levels of expression of blimp1/krox, foxA, gata-e and Endo16 in clusters 14 and 6 coincident with higher levels of expression of these genes in the midgut by in situ hybridization (Annunziata et al., 2014, Fig. 5B). We posit that cluster 14 and 6 give rise to anterior and posterior parts of midgut, respectively, based on an assumption in which cluster 14 is close to the Veg2-meso cluster 11 (Fig. 1A). Because cluster 11 of the SMC-related lineage exhibits high levels of gata-e in blastulae and Endo16 in blastulae and early gastrulae (Fig. 1C), it is likely that a portion of cluster 11 and cluster “SMC (gut subcluster)” as well (Fig. 1B) (Foster et al., 2020) is involved in the formation of foregut. These are forerunners in archenteron invagination. However, cells of these mesoderm clusters do not express SpBra. It is likely that sea urchin gastrulation is accomplished not only by physical

changes of first-invaginating cells but also by physical changes caused by SpBra in second-invaginating cells (Veg2-Endo and Veg1-Endo), that might push Veg2-meso inside the embryo.

3.3. Ectoderm cluster with SpBra expression and oral organizer activity

According to Li et al. (2015), the oral ectoderm of the sea urchin embryo is composed of four regions, stomodeal, near-apical, central, and veg1 oral ectoderm (Fig. 9A). This scRNA-seq study showed that the oral ectoderm is represented by four clusters, cluster 12 with cells of high level of SpBra expression, cluster 1 with cells of moderate level of SpBra expression, and cluster 4 and 13 with cells lacking SpBra expression (Figs. 3 and 4). Genes involved in the ventral organizer activity are expressed in cells of cluster 12 and 1 (Fig. 4). At least three results should be in mind to interpret the relationship of the ventral organizer and morphogenesis of oral invagination to form stomodeum. First, developmentally relevant genes that are involved in the ventral organizer activity such as chordin, BMP2/4 and not commence their expression at hatched blastula stage earlier than SpBra expression at gastrula stage (Figs. 3 and 4) (e.g., Lapraz et al., 2015). Second, in contrast, nearly 80% of the constituent cells in clusters 12 express chordin and not (Figs. 4 and 10), and only 25% and 40% of the cells express SpBra at early and late gastrula, respectively (Fig. 10). Third, most SpBra-positive cells are simultaneously expressing chordin and not (Fig. 10), indicating co-expression of SpBra, chordin and not in the same individual cells. The portion of cells with the three gene expression did not change when embryogenesis proceeds from

early to late gastrula stage. Therefore, it is likely that (1) the ventral organizer activity is confined to cells of clusters 12 and 1, (2) the organizer activity might commence by late blastula stage, and (3) a portion of the organizer cells begin to express SpBra at gastrula stage and invaginate inside the embryo to form stomodeum. That is, the stomodeum formation is accomplished by cells with the ventral organizer activity.

In this study, we assigned cluster 1 to be the central oral ectoderm (Fig. 9B). However, it is likely that a portion of cluster 1 forms near the surface stomodeum and the other portion forms the surface ectoderm, and these questions will be answered by future studies. Our scheme of relationship between lineage-based and scRNA-seq-based specification of embryonic cells also should be examined in future studies. Nevertheless, we would like to emphasize that the ventral organizer activity resonates well with morphogenesis during stomodeum formation.

3.4. SpBra; upstream cascades and downstream targets

SpBra is expressed in two different regions of early sea urchin embryos, namely in endoderm and oral ectoderm. Because the GRN for endoderm (or endomesoderm) specification differs from GRN for oral ectoderm specification, the genetic cascade leading to SpBra likely differs between the two lineages. The genes that possibly act upstream of SpBra expression in endoderm have been studied extensively (Fig. 5B) (Rast et al., 2002; Davidson et al., 2002; Oliveri et al., 2006; Li et al., 2013; Cui et al., 2014; EchinoWiki). The present scRNA-seq analysis focused on *hox11/13b*, *eve*, *blimp1/krox*, *foxA*, *gata-e*, and *hnf1* (Figs. 1C and 2). First, the timing of commencement

of *gata-e* and *hnf1* expression was later than that of *SpBra* (Fig. 2), suggesting that *gata-e* and *hnf* are not upstream components of *SpBra* expression (Fig. 5B). Second, *hox11/13b* and *eve* are expressed commonly in all three endoderm clusters, suggesting that these two transcription factors share upstream components of *SpBra* (Figs. 2 and 5B). On the other hand, the expression level of *blimp1/krox* and *foxA* in cells of cluster 8 (Veg1-Endo) was quite low compared with that in cells of clusters 14 and 6 (Veg2-Endo). This suggests that the function of *blimp1/krox* and *foxA* as upstream regulators of *SpBra* is limited to the cells of Veg2-Endo (Cui et al., 2014). However, so far, no studies have reported sequence-specific cis-regulatory modules upstream of sea urchin *Brachyury*, which is essential for future discussion of *SpBra* expression control in the endodermal cells.

In endoderm cells, Ben-Tabou et al. (2010) reported the presence of a *SpBra*-binding site on the *foxA* enhancer, which supports the upregulation of *foxA*. In the absence of *Bra*, *foxA* continued to be expressed at a low level, indicating that other transcription factors operated prior to *SpBra*, and as it turns out, after *SpBra* is extinguished from those cells. Superimposed t-SNE plot analysis indicated co-expression of *SpBra* and *foxA* in cells of clusters 12 and 1 (Fig. 6D), suggesting a possibility that *foxA* acts upstream of *SpBra* in cells of oral stomodeal region. On the other hand, *foxG* expression was evident in cells of clusters 12 and 1 at late blastula and gastrula stages (Fig. 4), suggesting that it could be a possible upstream regulator of *SpBra* (Fig. 7B). In addition, *gsc* is highly likely to act upstream of *SpBra* (Fig. 7), since *gsc* is specifically expressed in cells of clusters 12 and 1 (Figs. 3 and 4). *chordin*, *bmp2/4*, *nodal*, *not* and *Lim* are highly expressed in cells of clusters 12 and 1 but not in

cells of cluster 4 and 13. How do the genes responsible for the ventral organizer interact with SpBra? Whether the expression of chordin, bmp2/4, nodal, not and Lim is essential for SpBra transcription or independent of SpBra transcription, is one of key questions to be answered to understand the molecular mechanisms involved in the specification of early sea urchin embryonic cells.

