

Evolutionary transitions within Echinodermata

Adrian M. Reich

B.A., Cornell University, 2005

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of:

Doctor of Philosophy

In the Department of Molecular Biology, Cell Biology, and Biochemistry at:

Brown University

Providence, Rhode Island

May, 2014

© 2014 by Adrian M. Reich

This dissertation by Adrian M. Reich 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 _____

Gary M. Wessel, Ph.D., Advisor

Recommended to the Graduate Council

Date _____

Casey W. Dunn, Ph.D., Reader

Date _____

Mark Johnson, Ph.D., Reader

Date _____

Michael McKeown, Ph.D., Reader

Date _____

Kimberly Mowry, Ph.D., Reader

Date _____

Andrew Cameron, Ph.D., Reader, California Institute of Technology

Approved by the Graduate Council

Date _____

Peter M. Weber, Dean of the Graduate School

Curriculum Vitae

Adrian M. Reich

EDUCATION

- Ph.D. Candidate**, Brown University, Providence, RI, Expected graduation, 2014
Department of Molecular Biology, Cellular Biology and Biochemistry, Advisor: Dr. Gary Wessel, Doctoral Dissertation: “*Evolutionary transitions in Echinodermata*”
- B.A.**, Cornell University, Ithaca, NY, 2004
College of Arts and Sciences, Major: Biological Sciences; Concentration: Neurobiology and Behavior

TECHNICAL SKILLS

Laboratory Techniques: Illumina library sample prep, *de novo* transcriptome and genome assembly, SNP analysis, Proteomics, Alternative splicing, Genome resequencing, Optical and confocal microscopy, Transmission Electron Microscope, PCR, Western blot, Cloning, Microinjection, Electrophoresis, Tandem Affinity Purification, RNA synthesis, Typhoon, Protein purification

Computing Techniques: High-performance cluster computing, Linux, Comparative genomics, Pathway and gene ontology analysis

Software: Python, R, Velvet, Oases, Trinity, Bowtie, TopHat, samtools, CASAVA, Microsoft Office, File Maker Pro, Access, Photoshop, Illustrator, Dreamweaver

Languages: English (native speaker), French (fluent)

RESEARCH EXPERIENCE

- PhD Candidate**, Brown University, Providence, RI, 2007-2014
Conducting primary research and analysis examining the transcriptomes and genomes using high-throughput sequencing technologies.
My research spans the *de novo* assembly of transcriptomes and genomes of non-model organisms, and analysis of the transcriptomes of single cells in humans and mice. I have sequenced and analyzed over 25 *de novo* transcriptome projects, three genome projects, over 20 single cell samples; comprising greater than 500 gigabases of sequence data on four different next generation platforms.
- Research Assistant**, Brown University, Providence, RI, 2005-2007
Investigation of translational control and RNA localization in Xenopus laevis oocytes.
I conducted independent molecular biology research and also supported the projects of graduate students and post-docs. I worked on of the purification of a protein and mRNA complex undergoing active transport in the early *Xenopus* oocyte.
- Field Research Assistant**, Cornell University, Isla Socorro, Mexico, Spring 2005 & 2003
Collected data and managed database for a population survey of humpback whales.
- Co-Investigator**, Cornell University, Cranberry Lake Biological Field Station, NY, Summer 2004
Investigated recruitment rates of foragers in a honeybee hive as a function of varying comb substrate.

PUBLICATIONS

- Juliano, C.E., **Reich, A.**, Liu, N., Uman, S., Wessel, G.M., Steele, R.E., and Lin., H. 2013. Analysis of the PIWI-piRNA pathway in Hydra reveals broad function in somatic adult stem cells. *PNAS* (2014), 111(1), 337-342.
- Fresques, T., Zazueta-Novoa, V., **Reich, A.**, and Wessel, G.M. 2013. 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* (*Epub ahead of print*).
- Wessel, G.M., Brayboy, L., Fresques, T., Gustafson, E.A., Oulhen, N., Ramos, I., **Reich A.**, Swartz, S.Z., Yajima, M., and Zazueta, V. 2013. The biology of the germ line in echinoderms. *Mol Reprod Dev.* (*Epub ahead of print*).
- Oulhen, N., **Reich, A.**, Wong, J.L., Ramos, I., and Wessel, G.M. 2013. Diversity in the fertilization envelopes of echinoderms. *Evol Dev.* 15(1): 28-40.
- Reich, A.**, Neretti, N., Freiman, R.N., and Wessel, G.M. 2012. Transcriptome variance in single oocytes within, and between, genotypes. *Mol Reprod Dev.* 79(8): 502-3.
- Reich, A.**, Klatsky, P., Carson, S., and Wessel, G.M. 2011. The transcriptome of a human polar body accurately reflects its sibling oocyte. *JBC* 286: 40743-40749.
- Wessel, G.M., **Reich A.M.**, and Klatsky, P.C. 2010. Use of sea stars to study basic reproductive processes. *Syst Biol Reprod Med.* 56(3): 236-45.
- Seeley, T.D., **Reich, A.M.**, and Tautz, J. 2005. Does plastic comb foundation hinder waggle dance communication? *Apidologie* 36: 513-521.

PUBLISHED CODE

- Reich, A.** 2012. Centralize *de novo* transcriptome assembly into a database: <http://code.google.com/p/oases-to-csv/>

PRESENTATIONS

- CCV/EPSCoR Bioinformatics Workshop, Providence, RI, October 2013
"De novo assembly of transcriptomes and genomes"
- Developmental Biology of the Sea Urchin XXI, Woods Hole, MA, October 2012
"Comparative analysis of echinoderm ovary transcriptomes"
- CCV/EPSCoR Bioinformatics Workshop, Providence, RI, October 2012
"Identification of the sex chromosomes in S. purpuratus"
- 12th International Congress of Human Genetics, Montreal, Canada, October 2011
"The transcriptome of a human polar body accurately reflects its sibling oocyte"
- Developmental Biology of the Sea Urchin XXI, Woods Hole, MA, April 2011
"Comparative analysis of echinoderm ovary transcriptomes"
- Genome Assembly Special Forces Workshop, Providence, RI, May 2011
"De novo genome assembly 101"

HONORS AND AWARDS

- Invited speaker, Developmental Biology of the Sea Urchin XXI, 2012
- Trainee Research Award, American Society of Human Genetics, 2011
- Invited speaker, 12th International Congress of Human Genetics, 2011
- Invited speaker, Developmental Biology of the Sea Urchin XX, 2011
- Invited speaker, Genome Assembly Special Forces Workshop, 2011

TEACHING EXPERIENCE

- Graduate Student Advisor, iGEM, Brown University, 2008-2010

Principle advisor for four teams of undergraduates in the synthetic biology iGEM competition.

Teaching Assistant, Biology 1210, Brown University, 2009

Led discussions and gave lectures in the class Synthetic Biological Systems.

Graduate Student Advisor, Biology 1950, Brown University, 2007

Designed and taught a course to introduce undergraduate students to synthetic biology.

Led primary literature discussions and taught molecular biology lab techniques.

TRAINING AND CERTIFICATIONS

High Performance Computing – Brown University Center for Computation and Visualization, 2012

Hazardous Waste – Brown University, 2005-present

Laboratory Safety – Brown University, 2005-present

Radiation Safety – Brown University, 2005-2007

Acknowledgments

This thesis is dedicated to my wife Danielle and daughter Coralie.

This work would not have been possible without the aid and support of many people, foremost among them Danielle. Without her support these long years, it is unlikely you would be reading this, dear reader. I would also like to acknowledge the support from our family, specifically Linda, without whom Coralie would not be half as healthy or happy.

All past and present members of PRIMO have been amazing to work with and incredibly fulfilling. Special thanks to my graduate student brothers/sisters in arms: Celina, Eric, Zak, and Tara, with a specific thank you to Zak; together we experienced the whole journey together.

The scientific support at Brown University, especially by the staff of the Center for Computation and Visualization and the Brown Genomics Facility, has been nothing short of exemplary. I would particularly like to thank Mark Howison, Lingsheng Dong, Christoph Schorl, Hilary Hartlaub, and James Clifton.

The academic support I received at Brown University was second to none both by the administration and faculty. Thank you to Elaine Butler, Rosemarie Antoni, Christie Crozier Brown, and Tammy Glass. Among the numerous faculty that have supported and aided me, I would like to thank my committee members: Mark Johnson, Gary Wessel, Casey Dunn, Kimberly Mowry, Michael McKeown, and Andrew Cameron. Thank you especially to Kim, who first gave me a job in her lab and to Casey and members of his lab who all helped me tremendously with much of my work, especially Felipe Zapata. I would also like to thank the incredible academic training I received from Tricia Serio, Jeffrey Laney, Kimberly Mowry, and Alison DeLong.

I would especially like to thank my advisor, Gary Wessel for being an incredibly gifted scientist and a remarkable human being. I would not have been able to thrive in any other lab and I certainly would not have been able to accomplish a comparable body of work. Thank you.

This work would not have been possible without the financial support from very diverse sources including: National Institutes of Health, National Science Foundation, Brown University provost's office, Center for Reproduction and Infertility at Women & Infants Hospital of Rhode Island, Sigma-Aldrich and Novartis.

To everyone too numerous to list but no less critical, my most sincere gratitude.

Finally I would like to thank you dear reader, because without a source for this accumulated knowledge, all would have been in vain.