4. Conclusions

A deep-analysis of scRNA-seq datasets identified and characterized clusters of cells expressing SpBra in sea urchin embryos. SpBra expression was transient in blastulae, within cells of three clusters of endoderm and in gastrulae in cells of two of the four clusters of oral ectoderm. Pre-sumptive endodermal cells derived from Veg2 (clusters 14 and 6) simultaneously expressed SpBra, hox 11/13b, eve, blimp1/krox, foxA, gata-e, hnf1, and Endo16, whereas those derived from Veg1 (cluster 8) simultaneously expressed SpBra, hox11/13b, eve, and hnf1, but the expression of blimp1/krox, foxA, gata-e and Endo16 was present in low levels. Cells of cluster 14 and 6 are likely involved in mid- and/or foregut formation and cluster 8 in hindgut formation. Two (12 and 1) of the four clusters of oral ectodermal cells express SpBra together with chordin, bmp2/4, nodal, not, Lim1, and gsc. At a single-cell level, half of the cells with the ventral-organizer gene expression co-express SpBra, indicating that the two clusters serve as the ventral organizer and invaginate inside the embryo to form the stomodeum. In other words, the oral organizer activity was endowed in cells that can achieve morphogenetic movements to form the stomodeum. Two other clusters of oral ectoderm

without SpBra expression give rise to the oral surface epidermis. The present scRNA-seq analysis therefore highlights a significant role of Brachyury in early morphogenesis of sea urchin embryos at single cell exposure and opens a plethora of new candidates to test for SpBra expression and function.

5. Materials and methods

All scRNA-seq data used in this study have been deposited in NCBI database under accession no. GSE149221. Clustering of cells have already been reported by Foster et al. (2020), but the dataset was reanalyzed here. In this study, CellRanger gene expression matrices were analyzed using the R (v4.1.2; <http://www.R-project.org/>) package Seurat v 4.0.5 (Satija et al., 2015; Hao et al., 2021). First the number of constituent cells in each cluster at five different stages was counted (Table S1).

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 such as Violin plots and Feature plots were conducted as reported in Foster et al. (2020) (Fig. 6; Supplementary Figs. S1, S2, S4, S5, S6, S7, S9, S10 and S11).

Average expression was measured as count data of mRNAs normalized to library size and log transformed and used to dot-plot presentation (Figs. 1C, 2, 3, and 4) (Hao et al., 2021). To construct heatmap, log-normalized counts (>2 or higher) were pulled from the dataset (Supplementary Tables S2, S3 and S4) (Foster et al., 2020). Counted cell numbers were converted to bar graph (Figs. 8 and 10). The heatmap and

clustered matrix (Fig. 7; Supplementary Figs. S4 and S8) were created using R with the package Bioconductor (v3.14) and pheatmap (v1.0.12). Clustering for Fig. 7 was done with Ward's method (Ward, 1963), and Figs. S4 and S8 were sorted by log-normalized count of SpBra.

Author contributions

NS and GW conceived the experiment. KH, KN, HT, SM, NO and SF performed data analyses. NS and GW wrote the manuscript which was approved by all authors.

Declaration of competing interest

The authors declare no competing or financial interests.

Acknowledgments

OIST internal fund to Marine Genomics Unit (NS) supports this analysis. Members of Brown University are grateful for support from the National Institutes of Health (1R35GM140897 and 1P20GM119943).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ydbio.2022.01.005>.

References

- Andrikou, C., Iovene, E., Rizzo, F., Oliveri, P., Arnone, M.I., 2013. Myogenesis in the sea urchin embryo: the molecular fingerprint of the myoblast precursors. *EvoDevo* 4, 33, 2013.
- Annunziata, R., Perillo, M., Andrikou, C., Cole, A.G., Martinez, P., Arnone, M.I., 2014. Pattern and process during sea urchin gut morphogenesis: the regulatory landscape. *Genesis* 52, 251–268.
- Annunziata, R., Andrikou, C., Perillo, M., Cuomo, C., Cole, Arnone, M.I., 2019. Development and evolution of gut structures: from molecules to function. *Cell Tissue Res.* 377, 445–458.
- Ben-Tabou, S., Davidosn, E.H., 2010. Information processing at foxa node of the sea urchin endomesoderm specification network. *Proc. Natl. Acad. Sci. Unit. States Am.* 107, 10103–10108.
- Calestani, C., Rogers, D.J., 2010. Cis-regulatory analysis of the sea urchin pigment cell gene polyketide synthase. *Dev. Biol.* 340, 249–255.
- Cao, C., Lemaire, I.A., Wang, W., Yoon, P.H., Choi, Y.A., Parsons, L.R., Matese, J.C., Wang, W., Levine, M., Chen, K., 2019. Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature* 571, 349–354.
- Cavaliere, V., Spinelli, G., 2014. Early asymmetric cues triggering the dorsal/ventral gene regulatory network of the sea urchin embryo. *Elife* 3, e04664.
- Collura, R., Katula, K.S., 1992. Spatial pattern of expression of Cyl actin-p-galactosidase fusion genes injected into sea urchin eggs. *Dev. Growth Differ.* 34, 635–647.
- Croce, J., Lhomond, G., Gache, C., 2001. Expression pattern of Brachyury in the embryo of the sea urchin *Paracentrotus lividus*. *Dev. Gene. Evol.* 211, 617–619.
- Cui, M., Siriwon, N., Li, E., et al., 2014. Specific functions of the Wnt signaling system in gene regulatory networks throughout the early sea urchin embryo. *Proc. Natl. Acad. Sci. U. S. A.* 111, E5029–E5038.
- Davidson, E.H., 2006. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. Academic Press, San Diego, USA.
- Davidson, E.H., Rast, J.R., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Schilstra, M.J., Clarke, P.J.C., Rust, A.G., Pan, Z., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo develop. *Biol.* 246, 162–190.
- Duboc, V., Rottinger, E., Besnardeau, L., Lepage, T., 2004. Nodal and BMP2/4 signaling organizes the oral-aboral axis of the sea urchin embryo. *Dev. Cell* 6, 397–410.
- Duboc, V., Lapraz, F., Besnardeau, L., Lepage, T., 2008. Lefty acts as an essential modulator of nodal activity during sea urchin oral-aboral axis formation. *Dev. Biol.* 320, 49–59.
- EchinoWiki, 2020. https://wiki.echinobase.org/echinowiki/index.php/Main_Page.