Table of Contents

| | |
|---|-----|
| Title page | i |
| Copyright page..... | ii |
| Signature page..... | iii |
| Curriculum Vitae | iv |
| Acknowledgments..... | vii |
| Table of Contents..... | ix |
| List of Figures..... | xvi |
| List of Tables | xix |
| Chapter I: Introduction..... | 1 |
| Abstract..... | 2 |
| Introduction..... | 3 |
| Evolution of Echinoderms | 3 |
| Evolutionary differences in fertilization and early cleavage of echinoderms | 3 |
| Evolutionary history of larval development of echinoderms | 5 |
| Open questions..... | 7 |
| References..... | 9 |
| Chapter II: Phylogenetic analysis of extant echinoderms using <i>de novo</i> transcriptomes | 11 |
| Contribution | 12 |
| Abstract..... | 13 |
| Introduction..... | 14 |
| Results and Discussion | 15 |
| Transcriptome assemblies..... | 15 |
| Phylogenetic relationship of extant echinoderms | 16 |
| Resolution of the order Paxillosida in Asterozoa..... | 17 |
| Morphological and character trait changes | 18 |
| Conclusions..... | 19 |
| Future Directions | 19 |
| Positive selection in Asterozoa..... | 19 |
| Materials and Methods..... | 20 |

| | |
|--|----|
| RNA isolation and sequencing..... | 20 |
| Transcriptome assemblies and RefSeq data..... | 20 |
| Post assembly and phylogenetic analyses..... | 21 |
| Data Availability..... | 22 |
| References..... | 23 |
| Figures and tables | 25 |
| Supplemental Information | 28 |
| Chapter III: 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 | 36 |
| Contribution..... | 37 |
| Abstract..... | 38 |
| Introduction..... | 39 |
| Results and Discussion | 41 |
| Conserved germ-line determinants – Select expression in the Posterior Enterocoel | 42 |
| Gene regulatory molecules involved in inductive specification of germ cells in the mouse are conserved in echinoderms..... | 45 |
| Germ-line associated genes..... | 50 |
| Left/Right asymmetry molecules..... | 54 |
| Genomic maintenance during morphogenesis and early embryogenesis..... | 56 |
| Conclusions..... | 57 |
| Materials and Methods..... | 60 |
| Animals and embryo culture..... | 60 |
| RNA analysis | 60 |
| References..... | 61 |
| Figures and Tables | 67 |
| Chapter IV: Diversity in the fertilization envelopes of echinoderms | 86 |
| Contribution..... | 87 |
| Abstract..... | 88 |
| Introduction..... | 89 |
| Results..... | 93 |
| Sea star and sea urchin fertilization envelopes show differential permeability | 93 |
| Only three of the five proteins found in the sea urchin fertilization envelope are present in the sea star..... | 94 |
| The proteins involved in fertilization envelope formation differs among Echinoderms | 95 |

| | |
|---|-----|
| <i>Rendezvin</i> , <i>SFE9</i> , and <i>proteoliasin</i> transcripts are specifically expressed during the early oogenesis..... | 95 |
| Cortical granules translocate to the cell periphery during early oogenesis..... | 96 |
| Discussion..... | 97 |
| Materials and Methods..... | 101 |
| Animals..... | 101 |
| Permeability assays..... | 101 |
| Mass spectrometry analysis..... | 102 |
| Phylogenetic analysis..... | 102 |
| Whole mount RNA <i>in situ</i> hybridization (WMISH)..... | 102 |
| Real-time quantitative PCR (QPCR)..... | 103 |
| Antibody production..... | 103 |
| Western blot..... | 103 |
| Immunofluorescence..... | 104 |
| References..... | 105 |
| Figures..... | 108 |
| Supplemental Information..... | 117 |
| Chapter V: Synthesis and future directions..... | 123 |
| Introduction..... | 124 |
| Results..... | 124 |
| Discussion..... | 126 |
| Germline determination in echinoderms..... | 126 |
| How has germline determination evolved in Echinodermata?..... | 128 |
| Evidence of ancestral inductive germline determination in Euechinoids?..... | 131 |
| Conclusions..... | 134 |
| References..... | 135 |
| Figures..... | 137 |
| Appendix I: The transcriptome of a human polar body accurately reflects its sibling oocyte..... | 138 |
| Contribution..... | 139 |
| Abstract..... | 140 |
| Introduction..... | 141 |
| Methods..... | 142 |
| Human Oocyte Collection and Polar Body Biopsy..... | 142 |

| | |
|---|-----|
| Biopsy and WTA Amplification | 143 |
| Illumina Library Preparation and Sequencing | 143 |
| Mapping and Statistical Analysis..... | 144 |
| Results and Discussion | 144 |
| Analysis of Detected Genes and Gene Expression Levels..... | 144 |
| Examination of Gene Expression Profiles | 146 |
| Clinical Feasibility Test and Microarray Comparison | 148 |
| Conclusions..... | 150 |
| Data Availability..... | 150 |
| References..... | 151 |
| Figures and Tables | 153 |
| Supplemental Information | 159 |
| Appendix II: Transcriptome variance in single oocytes within, and between, genotypes | 169 |
| Contribution..... | 170 |
| Abstract..... | 171 |
| Results and Discussion | 172 |
| Conclusions..... | 174 |
| Data Availability..... | 174 |
| References..... | 175 |
| Figures | 176 |
| Supplemental Information | 178 |
| Appendix III: PIWI proteins and PIWI-interacting RNAs function in Hydra somatic stem cells | 182 |
| Contribution..... | 183 |
| Abstract..... | 184 |
| Introduction..... | 185 |
| Results..... | 186 |
| <i>Hydra</i> PIWI proteins, Hywi and Hyli, are expressed in multipotent stem cells. | 186 |
| Hywi and Hyli accumulate in perinuclear granules of epithelial stem/progenitor cells | 186 |
| Isolation and characterization of <i>Hydra</i> piRNAs reveals conserved mechanisms of piRNA biogenesis..... | 187 |
| The <i>Hydra</i> PIWI-piRNA targets transposon transcripts..... | 188 |
| Identification of candidate non-transposon PIWI-piRNA pathway targets | 189 |
| Hywi has an essential function in <i>Hydra</i> epithelial cells | 190 |

| | |
|--|-----|
| Discussion..... | 192 |
| Materials and Methods..... | 194 |
| Animals and Culturing Conditions | 194 |
| Hywi and Hyli Antibody Generation | 194 |
| Nuclear-Cytoplasmic Fractionation | 194 |
| Fluorescence Activated Cell Sorting (FACS)..... | 195 |
| Immunoprecipitation and piRNA Sequencing | 195 |
| Sequencing of Lineage-Specific Small RNAs | 195 |
| Assembly of the Hydra Transcriptome and Small RNA Mapping | 195 |
| Generation of Transgenic <i>Hydra</i> | 195 |
| Data Availability..... | 196 |
| References..... | 197 |
| Figures | 200 |
| Supplemental Information | 206 |
| <i>Hydra</i> strains and culturing conditions | 206 |
| Hywi and Hyli identification and antibody generation..... | 206 |
| Immunoblot and immunofluorescence analysis..... | 207 |
| Immuno-electron microscopy | 208 |
| Nuclear-Cytoplasmic Fractionation | 209 |
| Fluorescence Activated Cell Sorting (FACS)..... | 209 |
| Immunoprecipitation and piRNA sequencing..... | 209 |
| β -elimination and small RNA northern blot | 210 |
| Sequencing of lineage-specific small RNAs..... | 212 |
| Bioinformatic analysis and genomic mapping of small RNAs | 212 |
| Assembly of the <i>Hydra</i> transcriptome and mapping of piRNAs and lineage-specific small RNAs | 213 |
| Real-time quantitative PCR to test <i>hywi</i> knockdown levels | 214 |
| RNAi plasmid description and construction | 215 |
| Generation of transgenic <i>Hydra</i> | 215 |
| Supplemental References..... | 217 |
| Supplemental Figures and Tables | 219 |
| Appendix IV: Deadenylase depletion protects inherited mRNAs in primordial germ cells | 245 |
| Contribution | 246 |

| | |
|--|-----|
| Abstract..... | 247 |
| Introduction..... | 248 |
| Results..... | 249 |
| Differential expression analysis and identification of sMic enriched transcripts | 249 |
| The sMics are broadly transcriptionally repressed..... | 250 |
| CNOT6 transcript is selectively degraded in the sMics by a Nanos/Pumilio dependent mechanism | 251 |
| CNOT6 repression is required for retention of germ line determinants | 253 |
| Discussion..... | 254 |
| Materials and Methods..... | 256 |
| Animals..... | 256 |
| FACS isolation of sMics..... | 257 |
| Helicos sample preparation and deep sequencing..... | 257 |
| Illumina sample preparation, deep sequencing, and reference transcriptome assembly..... | 258 |
| Differential expression analysis..... | 258 |
| Whole mount in situ hybridization (WMISH) and immunofluorescence..... | 259 |
| Cloning and Reporter constructions..... | 260 |
| Morpholino antisense oligo (MASO) and mRNA microinjection..... | 260 |
| Western blot..... | 260 |
| Immunoprecipitation..... | 261 |
| Data Availability..... | 261 |
| References..... | 262 |
| Figures | 265 |
| Supplemental Information | 272 |
| Appendix V: A computational approach for the identification of the sex chromosomes in <i>S. purpuratus</i> | 288 |
| Contribution..... | 289 |
| Abstract..... | 290 |
| Introduction..... | 291 |
| Results and Discussion | 292 |
| Testing XY sex determination | 292 |
| Testing ZW sex determination..... | 294 |
| Future Directions | 294 |
| Materials and Methods..... | 296 |

| | |
|---|-----|
| Read processing | 296 |
| Genome mapping | 296 |
| <i>S. purpuratus</i> <i>de novo</i> transcriptome sequencing and assembly | 296 |
| Long single molecule DNA visualization | 297 |
| References..... | 298 |
| Figures and Tables | 299 |
| Appendix VI: Assembly of the genome of an early branching echinoderm, <i>Oxycomanthus japonicus</i> | |
| Contribution | 304 |
| Abstract..... | 305 |
| Introduction..... | 306 |
| Results and Discussion | 308 |
| Future Directions | 308 |
| Materials and Methods..... | 308 |
| Paired-end sequencing | 308 |
| Mate pair sequencing | 309 |
| BAC end sequencing..... | 309 |
| Genome assembly | 309 |
| De novo transcriptome sequence and assembly..... | 309 |
| Data Availability..... | 310 |
| References..... | 311 |
| Figures and Tables | 313 |
| Supplemental Information | 314 |

List of Figures

| | |
|---|-----|
| Chapter II | 25 |
| Figure 1: Two competing hypotheses of the phylogenetic relationship of extant echinoderms. ... | 25 |
| Figure 2: Phylogenetic relationship of extant echinoderms..... | 26 |
| Figure 3: Morphological and embryological trait changes in echinoderms..... | 27 |
| Supplemental Figure 1: Phylogenetic relationship of extant echinoderms..... | 28 |
| Supplemental Figure 2: Postassembly comparisons of RefSeq and <i>de novo</i> assembled datasets.. | 29 |
| Supplemental Figure 3: Sparse and dense supermatricies including all thirty taxa. | 30 |
| Supplemental Figure 4: Test of convergence of PhyloBayes chains. | 31 |
| | |
| Chapter III..... | 67 |
| Figure 1. Schematic representation of the developmental stages in sea star and sea urchin..... | 67 |
| Figure 2. Expression of conserved germ-line determinants during <i>P. miniata</i> embryonic development..... | 68 |
| Figure 3. Expression of genes involved in inductive germ-line specification during <i>P. miniata</i> embryonic development..... | 70 |
| Figure 4. Expression of germ-line associated genes during <i>P. miniata</i> embryonic development. | 72 |
| Figure 5. Expression of left/right asymmetry markers during <i>P. miniata</i> embryonic development. | 74 |
| Figure 6. Molecules involved in genomic regulation and maintenance during <i>P. miniata</i> embryonic development..... | 75 |
| Figure 7. Transcript dynamics during posterior enterocoel formation. | 76 |
| | |
| Chapter IV..... | 108 |
| Figure 1. The fertilization envelope is more permeable in sea star than in sea urchin. | 108 |
| Figure 2. In the sea star <i>Pm</i> , the fertilization envelope is composed of three major proteins: SFE9, rendezvin, and proteoliasin. | 109 |
| Figure 3. Phylogenetic trees representing the proteins involved in the formation of the fertilization envelope. | 110 |
| Figure 4. <i>Pm</i> rendezvin, <i>SFE9</i> , and <i>proteoliasin</i> mRNAs are highly and uniformly expressed during early oogenesis. | 111 |
| Figure 5. <i>Pm-SFE9</i> , <i>proteoliasin</i> , and <i>rendezvin</i> RNA levels decrease during oogenesis. | 112 |
| Figure 6. <i>Pm-SFE9</i> antibody specifically recognizes one high molecular weight bands. | 113 |
| Figure 7. The protein <i>Pm-SFE9</i> is present throughout oogenesis, maturation, and fertilization. | 114 |
| Figure 8. Sea star cortical granules move to the periphery of the cell during early oogenesis. ... | 115 |
| Figure 9. The cortex of immature oocytes and ultrastructural immunolocalization of SFE9 in cortical granules..... | 116 |
| Supplemental Figure 1. <i>Pm-SFE9</i> antibody specificity..... | 117 |
| Supplemental Figure 2. After fertilization, <i>Pm-SFE9</i> is incorporated in the fertilization envelope. | 118 |
| Supplemental Figure 3. <i>Pm-SFE9</i> antibody specifically labels the early developmental stage... 119 | 119 |

| | |
|--|-----|
| Chapter V | 137 |
| Figure 1. Detailed trait changes in echinoderms. | 137 |
| Appendix I | 153 |
| Figure 1: Amplification and sequencing strategy. | 153 |
| Figure 2: Sample to sample overlap and comparison. | 154 |
| Figure 3: The most abundant genes in oocytes are compared with the most abundant genes in polar bodies. | 156 |
| Figure 4: Comparison of our data with previously published data. | 157 |
| Supplemental Figure 1: Oocyte gene expression extrapolation. | 159 |
| Supplemental Figure 2: Pair-wise comparison of all oocytes samples. | 160 |
| Supplemental Figure 3: Test for differential gene expression between oocytes and polar bodies. | 161 |
| Supplemental Figure 4: Select KEGG pathway maps. | 162 |
| Appendix II | 176 |
| Figure 1: KO and WT morphological and molecular comparisons. | 176 |
| Figure 2: Dendrogram of 5 wild-type samples and 16 Taf4b-knockout samples. | 177 |
| Supplemental Figure 1: Heatmap comparing significant genes between KO and WT. | 178 |
| Supplemental Figure 2: Knockout versus wildtype standard deviations. | 179 |
| Supplemental Figure 3: A single mouse oocyte superimposed on a heatmap of differentially expressed genes. | 180 |
| Appendix III | 200 |
| Figure 1: PIWI proteins are expressed in the interstitial stem cells and are enriched in perinuclear granules. | 200 |
| Figure 2: PIWI proteins are cytoplasmic and expressed in the mitotically active somatic epithelial cells. | 201 |
| Figure 3: Sequencing and mapping of Hywi- and Hyli-bound piRNAs reveals conserved mechanisms of piRNA biogenesis and candidate post-transcriptional targets. | 202 |
| Figure 4: <i>Hywi</i> has an essential function in <i>Hydra</i> epithelial cells. | 204 |
| Supplemental Figure 1: Generation and characterization of antibodies against Hywi and Hyli. | 219 |
| Supplemental Figure 2: Hyli protein is expressed in interstitial stem cells and mitotically active epithelial stem/progenitor cells. | 221 |
| Supplemental Figure 3: <i>Hydra</i> PIWI proteins are expressed in developing nematoblasts. | 222 |
| Supplemental Figure 4: FACS isolation of ectodermal and endodermal cells. | 223 |
| Supplemental Figure 5: Isolation, deep-sequencing, and mapping of Hywi and Hyli bound piRNAs to the <i>Hydra</i> genome and beta elimination assay. | 224 |
| Supplemental Figure 6: Analysis of lineage-specific small RNAs. | 225 |
| Supplemental Figure 7: Transmission of the <i>hywi</i> RNAi-1 transgene through the germline and knockdown of <i>hywi</i> in the epithelial cells of F1 hatchlings. | 227 |
| Appendix IV | 265 |
| Figure 1. FACS isolation and deep sequencing of sMics. | 265 |

| | |
|--|-----|
| Figure 2. Transcriptional repression in sMics..... | 266 |
| Figure 3. Localization of select differentially expressed transcripts..... | 267 |
| Figure 4. CNOT6 mRNA is depleted in the sMics by Nanos..... | 268 |
| Figure 5. CNOT6 depletion is required for sMic Vasa protein expression..... | 269 |
| Figure 6. CNOT6 mediates selective enrichment of Seawi transcript in the sMics..... | 270 |
| Figure 7. CNOT6 regulates general retention of exogenous RNA in the sMics..... | 271 |
| Supplemental Figure 1: FACS isolation of sMics and transcriptomic analysis..... | 272 |
| Supplemental Figure 2: WMISH for selected transcripts. | 274 |
| Supplemental Figure 3: Interaction between Nanos and Pumillio and CNOT6 knockdown..... | 275 |
| Supplemental Figure 4: Time capsule model for germ line development. | 276 |
| | |
| Appendix V..... | 299 |
| Figure 1: Modeling run of fold coverage difference between sex chromosomes and autosomes. | 299 |
| Figure 2: Number of mapped read fragments per Kb of known sequence of each scaffold. | 300 |
| Figure 3: High copy number variation is seen in an individual scaffold. | 301 |
| | |
| Appendix VI..... | 314 |
| Supplemental Figure 1: Insert sizes of genome mate pair libraries. | 314 |

List of Tables

| | |
|--|-----|
| Chapter II | 32 |
| Supplemental Table 1: Assembly and post assembly summary by taxa..... | 32 |
| Supplemental Table 2: Classification of sea stars..... | 33 |
| Supplemental Table 3: Morphological trait changes within echinoderms..... | 34 |
| Supplemental Table 4: High copy sequences trimmed from <i>O. wendtii</i> 454 assembly..... | 35 |
| | |
| Chapter III..... | 77 |
| Table 1. Conserved germ-line determinants..... | 77 |
| Table 2. Genes involved in the inductive mechanisms of germ-line specification..... | 79 |
| Table 3. Germ-line associated genes..... | 81 |
| Table 4. Left-Right asymmetry molecules..... | 83 |
| Table 5. Regulation and genomic maintenance during morphogenesis and early embryogenesis..... | 84 |
| | |
| Chapter IV..... | 120 |
| Supplemental Table 1. Primers used to analyze the expression of the transcripts <i>Pm-SFE9</i> , <i>rendezvin</i> and <i>proteoliaisin</i> | 120 |
| Supplemental Table 2. Transcripts encoding for proteins involved in the formation of the fertilization envelope in sea urchins, pencil urchin, and sea stars..... | 121 |
| Supplemental Table 3. <i>Pm-SFE9</i> , <i>proteoliaisin</i> , and <i>rendezvin</i> mRNA are highly expressed in young oocytes..... | 122 |
| | |
| Appendix I | 158 |
| Table 1: 22 oocytes and sibling PBs were divided into eight sequencing reactions..... | 158 |
| Supplemental Table 1: The 279 genes that are expressed in all four paired samples..... | 163 |
| | |
| Appendix II..... | 181 |
| Supplemental Table 1: Database of KO vs WT..... | 181 |
| | |
| Appendix III..... | 229 |
| Supplemental Table 1: piRNA mapping to the <i>Hydra</i> transcriptome..... | 229 |
| Supplemental Table 2: Gene ontology analysis of transcripts with greater than 10 Hywi-bound piRNAs mapped..... | 230 |
| Supplemental Table 3: Gene ontology analysis of transcripts with greater than 10 Hyli-bound piRNAs mapped..... | 234 |
| Supplemental Table 4: Gene ontology analysis of putative lineage-specific targets of the PIWI- piRNA pathway..... | 238 |
| Supplemental Table 5: Real-time quantitative PCR to test <i>hywi</i> knockdown levels..... | 244 |
| | |
| Appendix IV | 277 |
| Supplemental Table 1: Gene set enrichment analysis of the union sets of 78 sMic enriched transcripts, and 152 sMic-depleted transcripts..... | 277 |

| | |
|--|-----|
| Supplemental Table 2: Yields of cell sorting runs. | 278 |
| Supplemental Table 3: Run statistics for Helicos sequencing runs. | 279 |
| Supplemental Table 4: Transcripts that are differentially enriched in sMics compared to non-sMics (replicate 1 excluded). | 280 |
| Supplemental Table 5: Transcripts that are differentially depleted in sMics compared to non-sMics (replicate 1 excluded). | 282 |
| Supplemental Table 6: PCR primers for WMISH probes. | 285 |
| Supplemental Table 7: Primers used for generating constructs. | 286 |
| Supplemental Table 8: Custom morpholino sequences. | 287 |
| Appendix V. | 302 |
| Table 1: Mapping efficiency of read fragments to genome. | 302 |
| Appendix VI. | 313 |
| Table 1: Total read breakdown and coverage used to assemble genome. | 313 |