- Erkenbrach, E.M., 2016. Divergence of ectodermal and mesodermal gene regulatory network lineages in early development of sea urchins. *Proc. Natl. Acad. Sci. U.S.A.* 113, E7202–E7211.
- Ettensohn, C.A., 2020. The gene regulatory control of gastrulation. *Mech. Dev.* 162, 103599.
- Fang, H., Brandhorst, B.P., 1994. Evolution of actin gene family of sea urchins. *J. Mol. Evol.* 39, 347–356.
- Foster, S., Oulhen, N., Wessel, G., 2020. A single cell RNA sequencing resource for early sea urchin development. *Development* 147, dev191528. <https://doi.org/10.1242/dev.191528>.
- Gross, J.M., McClay, D.R., 2001. The role of Brachyury (T) during gastrulation movements in the sea urchin *Lytechinus variegatus*. *Dev. Biol.* 239, 132–147.
- Hao, Y., Hao, S., Andersen-Nissen, E., et al., 2021. Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587.
- Harada, Y., Yasuo, H., Satoh, N., 1995. A sea urchin homologue of the chordate Brachyury (T) gene is expressed in the secondary mesenchyme cell found cells. *Development* 121, 2747–2754.
- Howard-Ashby, M., Materna, S.C., Brown, C.T., et al., 2006. Gene families encoding transcription factors expressed in early development of *Strongylocentrotus purpuratus*. *Dev. Biol.* 300, 90–107.
- Lapraz, F., Besnardeau, L., Lepage, T., 2009. Dorsal-ventral patterning in echinoderms: insights into the evolution of the BMP-Chordin signaling Network. *PLoS Biol.* 7, 1–25.
- Lapraz, F., Haillot, E., Lepage, T., 2015. A deuterostome origin of the Spemann organiser suggested by Nodal and ADMPs functions in Echinoderms. *Nat. Commun.* 6, 8434. <https://doi.org/10.1038/ncomms9434>.
- Li, E., Materna, S.C., Davidson, E.H., 2013. New regulatory circuit controlling spatial and temporal gene expression in the sea urchin embryo oral ectoderm GRN. *Dev. Biol.* 382, 268–279.
- Li, E., Cui, M., Peter, I.S., et al., 2014. Encoding regulatory state boundaries in the pregastrular oral ectoderm of the sea urchin embryo. *Proc. Natl. Acad. Sci. U. S. A.* 111, E906–E913.
- Materna, S.C., Ransick, A., Li, E., Davidson, E.H., 2013. Diversification of oral and aboral mesodermal regulatory states in pregastrular sea urchin embryos. *Dev. Biol.* 375, 92–104.
- McClay, D.R., 2011. Evolutionary crossroads in developmental biology: sea urchins. *Development* 138, 2639–2648.
- McCoon, P.E., Angerer, R.C., Angerer, L.M., 1996. SpFGFR, a new member of the fibroblast growth factor receptor family, is developmentally regulated during early sea urchin development. *J. Biol. Chem.* 271, 20119–20125.
- Nocente-McGrath, C., Brenner, C.A., Ernst, S.G., 1989. Endo16, a lineage-specific protein of the sea urchin embryo, is first expressed just prior to gastrulation. *Dev. Biol.* 136, 264–272.
- Oliveri, P., Walton, K.D., Davidson, E.H., McClay, D.R., 2006. Repression of Mesodermal Fate by Foxa, a Key Endoderm Regulator of the Sea Urchin Embryo Development, vol. 133, pp. 4173–4181.

- Perillo, M., Oulhen, N., Foster, S., Spurrell, M., Calestani, G., Wessel, G., 2020. Regulation of dynamic pigment cell states at single-cell resolution. *Elife* 9, e60388. <https://doi.org/10.7554/eLife.60388>.
- Peter, I.S., Davidson, E.H., 2010. The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. *Dev. Biol.* 340, 188–199.
- Peterson, K.J., Harada, Y., Cameron, R.A., Davidson, E.H., 1999. Expression pattern of Brachyury and Not in the sea urchin: comparative implications for the origins of mesoderm in the basal deuterostomes. *Dev. Biol.* 207, 419–431.
- Ransick, A., Ernst, S., Britten, R.J., 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.
- Ransick, A., Davidson, E.H., 1998. Late specification of Veg1 lineages to endodermal fate in the sea urchin embryo. *Dev. Biol.* 195, 38–48. 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.
- Rast, J.P., Cameron, R.A., Poustka, A.J., Davidson, E.H., 2002. Brachyury target genes in the early sea urchin embryo isolated by differential microarray screening. *Dev. Biol.* 246, 191–208.
- Robertson, A.J., Howard, J.T., Dominski, Z., Schnackenberg, B.J., Sumerel, J.L., McCarthy, J.J., Coffman, J.A., Marzluff, W.F., 2004. The sea urchin stem-loop-binding protein: a maternally expressed protein that probably functions in expression of multiple classes of histone mRNA. *Nucleic Acids Res.* 32, 811–818.
- Roëttinger, E., Saudemont, A., Duboc, V., Besnardeau, L., McClay, D., Lepage, T., 2008. FGF signals guide migration of mesenchymal cells, control skeletal morphogenesis and regulate gastrulation during sea urchin development. *Development* 135, 353–365.
- Satija, R., Farrell, J., Gennert, D., et al., 2015. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502.
- Sethi, A.J., Wikmanayakem, R.M., Angerer, R.C., Range, R.C., Angerer, L.M., 2012. Sequential signaling crosstalk regulates endomesoderm segregation in sea urchin embryos. *Science* 335, 590–593.
- Soltysik-Espanola, M., Klinzing, D.C., Pfarr, K., Burke, R.D., Ernst, S.G., 1994. Endo16, a large multidomain protein found on the surface and ECM of endodermal cells during sea urchin gastrulation, binds calcium. *Dev. Biol.* 165, 73–85.
- Stuart, T., Satija, R., 2019. Integrative single-cell analysis. *Nat. Rev. Genet.* 20, 257–272.
- Su, Y.H., Li, E., Geiss, G.K., et al., 2009. A perturbation model of the gene regulatory network for oral and aboral ectoderm specification in the sea urchin embryo. *Dev. Biol.* 329, 410–421.
- Ward Jr., J.H., 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58, 236–244.
- Wei, Z., Range, R., Angerer, R., Angerer, L., 2012. Axial patterning interactions in the sea urchin embryo: suppression of nodal by Wnt1 signaling. *Development* 139, 1662–1669.

- Yaguchi, S., Yaguchi, J., Angerer, R.C., Angerer, L.M.A., 2008. Wnt-FoxQ2-nodal pathway links primary and secondary axis specification in sea urchin embryos. *Dev. Cell* 14, 97–107.
- Yuh, C.-H., Bolouri, H., Davidson, E.H., 2001. Cis-regulatory logic in the *endo16* gene: switching from a specification to a differentiation mode of control. *Development* 128, 617–629.