Chapter I: Introduction

Adrian Reich

ABSTRACT

One of the fundamental questions in biology that has generated significant scholarly debate is the succession of generations. Few organisms that are not asexual or that can divide parthenogenetically can claim immortality; therefore, organisms evolved the “immortal germline” to pass their fitness advantages on to progeny. In recent years, the discussion is framed in the context of separation of the germline from somatic tissue, whereby somatic tissue is lost in each successive generation, while the germline is competent to form an entirely new organism, including regeneration of the germline. Much of the research in germline determination has been concentrated in a handful of model organisms including *D. melanogaster*, *C. elegans* (members of Ecdysozoa) and the mouse (a deuterostome), among others. These well-studied model organisms exhibit a range of germline specification mechanisms, from an inherited germline in *D. melanogaster* to an induced germline in the mouse. However, these organisms are very distantly related with extremely different life histories and methods of reproduction where it can be difficult to study fertilization and early development (specifically internal fertilization and relatively few embryos).

Echinodermata is a group of organisms within Deuterostomia which includes: sea urchins, sea stars, sea cucumbers, brittle stars, and sea lilies; and have been invaluable in studying developmental biology for hundreds of years (Derbès, 1847; Ernst, 1997). Echinoderms have numerous advantages for the study of fertilization and early development (including external fertilization of millions of gametes), however the phylogenetic relationships of extant taxa is contentious. This phylum has yielded fascinating results on the evolution of germline determination and evolutionary transitions, but much remains to be discovered. With a well-supported phylogenetic tree of Echinodermata, studies of the evolution of germline determination and early development will become more informative and valuable.

INTRODUCTION

Evolution of Echinoderms

Echinoderms branched away from other deuterostomes more than 500 million years ago and rapidly diversified (Smith *et al.*, 2013), leaving a rich fossil record. Members of Echinodermata are a diverse group of organisms; in Deuterostomia they comprise the second most speciose group after chordates. Echinoderms are only found in marine habitats, but they occupy all benthic habitats, ranging from intertidal to deep sea. The adult body plan is unique among bilaterians as echinoderms demonstrate pentameric symmetry, while the larvae are bilaterally symmetrical. During larval stages, the adult rudiment develops inside the larva and upon metamorphosis, larval structures are typically lost. The adult body of extant echinoderms is supported by a mesoderm-derived biomineralized skeleton that is calcareous, though evidence for magnesium carbonate skeletons exist in the fossil record (Kouchinsky *et al.*, 2012). The biomineralized skeleton can take many forms, from a single structure or test consisting of fused plates in the case of sea urchins and sand dollars, to segmented arms with a full range of motion in the case of sea stars and especially brittle stars. All echinoderms exhibit robust regenerative abilities, both as larvae and adults, though brittle stars and crinoids are especially adept at regeneration, especially in the adult (Burns *et al.*, 2013; Candia Carnevali and Bonasoro, 2001; Gahn and Baumiller, 2010).

Evolutionary differences in fertilization and early cleavage of echinoderms

Sea urchins store the female gametes as fertilization competent eggs and are arrested in post-meiotic interphase (Pearse and Cameron, 1991). Sea urchin eggs range from 80-150 μ m in diameter, though some species have much larger eggs. In many species of echinoderms, including sea urchins, females can store millions of eggs that are released into the water column prior to external fertilization.

In sea stars by contrast, the female germ cells that are stored in the ovary are oocytes that are arrested in prophase of meiosis I (Chiba, 2000). Typically, sea star oocytes are larger than sea

urchins, 100-200µm in diameter, though in the case of lecithotrophic development, the oocytes can be as large as 1,100µm. The signal to re-initiate and complete meiosis and to mature the oocyte into a fertilization competent egg is 1-methyladenine (1-MA; Kanatani *et al.*, 1969). The maturation process initiated by 1-MA triggers many different pathways including germinal vesicle breakdown (reviewed in Chiba, 2000) and a MAP kinase apoptotic pathway that will induce cell death if the egg is not fertilized within 9-12 hours (Sasaki and Chiba, 2001). Due to the exquisite control over the sea star maturation process, oocytes can be freely manipulated prior to egg maturation and fertilization (Wessel *et al.*, 2010).

Upon fusion of the sperm and egg, a fertilization envelope is raised to help prevent multiple sperm fusing with the egg or polyspermy. The raising of the fertilization envelope takes approximately 30 seconds and is very efficient in the occlusion of sperm and particles much smaller in size (Wong and Wessel, 2006a). The fertilization envelope is constructed using the contents of the cortical granules which are secretory vesicles that are prepositioned in close proximity to the plasma membrane. When cortical granules are stimulated to fuse with the plasma membrane they release their contents into the environment whereupon the secreted factors react with the extracellular matrix (Wong and Wessel, 2006b). Upon exocytosis of the cortical granules, a series of biochemical pathways are initiated, including: proteolysis, transamidation, hydrogen peroxide synthesis, and dityrosine crosslinking (Wong and Wessel, 2008). Furthermore, the rapid raising of the fertilization envelope is due to the hemifused nature of the cortical granules in sea urchins (Wong *et al.*, 2007).

Sea urchins share several similar characteristics with sea stars during early development, but the two methods of development quickly diverge. Until the 4th cell cleavage, all blastomeres divide symmetrically, yielding 8 equal blastomeres. At the 4th cell cleavage, the blastomeres at the animal pole divide evenly, while those at the vegetal pole do not. These four blastomeres divide unequally, forming four macromeres and four micromeres; the four micromeres are the most vegetal level of cells. At the 5th cell cleavage, the micromeres divide asymmetrically and

slightly asynchronously compared to the rest of the embryo which produces an intermediate 28 cell embryo (Tanaka and Dan, 1990). Once the micromere 5th cell cleavage is complete it produces the large and small micromere lineages, which have completely different fates. The large micromeres are fated to become the primary mesenchyme cells; responsible for the formation of the larval skeleton. The small micromeres have a completely different fate; mounting evidence supports the hypothesis that the small micromeres are presumptive germ cells (Yajima and Wessel, 2011). Even though a single cell division earlier, the small micromeres shared cytoplasm with a completely somatic cell lineage. Although the small micromere lineage is unique to Euechinoids (sea urchins and sea biscuits), the earlier branching cidaroids (pencil urchins) also share the micromere lineage. In the more basal Cidaroida, the number of micromeres is variable, producing between 1 and 4 micromeres (Bennett *et al.*, 2012).

In comparison to sea urchins, not as much is known about the fertilization and early development in sea stars. After the oocytes have matured into fertilization competent eggs and fertilized, a robust fertilization envelope forms. The formation of the fertilization envelope in sea stars generally takes much longer than sea urchins, approximately 10 minutes depending on the species.

In contrast to the sea urchin, the blastomeres of sea stars divide symmetrically throughout development. Due to the lack of a hyaline layer in sea stars during the first several hours of development (Cameron and Holland, 1983), the fertilization envelope is critical in the development of sea star embryos. Upon removal of the fertilization envelope, blastomeres often disassociate and development is arrested (Matsunaga *et al.*, 2002). During gastrulation, a pocket of cells evaginates from the wall of the archenteron and migrates to the left side during late gastrula. This pocket of cells forms the posterior enterocoel, a structure implicated in germ line formation (Inoue *et al.*, 1992).

Evolutionary history of larval development of echinoderms

Echinoderms diverged from the rest of Bilateria more than 500 million years ago and the adult body plans that we observe in extant echinoderms rapidly radiated only 10-15 million years later (Smith *et al.*, 2013). As such, there is a rich history of adult echinoderms appearing in the fossil record due to the biomineralized skeletons. These fossils not only identify gross adult morphologies but also can identify when specific gene regulatory networks first evolved (Bottjer *et al.*, 2006). The stereom is a mesh-like structure seen in extant echinoderms and first appears in the fossil record during the Cambrian explosion, approximately 520 million years ago (Bottjer *et al.*, 2006; Sevastopulo and Keegan, 1980). Although the adult echinoderms, are well preserved in the fossil record, larval morphologies are not for many reasons, primarily because they are small and comprised almost entirely of soft tissue. However, some characteristics of larval morphology do carry over into adult fossils; in particular the orientation of the spicules of the larval skeleton can inform the development of the adult skeleton in the rudiment if the larval skeleton is not reabsorbed during metamorphosis (Emler, 1985; Emler, 1989; Yajima and Kiyomoto, 2006). In addition to the fossil record, many aspects of larval development can be inferred from evolution.

In extant echinoderms, larval development falls into three categories, although two predominate. The first prevalent form is a lecithotrophic larva, meaning that the larvae do not feed in larval form and subsist on maternally deposited yolk stores until metamorphosis, which normally occurs within hours or days after fertilization. The second dominant form is a filter feeding, planktotrophic larva, which develop functional feeding structures and digestive system and often stay for extended periods of time (days to months) in the water column between fertilization and metamorphosis. A third and fairly rare development strategy is that of a facultative feeder. In this mode of development there is often a large maternal store of yolk but the larva does develop functional feeding apparatuses which are not necessary for larval development and subsequent metamorphosis (reviewed in Smith *et al.*, 2007). Both major strategies come with fitness tradeoffs: large maternal contributions into few progeny provides a high chance of the offspring to mature into adults, while little maternal contribution into

numerous progeny can potentially yield a much greater number of surviving offspring in favorable environmental conditions. A complimentary mode of development to the above is the evolution of nonplanktotrophic larval development. This form of development is also called brooding and is a recent evolution which occurred in the last 100 million years but has poor representation in the fossil record (Smith, 1997).

There is little evidence that the ancestral bilaterian was a planktotrophic form similar to modern larvae which later evolved a benthic adult life stage (Raff, 2008). Rather, the evidence supports the conclusion that larvae evolved in multiple lineages by convergent evolution and co-opted adult genes to form larval structures (Raff, 2008). Furthermore, the ancestral developmental mode of extant echinoderm larva was a planktotrophic feeding larva (Raff and Byrne, 2006). A non-feeding lecithotrophic larva is a more recent evolutionary adaptation and has evolved many independent times (Raff, 1987). There are several cases in which sister taxa of recently diverged species demonstrate lecithotrophic and feeding, planktotrophic larvae (Raff and Byrne, 2006; Smith *et al.*, 2007), but in all examined cases, it appears that the lecithotrophic mode is secondarily gained.

OPEN QUESTIONS

Broadly speaking, before making any comparisons between any two organisms, the most critical step is identifying the exact phylogenetic relationship between the two. Following the establishment of a clear evolutionary history, specific character changes can be examined in an evolutionary context. Instead of simply observing qualitative differences between different organisms, the observations can be tested and interpreted quantitatively. Furthermore, an accurate phylogenetic tree allows for the testing of hypotheses outside of select model organisms because the investigator can select an informative evolutionary node and design an experiment that not

only addresses the question in the organism of choice but also informs results from closely related and well-studied model organisms.

The first question that I must address is: how are extant echinoderms related to each other? Although a significant amount of work has been invested in studying the developmental biology and fertilization of echinoderms (and sea stars and sea urchins in particular), it has always been difficult to frame the conclusions from these studies in an evolutionary context. Several conflicting phylogenetic trees for echinoderms have been proposed, but without having a clear understanding of the evolutionary history of Echinodermata, any conclusions drawn from the data are severely limited. The second question that follows the establishment of a well-supported phylogeny: how has germline specification evolved in Echinodermata? With a well-supported phylogeny, one can begin to test the differences of germline specification and potentially infer the ancestral mechanism. Finally, how have other features of early development evolved within the phylum? Furthermore, the tree will also allow for the identification of other characteristics that have evolved independently or those inherited from a common ancestor.

I address in this body of work my efforts to establish the clear phylogenetic relationships of Echinodermata and the subsequent experiments to study evolutionary transitions of early development and germline determination in the phylum.

REFERENCES

- Bennett, K.C., Young, C.M., and Emlet, R.B. (2012). Larval development and metamorphosis of the deep-sea cidaroid urchin *Cidaris blakei*. *Biol Bull* 222, 105-117.
- Bottjer, D.J., Davidson, E.H., Peterson, K.J., and Cameron, R.A. (2006). Paleogenomics of echinoderms. *Science* 314, 956-960.
- Burns, G., Thorndyke, M.C., Peck, L.S., and Clark, M.S. (2013). Transcriptome pyrosequencing of the Antarctic brittle star *Ophionotus victoriae*. *Mar Genomics* 9, 9-15.
- Cameron, R.A., and Holland, N.D. (1983). Electron microscopy of extracellular materials during the development of a sea star, *Patiria miniata* (Echinodermata: Asteroidea). *Cell Tissue Res* 234, 193-200.
- Candia Carnevali, M.D., and Bonasoro, F. (2001). Microscopic overview of crinoid regeneration. *Microsc Res Tech* 55, 403-426.
- Chiba, K. (2000). Meiosis reinitiation in starfish oocyte. *Zoological Science* 17, 413-417.
- 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.
- Emlet, R.B. (1985). Crystal axes in recent and fossil adult echinoids indicate trophic mode in larval development. *Science* 230, 937-940.
- Emlet, R.B. (1989). Apical skeletons of sea urchins (Echinodermata: Echinoidea): two methods for inferring mode of larval development. *Paleobiology*, 223-254.
- Ernst, S.G. (1997). A century of sea urchin development. *American zoologist* 37, 250-259.
- Gahn, F.J., and Baumiller, T.K. (2010). Evolutionary history of regeneration in crinoids (Echinodermata). *Integr Comp Biol* 50, 514a-514m.
- Inoue, C., Kiyomoto, M., and Shirai, H. (1992). Germ cell differentiation in starfish: the posterior enterocoel as the origin of germ cells in *Asterina pectinifera*. *Dev Growth Differ* 34, 413-418.
- Kanatani, H., Shirai, H., Nakanishi, K., and Kurokawa, T. (1969). Isolation and identification of meiosis inducing substance in starfish *Asterias amurensis*.
- Kouchinsky, A., Bengtson, S., Runnegar, B., Skovsted, C., Steiner, M., and Vendrasco, M. (2012). Chronology of early Cambrian biomineralization. *Geological Magazine* 149, 221.
- Matsunaga, M., Uemura, I., Tamura, M., and Nemoto, S.I. (2002). Role of specialized microvilli and the fertilization envelope in the spatial positioning of blastomeres in early development of embryos of the starfish *Astropecten scoparius*. *Biological Bulletin* 202, 213-222.
- Pearse, J.S., and Cameron, R.A. (1991). Echinodermata: echinoidea.
- Raff, R.A. (1987). Constraint, flexibility, and phylogenetic history in the evolution of direct development in sea urchins. *Dev Biol* 119, 6-19.
- Raff, R.A. (2008). Origins of the other metazoan body plans: the evolution of larval forms. *Philos Trans R Soc Lond B Biol Sci* 363, 1473-1479.
- Raff, R.A., and Byrne, M. (2006). The active evolutionary lives of echinoderm larvae. *Heredity (Edinb)* 97, 244-252.
- Sasaki, K., and Chiba, K. (2001). Fertilization blocks apoptosis of starfish eggs by inactivation of the MAP kinase pathway. *Dev Biol* 237, 18-28.
- Sevastopulo, G., and Keegan, J. (1980). A technique for revealing the stereom structure of fossil crinoids. *Palaeontology* 23, 749-756.
- Smith, A.B. (1997). Echinoderm larvae and phylogeny. *Annual review of ecology and systematics* 28, 219-241.
- Smith, A.B., Zamora, S., and Alvaro, J.J. (2013). The oldest echinoderm faunas from Gondwana show that echinoderm body plan diversification was rapid. *Nat Commun* 4, 1385.
- Smith, M.S., Zigler, K.S., and Raff, R.A. (2007). Evolution of direct-developing larvae: selection vs loss. *Bioessays* 29, 566-571.

- Tanaka, S., and Dan, K. (1990). Study of the lineage and cell cycle of small micromeres in embryos of the sea urchin, *Hemicentrotus pulcherrimus*. *Dev Growth Differ* 32, 145-156.
- Wessel, G.M., Reich, A.M., and Klatsky, P.C. (2010). Use of sea stars to study basic reproductive processes. *Syst Biol Reprod Med* 56, 236-245.
- Wong, J.L., Koppel, D.E., Cowan, A.E., and Wessel, G.M. (2007). Membrane hemifusion is a stable intermediate of exocytosis. *Developmental Cell* 12, 653-659.
- Wong, J.L., and Wessel, G.M. (2006a). Defending the zygote: search for the ancestral animal block to polyspermy. *Curr Top Dev Biol* 72, 1-151.
- Wong, J.L., and Wessel, G.M. (2006b). Rendezvin: An essential gene encoding independent, differentially secreted egg proteins that organize the fertilization envelope proteome after self-association. *Mol Biol Cell* 17, 5241-5252.
- Wong, J.L., and Wessel, G.M. (2008). Free-radical crosslinking of specific proteins alters the function of the egg extracellular matrix at fertilization. *Development* 135, 431-440.
- Yajima, M., and Kiyomoto, M. (2006). Study of larval and adult skeletogenic cells in developing sea urchin larvae. *Biol Bull* 211, 183-192.
- Yajima, M., and Wessel, G.M. (2011). Small micromeres contribute to the germline in the sea urchin. *Development* 138, 237-243.

**Chapter II: Phylogenetic analysis of extant echinoderms using *de novo*
transcriptomes**

Adrian Reich

Unpublished

CONTRIBUTION

I conducted all experiments and analyses.

ABSTRACT

Echinoderms (sea urchins, sea stars, brittle stars, sea lilies and sea cucumbers) are a very rich group of organisms, second only in the number of species in Deuterostomia after chordates. Echinoderms serve as excellent model systems for developmental biology due to their diverse developmental mechanisms, tractable laboratory use, and close phylogenetic distance to chordates. In addition, echinoderms are very well represented in the fossil record, including some larval features, making echinoderms a valuable system for studying evolutionary development. However, the phylogenetic relationships within the phylum have been contentious; generating little overlap between molecular, morphological and combined analyses. In order to resolve the controversies, we sequenced 23 *de novo* transcriptomes from all five classes of echinoderms. Using multiple phylogenetic methods at a variety of sampling depths we have constructed a well-supported phylogenetic tree of Echinodermata, including support for the sister groups of Asterozoa (sea stars and brittle stars) and Echinozoa (sea urchins and sea cucumbers). The larger of the two analyzed datasets includes 630,945 amino acid sites across 4,645 peptide sequences and 30 taxa. These results will help inform developmental and evolutionary studies specifically in echinoderms and deuterostomes in general.

INTRODUCTION

Echinoderms are an important group of organisms that are closely related to chordates and have been an invaluable model system for developmental biology for hundreds of years (Derbès, 1847; Ernst, 1997). However, the phylogenetic relationships of these organisms within the phylum have been controversial (Janies, 2001; Janies *et al.*, 2011; Pisani *et al.*, 2012; Smith, 1984). This is due in part, to the very rapid radiation of Echinodermata that occurred only 10-15 million years after the emergence of echinoderms more than 500 million years ago (Smith *et al.*, 2013).