Figures

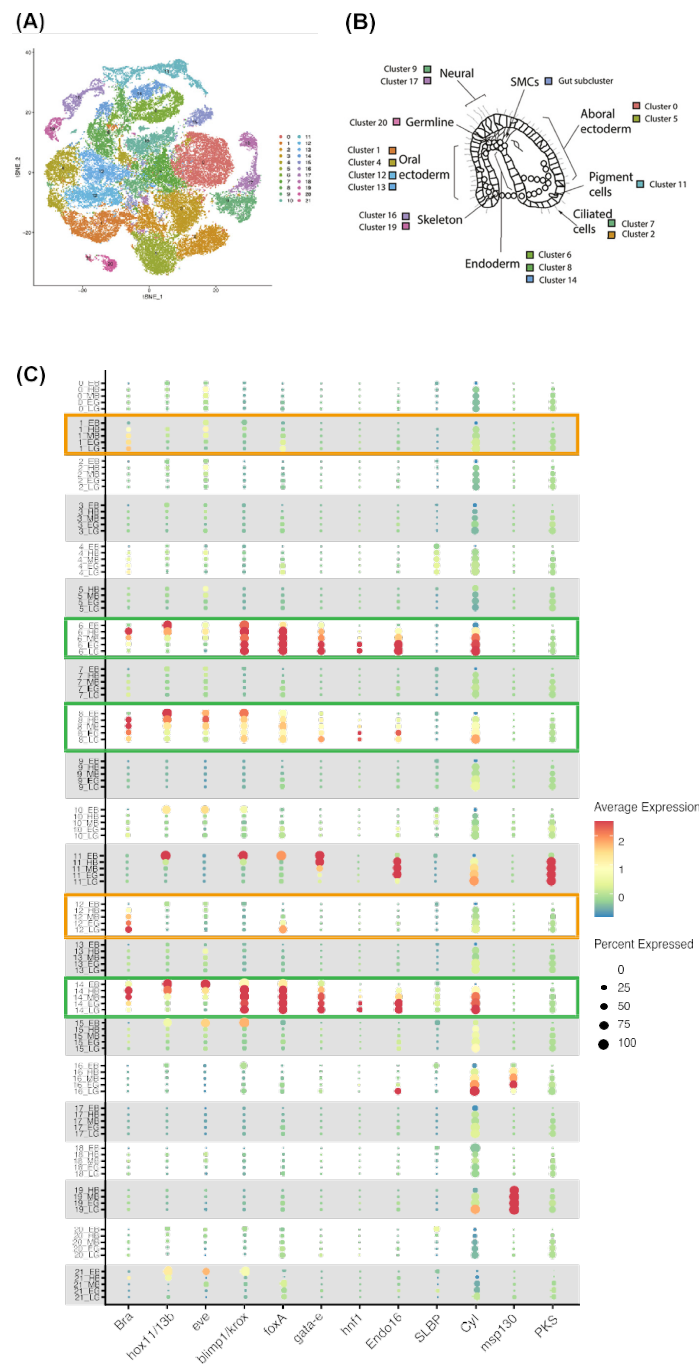


Fig. 1. Identification of cell clusters with expression of Brachyury (Bra) in early embryos of *Strongylocentrotus purpuratus*. (A) t-SNE plot of integrated dataset with clustering, showing 22 cell-clusters numbered from 0 to 21. t-SNE plot of integrated dataset with clustering was performed at a resolution of 0.5 (from Foster et al., 2020). (B) Drawing of late gastrula and mapping cell clusters identified by scRNA-seq to the embryo. Colors match cell clusters seen in A (from Foster et al., 2020). Clusters 6, 8

and 14 are endoderm and clusters 1, 4, 12 and 13 oral ectoderm. (C) Transcriptome trajectories for 22 clusters (from 0 to 21) comprising blastula and gastrula stages of *Strongylocentrotus purpuratus* embryos. Major genes that are preferentially expressed in endoderm lineages include *Bra*, *hox11/13b*, *eve*, *blimp1/k-rox*, *foxA*, *gata-e*, *hnf1*, *Endo16*, *SLBP*, *Cyl*, *msp130*, and *pks1*. Dot size represents the percentage of cells that express genes for transcription factors, signaling molecules, and structural proteins. Dot color shows the averaged level of expression. EB, early blastula; HB, hatched blastula; MB, mesenchyme blastula; EG, early gastrula; and LG, late gastrula. Clusters enclosed by green belong to the endoderm with *SpBra* expression, numbers 6, 8 and 14, and those enclosed by brown belong to the oral ectoderm with *SpBra* expression, number 1 and 12.

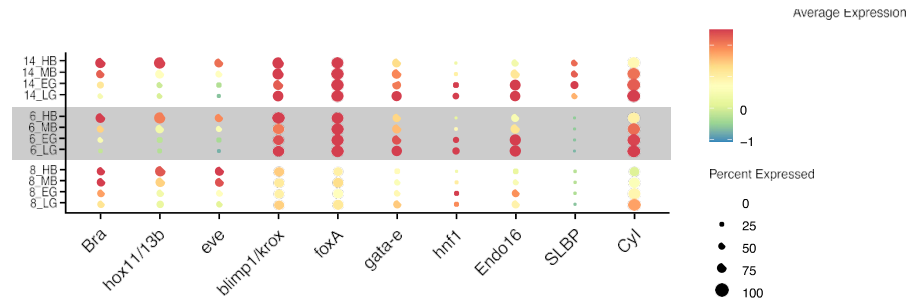


Fig. 2. Enlargement of clusters 14, 6 and 8 to show transcriptome trajectories of Bra, hox11/13b, eve, blimp1/krox, foxA, gata-e, hnf1, Endo16, SLBP, and Cyl. Dot size represents the percentage of cells that express genes for transcription factors, signaling molecules, and structural proteins. Dot color shows the averaged level of expression. EB, early blastula; HB, hatched blastula; MB, mesenchyme blastula; EG, early gastrula; and LG, late gastrula