In recent years, two competing hypotheses have emerged, (Fig. 1) with the major point of contention being the placement of the ophiuroids (brittle stars). Different methods used for classifying the phylogenetic relationships (molecular, morphology, embryological or combined analyses) recover conflicting tree topologies. Morphological and embryological analyses support the Cryptosyringida hypothesis, (Janies, 2001; Raff and Byrne, 2006), a monophyletic clade comprised of Echinoidea, Holothuroidea and Ophuroidea (Fig. 1). This was first formalized in name by Smith (1984), but was originally proposed much earlier (Mac Bride, 1906). In contrast, the Asterozoan hypothesis is supported by combined analyses, using molecular and DNA analyses (Janies, 2001), and was first proposed by Bather (1900). A more recent study found that using micro RNAs was surprisingly uninformative, however using other molecularly based analyses provided support for the Cryptosyringida hypothesis (Pisani *et al.*, 2012). The difficulties in previous studies are twofold for both morphological and molecular approaches. In morphological or embryological analyses, a large number of traits can be used but they are often binary, leaving the analyses sensitive to perhaps a single trait; furthermore these analyses are not very efficient at resolving relationships within a clade. In molecular analyses (using conserved genes, ribosomal sequences or mitochondrial sequences and/or gene order), sampling within an individual taxon is often very shallow (often no more than a dozen genes) and the total number of taxa can also be very low; sometimes only one or two representatives from each clade. The trees

recovered from these molecular phylogenies are therefore very sensitive to the analysis software and parameters used; recently, Janies *et al.* (2011) recovered three different tree topologies and Pisani *et al.* (2012) found support for two.

Deep sequencing and *de novo* assembly of transcriptomes can simultaneously analyze large numbers of sequences, across many taxa and can robustly recover relationships of taxa, both within and between clades (Dunn *et al.*, 2008; Dunn *et al.*, 2013; Hejnol *et al.*, 2009), at the expense of greatly increasing computational resources. In addition, high throughput sequencing technologies can identify rapidly evolving sequences including those under positive selection (Claw *et al.*, 2014; Palmer *et al.*, 2013). In the current study, we sequenced and assembled *de novo* transcriptomes of ovary tissue of 23 different species of echinoderms, 18 of which had never been sequenced before, and compared them to 7 RefSeq datasets including 1 echinoderm and 6 outgroups. RefSeq was chosen as the database because they are non-redundant and manually curated in a standard manner, and contain numerous species. In addition, we sequenced and assembled testis transcriptomes of two different sea stars to test for sperm and egg sequences that are co-evolving.

RESULTS AND DISCUSSION

Transcriptome assemblies

The assembled raw transcripts from the individual datasets were reduced by comparing the transcripts against the curated and annotated SwissProt database by BLAST. This data reduction was done to remove any spuriously assembled transcripts but was also necessary to efficiently run downstream analyses. The assembly of the *de novo* transcriptomes (excluding the two transcriptomes assembled using high throughput pyrosequencing (454, Roche)) yielded on average 12,000 high quality contigs with a match to SwissProt (Supplemental Table 1). The number of SwissProt transcripts in the *de novo* transcriptomes was comparable to the RefSeq datasets (Supplemental Fig. 2a). This was quite surprising given that the transcriptomes were

assembled from a single tissue and therefore are expected to be a subset of the total transcriptome. In contrast, the RefSeq samples are gene predictions from the entire sequenced genome, which can contain all possible transcripts. Furthermore, the number of SwissProt sequences did not appear to be artificially inflated, which could indicate a fragmented transcriptome (Supplemental Fig. 2b). With a fragmented transcriptome, two truncated RNA fragments from the same mRNA could match the same SwissProt sequence, thereby double counting the hit and having an incomplete mRNA model. Another assembly metric is N50, or the contig size in which 50% of the total transcriptome length is found in contigs smaller than the N50 contig size and 50% is in contigs larger than the N50 contig size. The average N50s of the SwissProt sequences from the *de novo* transcriptomes was 2.2kb, only slightly smaller than the 3.3kb average of the RefSeq datasets (Supplemental Fig.2b and Supplemental Table 1). The comparable number of SwissProt sequences and comparable size of the N50s of the *de novo* transcriptomes in relation to the RefSeq datasets, both suggest that the assembled transcriptomes are of high quality and complete.

Phylogenetic relationship of extant echinoderms

Our results strongly support the Asterozoan hypothesis of echinoderm evolution (Fig. 1 and 2) using multiple phylogenetic methods and sampling depths. In addition, the 30 taxa represented in our analyses span more than 550 million years of evolution including a non-bilaterian outgroup, and yet 85% of all the nodes have full support from all three analyses (Fig. 2). Two minor differences occur between the results generated from the two different RAxML analyses and the PhyloBayes analysis. The first is the incorrect placement of *A. californica* in the PhyloBayes analysis. *A. californica* (a Lophotrochozoan) was deemed more closely related to the chordate outgroups (*e.g. H. sapiens*) than *Saccoglossus kowalevskii* (a hemichordate; Fig. 2). The second minor difference is the internal tree arrangement of the brittle stars. In the two RAxML analyses, the genus *Ophiocoma* is polyphyletic, while in the PhyloBayes analysis the members of the genus are sister to each other. These minor differences are likely due to different models of

sequence evolution used in the RAxML and PhyloBayes analyses (WAG and CAT, respectively). With respect to the paraphyletic *Ophiocoma* genus in RAxML, this is likely due to poor sampling in 2 of the 3 brittle star taxa (<2,000 SwissProt sequences for *O. wendtii* and *O. victoriae* compared with >20,000 SwissProt sequences in *O. echinata*; Supplemental Table 1). However, even with the shallower sampling, this does not diminish the confidence of the placement of Ophuroidea as sister to Asteroidea. All three phylogenetic analyses (with different sampling depth, aligned genes, analysis programs and matrix occupancy) identified a monophyletic Ophuroidea that was sister to a monophyletic Asteroidea (Fig. 2).

Resolution of the order Paxillosida in Asteroidea

A particularly interesting result is the internal phylogenetic relationships of the sea stars; which differed significantly with a number of previous analyses (Hart *et al.*, 1997; Knott and Wray, 2000; Mah, 2000; Matsubara *et al.*, 2005; Wada *et al.*, 1996). Furthermore, a number of other analyses were inconclusive due to a combination of insufficient overlap of taxa and/or polytomic relationships (Byrne, 2006; Hart *et al.*, 2004; Mah and Foltz, 2011; Matsubara *et al.*, 2004; O'Loughlin and Waters, 2004). We recovered monophyletic groups for the orders Valvatida, Spinulosida, and Forcipulatida (Fig. 2, Supplemental Table 2). However, we found evidence against the subclass Valvatacea as the three organisms in this subclass are paraphyletic relative to Spinulosacea with a minimum support of 98 from all three analyses. Similarly to the recent analysis by Foltz *et al.* (2007) and others, (Clark and Downey, 1992; Fisher, 1928), we recovered a monophyletic Forcipulatacea subclass (Supplemental Table 2). This was not surprising given that all members in this study came from the same family (Asteroidea), however, even within Forcipulatacea, we identified discrepancies with previous studies (Mah, 2000), specifically the genus *Pisaster*. All three analyses in this study supported the genus *Pisaster* as sister group to the common ancestor of the genera *Asterias* and *Leptasterias* (Fig. 2).

The largest difference between many of previous studies and these results was the placement of the order Paxillosida (represented in this study by *Luidia clathrata*). In the analyses

reported here, Paxillosida was sister to the superorder of Valvatida (e.g. *Patiria miniata*) and Spinulosida (e.g. *Henricia* sp.), which were in turn sister to each other (Fig. 2, Supplemental Table 2). This is in contrast to the previous placement of Paxillosida as basal to Forcipulatida (Knott and Wray, 2000), as basal to all of Asteroidea (Wada *et al.*, 1996), or as more derived (Matsubara *et al.*, 2005). The placement of Paxillosida as a more basal with respect to Valvatida and Spinulosida presents several interesting hypotheses considering the morphological features unique to Paxillosida (lack of anus and suckers on tube feet; Gale, 1987; Jangoux and Lawrence, 1982). Either the ancestral Asteroidea evolved these features and were secondarily lost in Paxillosida (two changes), or that these features independently evolved in Forcipulatida and the common ancestor of Valvatida and Spinulosida (also two changes).

Morphological and character trait changes

There are many documented character changes in echinoderms that in the past have been used to study phylogenetic relationships of echinoderms. Instead of using the character traits to infer relationships (Janies, 2001), we mapped the character changes to the phylogenetic tree to identify instances of convergent evolution. Mapping these morphological characteristics, a significant amount of homoplasy has occurred; especially between Ophuroidea (brittle stars) and Echinoidea (sea urchins and sand dollars; Janies, 2001; Fig. 3 and Supplemental Table 3). Of six character traits that have changed along the Ophuroidea branch, three of those same characteristics are also common to sea urchins (which has seven trait changes in the Echinoidea branch). It may also be possible that these shared characteristics have not evolved independently in these two classes but were found in the last common ancestor of Asterozoa and Echinozoa and were secondarily lost in Asteroidea and Holothuridea. There is also some convergent evolution that has occurred between Crinoidea (sea lilies and feather stars) and Holothuridea (sea cucumbers); two of three character traits in sea cucumbers are also found in crinoids (total of four trait changes).

CONCLUSIONS

The phylogenetic tree constructed from our data identifies with strong support the overall relationships of all five extant classes of echinoderms. In addition several contentious relationships within Asteroidea in particular have also been resolved with strong support. Previously, the sea urchin *S. purpuratus* was the only echinoderm with a significant number of available annotated sequences (Sea Urchin Genome Sequencing *et al.*, 2006) and was the *de facto* echinoderm representative in phylogenetic studies of deuterostomes. However, *S. purpuratus* has a number of derived characteristics that do not necessarily represent echinoderms as a whole (Fig. 3 and Supplemental Table 3). As such, the data presented here will more accurately represent the echinoderm phylum in studying the evolutionary history of deuterostomes, and greatly facilitate the exploration of the diverse biology within Echinodermata.

FUTURE DIRECTIONS

Positive selection in Asteroidea

Beyond establishing the phylogenetic relationships of echinoderms, this dataset can be used for additional experiments that have great potential to yield interesting results. Of particular interest is to test for genes that are rapidly evolving and undergoing positive selection. In designing this experiment, ovary tissue was selected for several reasons; foremost was the hypothesis that the maternally deposited mRNA in the eggs would be enriched for female specific gamete recognition proteins. A number of closely related species is critical for this analysis in order to minimize the effects of long branch attraction or the apparent similarity of an amino acid between organisms when two or more mutations could have occurred (a change and a subsequent reversion to the original state). Examining Asteroidea would yield the cleanest results because there is a great degree of phylogenetic resolution (11 taxa represented). Furthermore, the sea star dataset has the best potential of yielding testable hypotheses, because for two of the species (*A. amerensius* and *A. pectinifera*), I have sequenced, assembled and annotated *de novo* testis

transcriptomes (data not shown). Using the same methods as above, I can test these samples for sequences undergoing positive selection and subtract all sequences that are expressed in both ovary and testis. This could potentially identify sequences that are gamete specific and under positive selection, both hallmarks of sperm/egg recognition sequences.

These ovary and testis specific sequences that are under positive selection can also be tested for co-evolution by analyzing the correlation of d_N/d_S values between pairs of male and female specific sequences that are under positive selection (Claw *et al.*, 2014).

MATERIALS AND METHODS

RNA isolation and sequencing

Whole ovary was dissected from gravid females and put in Trizol (Invitrogen). RNA was extracted and then cleaned with a Qiagen RNeasy Micro column with on-column DNA digestion. The sequencing libraries were prepared with Illumina reagents, (mRNA-Seq Sample Prep Kit for GAIx samples or TruSeq Sample Prep Kit for HiSeq samples) with the maximum recommended RNA input. The protocol was followed exactly with the addition of a gel selection step of 400-500bp, (agarose gel for GAIx samples or Caliper LabChip XT for HiSeq samples) prior to PCR amplification.

Transcriptome assemblies and RefSeq data

Reads were processed and assembled using the agalma pipeline (ver. 0.3.5, <https://bitbucket.org/caseywdunn/agalma>; Dunn *et al.*, 2013; Howison *et al.*, 2012) which wraps the trinity *de novo* transcriptome assembler (ver. r2013_08_14; Grabherr *et al.*, 2011; Haas *et al.*, 2013). The default settings were used within agalma to assemble all transcriptomes. A total of three echinoderm public datasets were also *de novo* assembled and used in this analysis. One dataset consisting of combined 2 day (SRR496203) and 6 day (SRR496204) old larvae of *Parastichopus parvimensis* (Le *et al.*, 2013) was assembled as above. Two additional brittle star 454 datasets were assembled using newbler (DataAnalysis ver. 2.9). Reads isolated from

regenerating arms of *Ophionotus victoriae* (Burns *et al.*, 2013; SRR500294) were assembled as follows: “runAssembly -cdna”; reads from gastrula larvae of *Ophiocoma wendtii* (Vaughn *et al.*, 2012; Brian Livingston personal communication) were assembled as follows: “runAssembly -cpu 8 -vt Adapters.fasta -cdna” (for Adapters.fasta, see Supplemental Table 4). Exemplar transcripts were then selected from the assembled 454 datasets as previously described (Smith *et al.*, 2011). Seven RefSeq datasets were downloaded: *Aplysia californica*, *Branchiostoma floridae*, *Gallus gallus*, *Homo sapiens*, *Nematostella vectensis*, *Saccoglossus kowalevskii*, and *Strongylocentrotus purpuratus*. All except for the human dataset were from RefSeq release 62, dated Nov. 10, 2013; the human dataset was downloaded Dec. 9, 2013.

Post assembly and phylogenetic analyses

All 30 datasets (23 *de novo* assembled transcriptomes and 7 RefSeq datasets; Supplemental Table 1) were translated and compared against NCBI SwissProt using BLASTp with a cutoff of 0.00001 (Dunn *et al.*, 2013). Similar sequences were identified by a pair wise BLASTp followed by clustering using MCL (Enright *et al.*, 2002) to identify orthologous genes. Two different supermatricies were used in the phylogenetic analyses: ‘sparse’ and ‘dense’. The ‘sparse’ supermatrix has all 30 taxa and 34% matrix occupancy, (Supplemental Fig. 3a) with 4,645 peptide sequences and 630,945 amino acid sites. The ‘dense’ supermatrix has all 30 taxa and 70% matrix occupancy, (Supplemental Fig. 3b) with 1,125 peptide sequences and 101,652 amino acid sites. Maximum likelihood phylogenetic analyses were done with RAxML (WAG model) with 1,000 boot strap iterations on the ‘dense’ supermatrix, and using the same model, 100 boot strap iterations on the ‘sparse’ supermatrix.

Baysian phylogenetic analyses were done with PhyloBayes 1.3b-mpi (Lartillot *et al.*, 2009) on the ‘dense’ supermatrix using the CAT-GTR model with the following command: “pb_mpi -S -d supermatrix.dense.phylip -cat -gtr outputFile”. A total of 31,623 generations were run over three chains. All three chains converged within 2,000 generations and after removing

these 2,000 from each chain and sampling every 10 trees, the maximum difference was 1.17×10^{-3} , (2,561 sampled trees), and a majority consensus tree was constructed.

DATA AVAILABILITY

The raw reads and assembled transcripts reported in this paper have been deposited in the GenBank database (NCBI BioProject no. PRJNA236087). Assembly statistics and agalma resource reports can be found at: <https://bitbucket.org/AdrianReich/phylogenetic-analysis-of-echinoderms>. The *de novo* transcripts can also be accessed at: <http://www.echinobase.org/>.