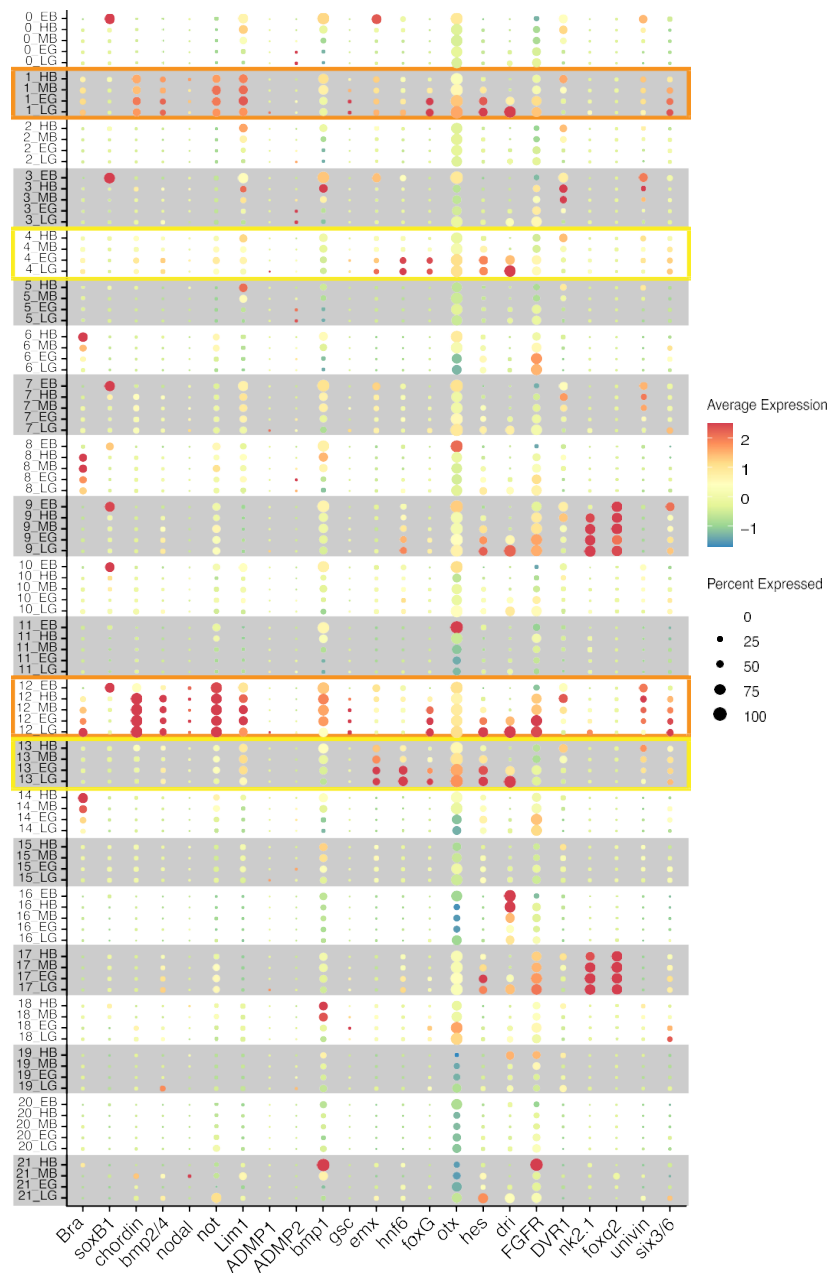


Fig. 3. Transcriptome trajectories for 22 clusters comprising blastula and gastrula stages of *Strongylocentrotus purpuratus* embryos. Major genes that are preferentially expressed in oral ectoderm lineages include Bra, SoxB1, chordin, bmp2/4, nodal, not, Lim1, ADMP1, ADMP2, bmp1 gsc, emx, hnf6, foxG, otx, hes, dri, FGFR, DVR1, nk2.1, foxq2, univin, and six3/6. Dot size represents the percentage of cells that express genes for transcription factors, signaling molecules, and structural proteins. Dot color shows the averaged level of expression. EB, early blastula; HB, hatched blastula; MB, mesenchyme blastula; EG, early gastrula; and LG, late gastrula. Clusters enclosed are ectoderm-lineage, 1 and 12 with SpBra expression (brown) and 4 and 13 with no SpBra expression (yellow).

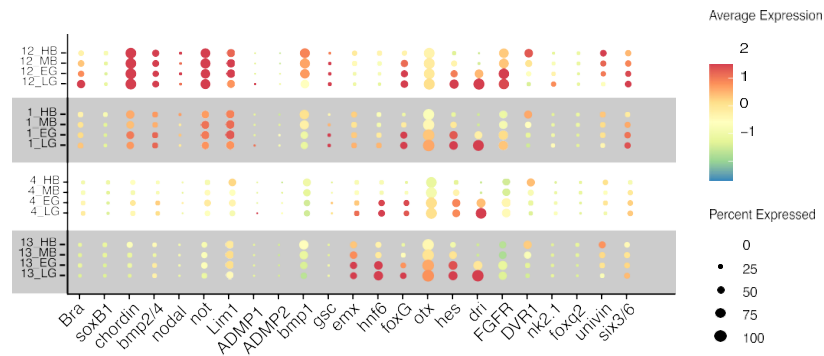


Fig. 4. Enlargement of clusters 12, 1, 4, and 13 to show transcriptome trajectories of Bra, chordin, bmp2/4, nodal, not, Lim1, ADMP1, ADMP2, bmp1 gsc, emx, hnf6, foxG, otx, hes, dri, FGFR, DVR1, nk2.1, foxq2, univin, and six3/6. Dot size represents the percentage of cells that express genes for transcription factors, signaling molecules, and structural proteins. Dot color shows the averaged level of expression. EB, early blastula; HB, hatched blastula; MB, mesenchyme blastula; EG, early gastrula; and LG, late gastrula

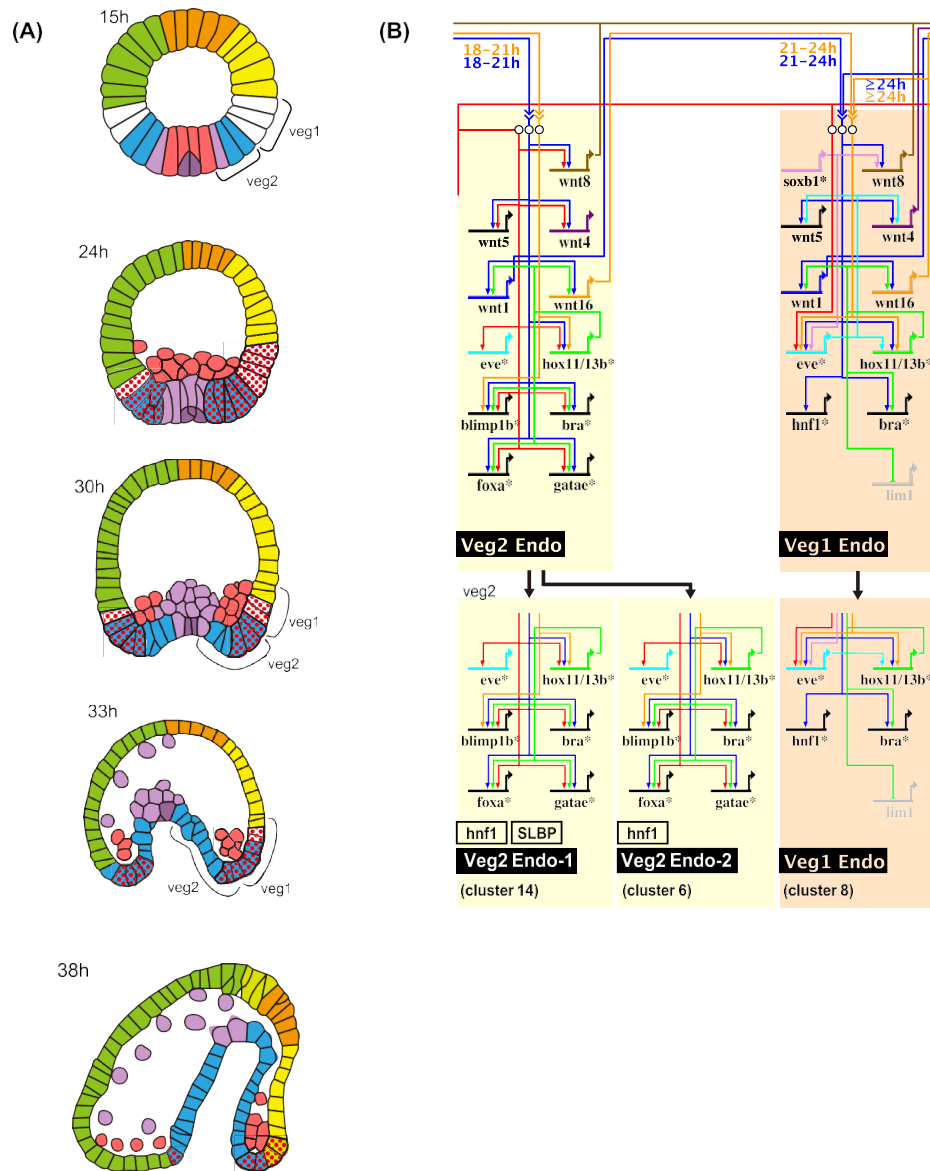
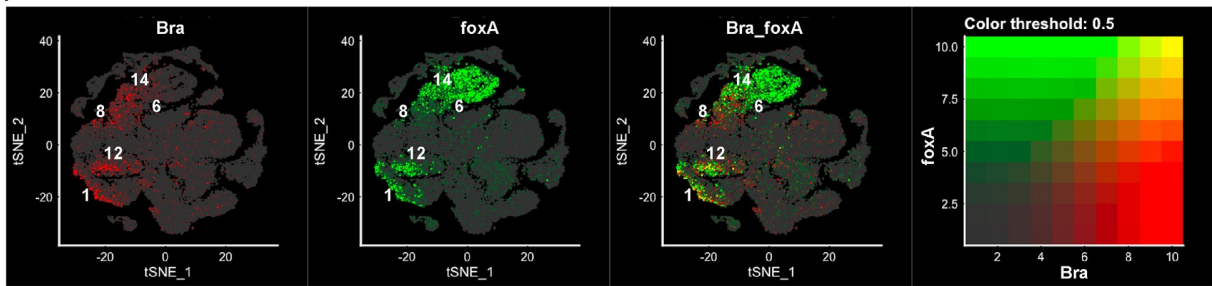


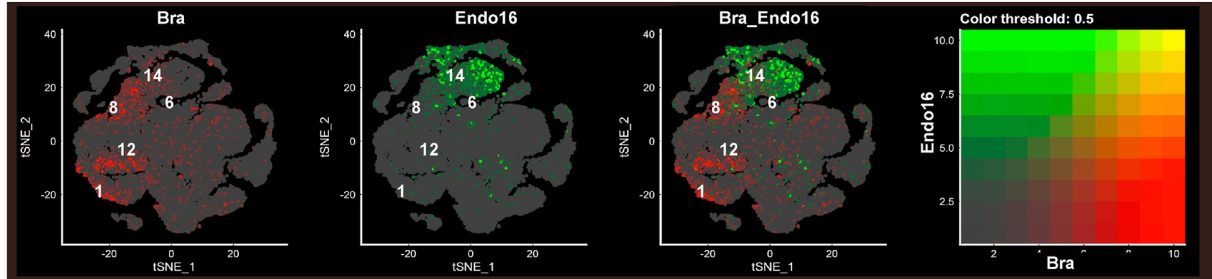
Fig. 5. (A) Diagrams to show specification in the sea urchin embryo. Color-coded tracings from photomicrographs of the embryo of *Strongylocentrotus purpuratus* are shown. veg1 and veg2 are rings of eight cells each, arising from their parental cells at the horizontal 6th cleavage. From veg1 derives ectoderm (mainly) plus hindgut endoderm; and from veg2 non-skeletogenic (secondary) mesenchyme (mesodermal cell types) plus gut endoderm. Skeletogenic mesenchyme lineage, red; endoderm, blue; secondary mesenchyme, violet; oral ectoderm, yellow; apical oral ectoderm, hatched yellow; aboral ectoderm, green; unspecified cells, white. 15 h, blastula stage; 20 and 24 h, mesenchyme blastula; 30, 33, 38 h, gastrula stages. The original drawing of embryos by Howard-Ashby et al. (2006) is cited (A. Ransick and E. Davidson; reproduced from Davidson, 2006, copyright Elsevier Inc.). SpBra expression was shown by red dots. (B) Gene regulatory network (GRN) in endoderm specification of sea urchin embryos. The original drawing of embryos by Cui et al. (2014) is cited. Endoderm comprises of Veg2-derived (Veg2-Endo) and Veg1-derived (Veg1-Endo). The present scRNA-seq analysis

suggests two classes of Veg2-derived cells, Veg2-Endo1 (cluster 14) and Veg2-Endo2 (cluster 6), which is distinguishable by exclusive expression SLBP in the former. Genes enclosed are findings of this study.

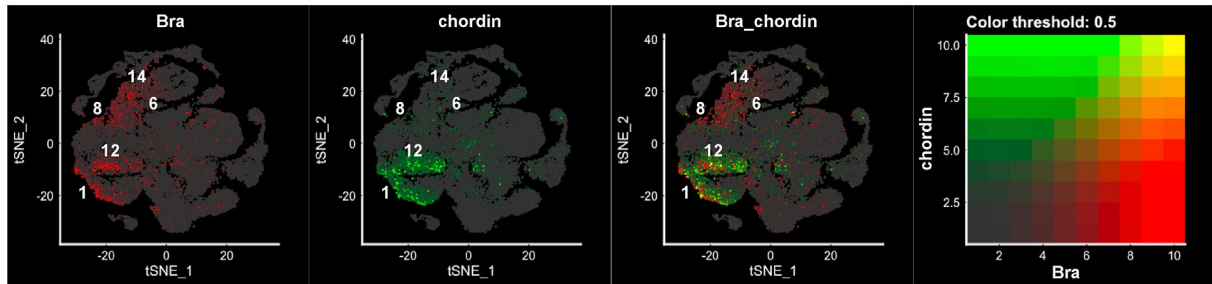
(A)



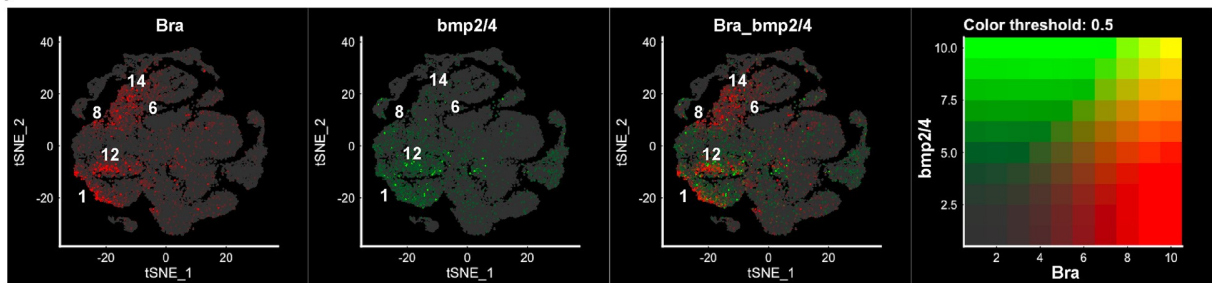
(B)



(C)



(D)



(E)

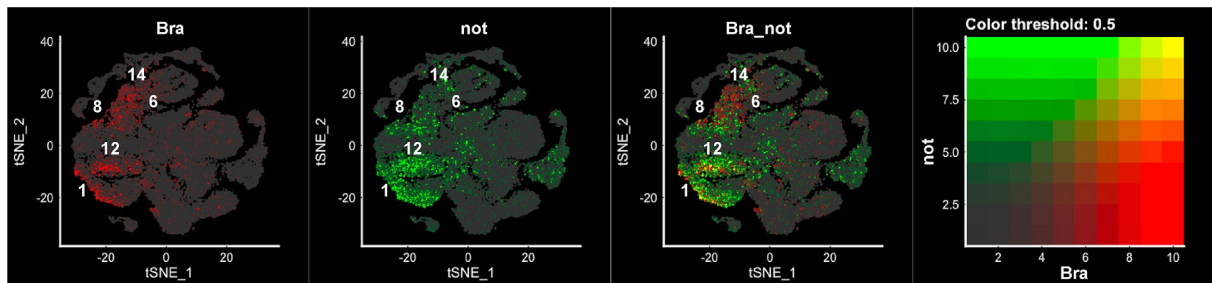


Fig. 6. t-SNE plots showing co-expression of Brachyury with (A) *foxA*, (B) *Endo16*, (C) *chordin*, (D) *bmp2/4*, and (E) *not*. Cells with *SpBra* expression are shown by red dots, the other genes by green dots, and co-expression of the two genes by yellow dots. Clusters 14, 6 and 8 are of endoderm and clusters 12 and 1 of oral ectoderm.

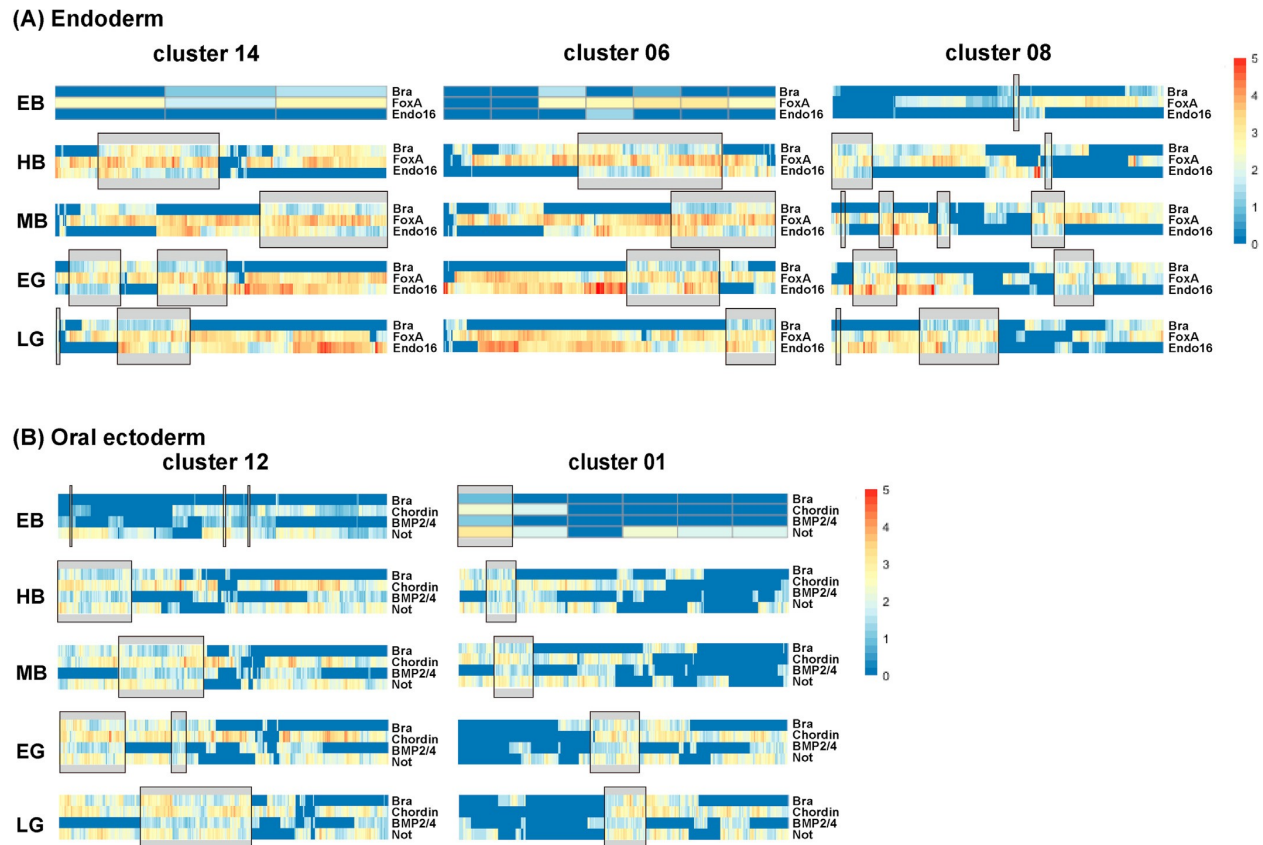
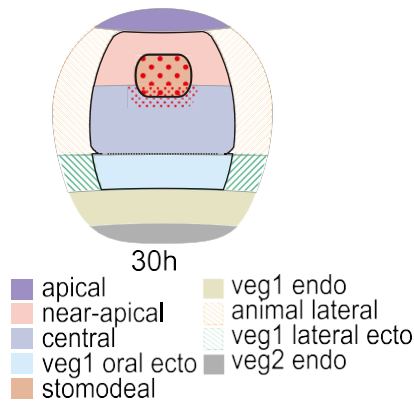


Fig. 7. Heatmap to show the co-expression level of Bra and other genes. (A) The expression of Bra, foxA and Endo16 in clusters 14, 6 and 8. Clusters of cells that simultaneously expressed Bra, foxA and Endo16 are boxed. (B) The expression of Bra, chordin, BMP2/4 and Not in clusters 12 and 1. Clusters of cells that simultaneously expressed Bra, chordin, BMP2/4 and Not are boxed. EB, early blastula; HB, hatched blastula; MB, mesenchyme blastula; EG, early gastrula; and LG, late gastrula.



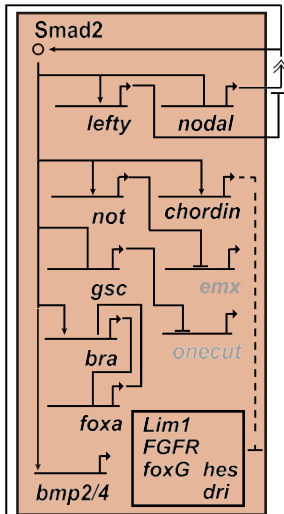
Fig. 8. Cells highly expressing Bra, FoxA or Endo16, and those co-expressing Bra and FoxA, Bra and Endo, or Bra, FoxA and Endo16 are shown as percentages of the cells among constituent cell numbers of clusters (A) 14, (B) 6 and (C) 8, respectively.

(A)

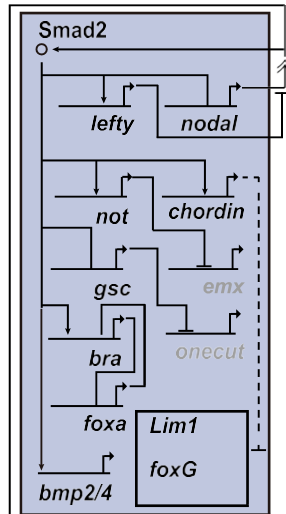


(B)

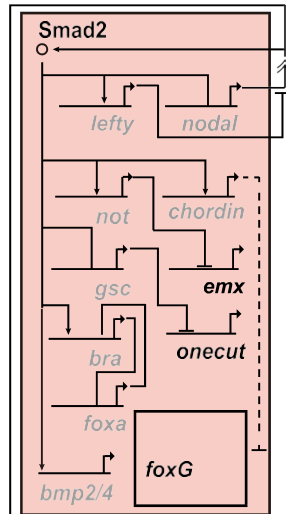
stomodeum
(cluster 12)



central
(sub-stomodeum)
(cluster 1)



near-apical
(cluster 4)



oral veg1 ecto
(cluster 13)

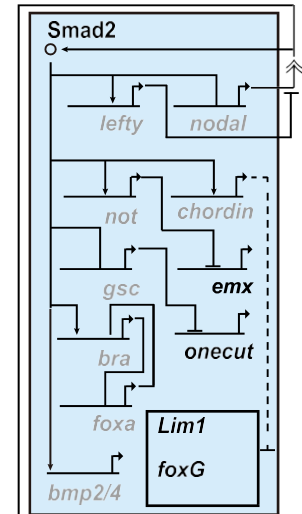


Fig. 9. A proposed GRN in four clusters of oral ectoderm. (A) Diagram showing territories of oral ectoderm (circled by solid line). Dotted line indicates the boundary between animal and vegetal hemisphere. The original drawing of embryos by Li et al. (2014) is cited. SpBra expression is shown by red dots. (B) Possible GNRs in cluster 12 (stomodeum), cluster 1 (central; sub-stomodeum), near-apical (cluster 4) and oral veg1 ectoderm (cluster 13). Genes with bold letter are expressed in the cluster while those with gray are not expressed in the cluster. Enclosed are findings of this study.

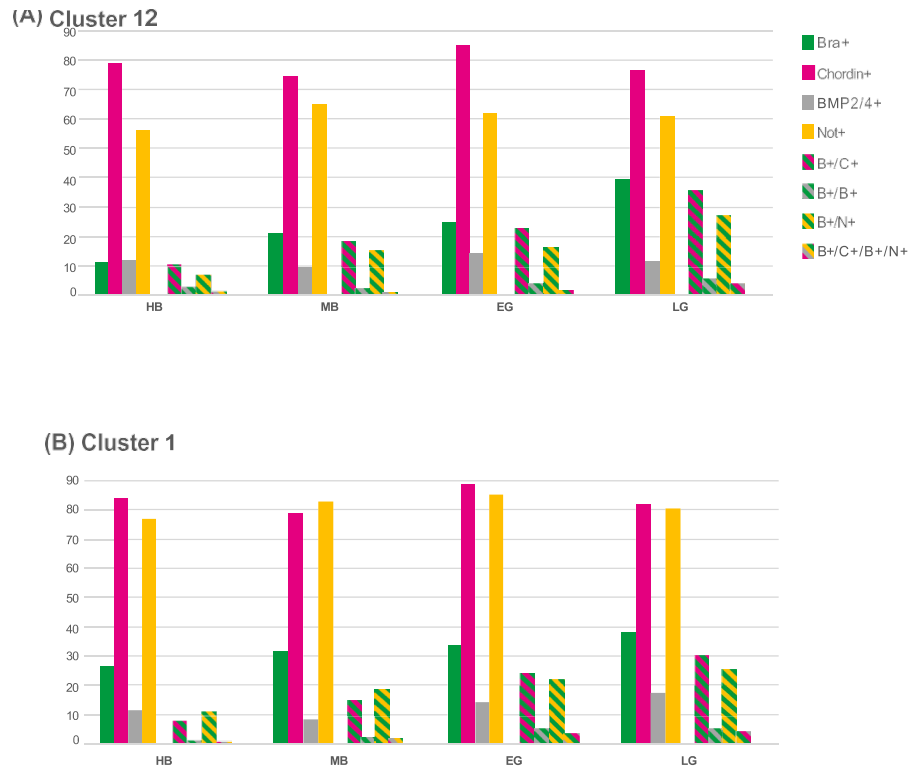


Fig. 10. Cells highly expressing Bra, chordin, BMP2/4 or Not, and those co-expressing Bra and chordin, Bra and BMP2/4, Bra and Not, or Bra, chordin, BMP2/4 and Not are shown as percentages of the cells among constituent cell numbers of clusters (A) 12 and (B) 1, respectively.