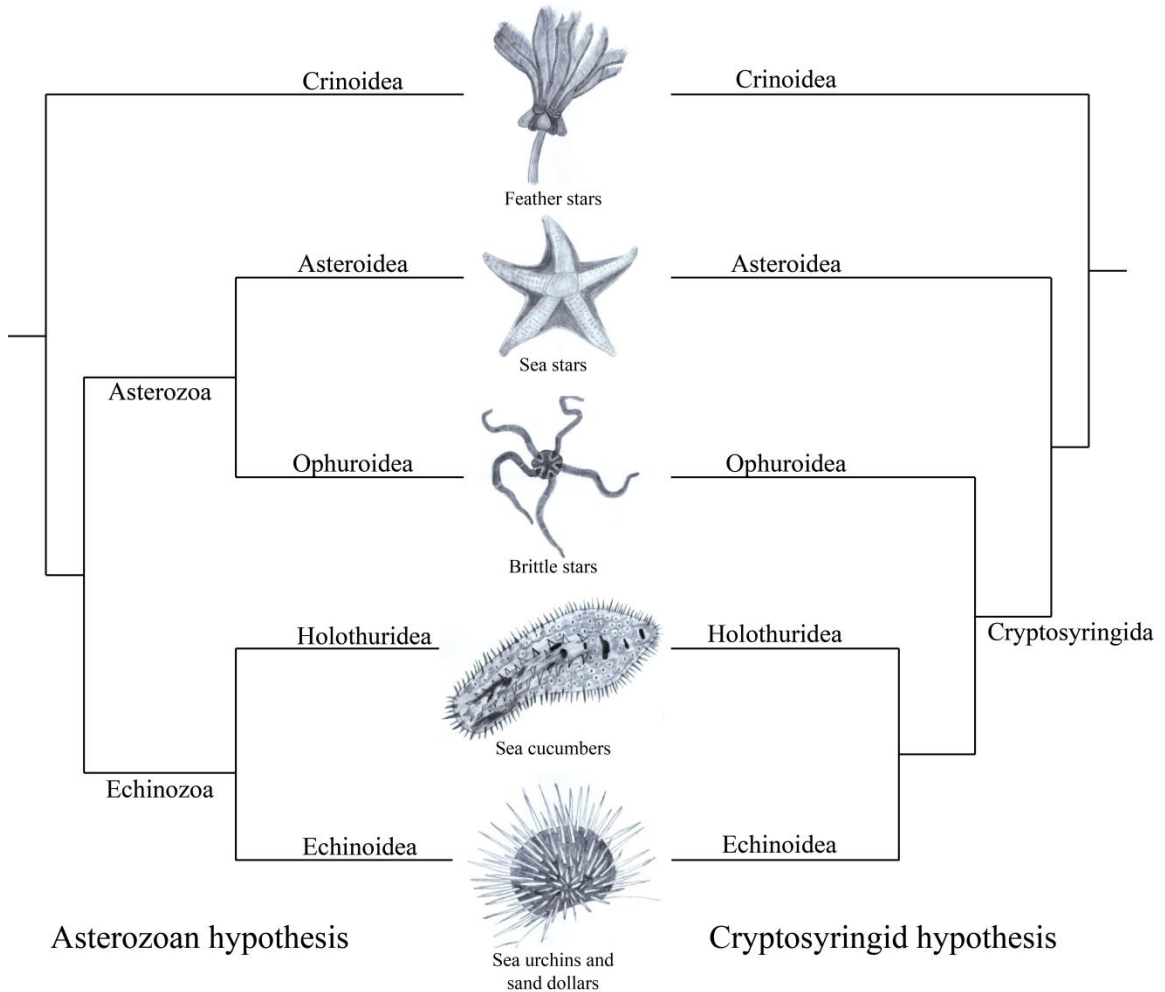
REFERENCES

- Bather, F. (1900). The Echinodermata: Treatise on Zoology, pt. 3.
- Burns, G., Thorndyke, M.C., Peck, L.S., and Clark, M.S. (2013). Transcriptome pyrosequencing of the Antarctic brittle star *Ophionotus victoriae*. *Mar Genomics* 9, 9-15.
- Byrne, M. (2006). Life history diversity and evolution in the Asterinidae. *Integrative and Comparative Biology* 46, 243-254.
- Clark, A.M., and Downey, M.E. (1992). Starfishes of the Atlantic (Chapman & Hall).
- Claw, K.G., George, R.D., and Swanson, W.J. (2014). Detecting coevolution in mammalian sperm-egg fusion proteins. *Mol Reprod Dev* *in press*.
- 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.
- Dunn, C.W., Hejnal, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., *et al.* (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452, 745-749.
- Dunn, C.W., Howison, M., and Zapata, F. (2013). Agalma: an automated phylogenomics workflow. *BMC Bioinformatics* 14, 330.
- Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30, 1575-1584.
- Ernst, S.G. (1997). A century of sea urchin development. *American zoologist* 37, 250-259.
- Fisher, W. (1928). Asteroidea of the North Pacific and adjacent waters: US Natl. Mus. Bull 76.
- Foltz, D.W., Bolton, M.T., Kelley, S.P., Kelley, B.D., and Nguyen, A.T. (2007). Combined mitochondrial and nuclear sequences support the monophyly of forcipulatacean sea stars. *Mol Phylogenet Evol* 43, 627-634.
- Gale, A.S. (1987). Phylogeny and classification of the Asteroidea (Echinodermata). *Zoological Journal of the Linnean Society* 89, 107-132.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., *et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29, 644-652.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., *et al.* (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8, 1494-1512.
- Hart, M.W., Byrne, M., and Smith, M.J. (1997). Molecular phylogenetic analysis of life-history evolution in asterinid starfish. *Evolution*, 1848-1861.
- Hart, M.W., Johnson, S.L., Addison, J.A., and Byrne, M. (2004). Strong character incongruence and character choice in phylogeny of sea stars of the Asterinidae. *Invertebrate Biology* 123, 343-356.
- Hejnal, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G.W., Edgecombe, G.D., Martinez, P., Baguna, J., Bailly, X., Jondelius, U., *et al.* (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc Biol Sci* 276, 4261-4270.
- Howison, M., Sinnott-Armstrong, N., and Dunn, C.W. (2012). BioLite, a lightweight bioinformatics framework with automated tracking of diagnostics and provenance. *Proceedings of the 4th USENIX Workshop on the Theory and Practice of Provenance (TaPP '12)*.
- Jangoux, M., and Lawrence, J.M. (1982). *Echinoderm nutrition* (CRC Press).
- Janies, D. (2001). Phylogenetic relationships of extant echinoderm classes. *Canadian Journal of Zoology* 79, 1232-1250.
- Janies, D.A., Voight, J.R., and Daly, M. (2011). Echinoderm phylogeny including *Xyloplax*, a progenetic asteroid. *Syst Biol* 60, 420-438.
- Knott, K.E., and Wray, G.A. (2000). Controversy and consensus in asteroid systematics: new insights to ordinal and familial relationships. *American zoologist* 40, 382-392.

- Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25, 2286-2288.
- Le, H.S., Schulz, M.H., McCauley, B.M., Hinman, V.F., and Bar-Joseph, Z. (2013). Probabilistic error correction for RNA sequencing. *Nucleic Acids Res* 41, e109.
- Mac Bride, E.W. (1906). *Echinodermata* (Macmillan & Company).
- Mah, C., and Foltz, D. (2011). Molecular phylogeny of the Valvatacea (Asteroidea: Echinodermata). *Zoological Journal of the Linnean Society* 161, 769-788.
- Mah, C.L. (2000). Preliminary phylogeny of the forcipulatacean Asteroidea. *American zoologist* 40, 375-381.
- Matsubara, M., Komatsu, M., Araki, T., Asakawa, S., Yokobori, S., Watanabe, K., and Wada, H. (2005). The phylogenetic status of Paxillosida (Asteroidea) based on complete mitochondrial DNA sequences. *Mol Phylogenet Evol* 36, 598-605.
- Matsubara, M., Komatsu, M., and Wada, H. (2004). Close relationship between Asterina and Solasteridae (Asteroidea) supported by both nuclear and mitochondrial gene molecular phylogenies. *Zoological Science* 21, 785-793.
- O'Loughlin, P.M., and Waters, J.M. (2004). A molecular and morphological revision of genera of Asterinidae (Echinodermata: Asteroidea). *Memoirs of Museum Victoria* 61, 1-40.
- Palmer, M.R., McDowall, M.H., Stewart, L., Ouaddi, A., MacCoss, M.J., and Swanson, W.J. (2013). Mass spectrometry and next-generation sequencing reveal an abundant and rapidly evolving abalone sperm protein. *Mol Reprod Dev* 80, 460-465.
- Pisani, D., Feuda, R., Peterson, K.J., and Smith, A.B. (2012). Resolving phylogenetic signal from noise when divergence is rapid: a new look at the old problem of echinoderm class relationships. *Mol Phylogenet Evol* 62, 27-34.
- Raff, R.A., and Byrne, M. (2006). The active evolutionary lives of echinoderm larvae. *Heredity (Edinb)* 97, 244-252.
- Sea Urchin Genome Sequencing, C., Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess, D.R., *et al.* (2006). The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314, 941-952.
- Smith, A.B. (1984). Classification of the Echinodermata. *Palaeontology* 27, 431-459.
- Smith, A.B., Zamora, S., and Alvaro, J.J. (2013). The oldest echinoderm faunas from Gondwana show that echinoderm body plan diversification was rapid. *Nat Commun* 4, 1385.
- Smith, S.A., Wilson, N.G., Goetz, F.E., Feehery, C., Andrade, S.C., Rouse, G.W., Giribet, G., and Dunn, C.W. (2011). Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* 480, 364-367.
- Vaughn, R., Garnhart, N., Garey, J.R., Thomas, W.K., and Livingston, B.T. (2012). Sequencing and analysis of the gastrula transcriptome of the brittle star *Ophiocoma wendtii*. *Evodevo* 3, 19.
- Wada, H., Komatsu, M., and Satoh, N. (1996). Mitochondrial rDNA phylogeny of the asteroidea suggests the primitiveness of the paxillosida. *Mol Phylogenet Evol* 6, 97-106.

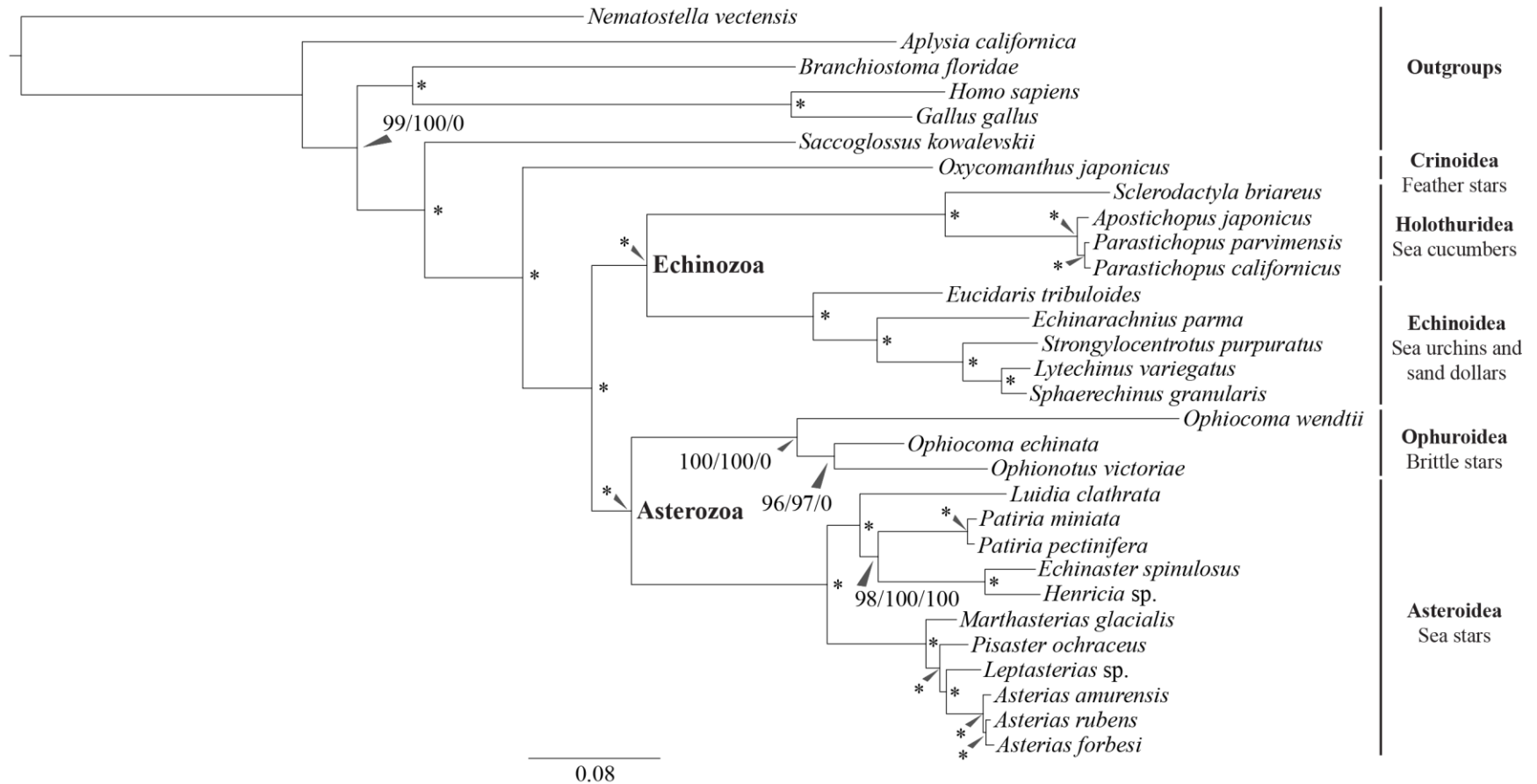
FIGURES AND TABLES

Figure 1: Two competing hypotheses of the phylogenetic relationship of extant echinoderms.



The predominant, mutually exclusive hypotheses of how extant echinoderms are related. Most phylogenetic studies place Echinoidea and Holothuridea as sister group, and Crinoidea as the earliest branching. The difficulty lies in the placement of Ophuroidea; different methods favor different positions for brittle stars.

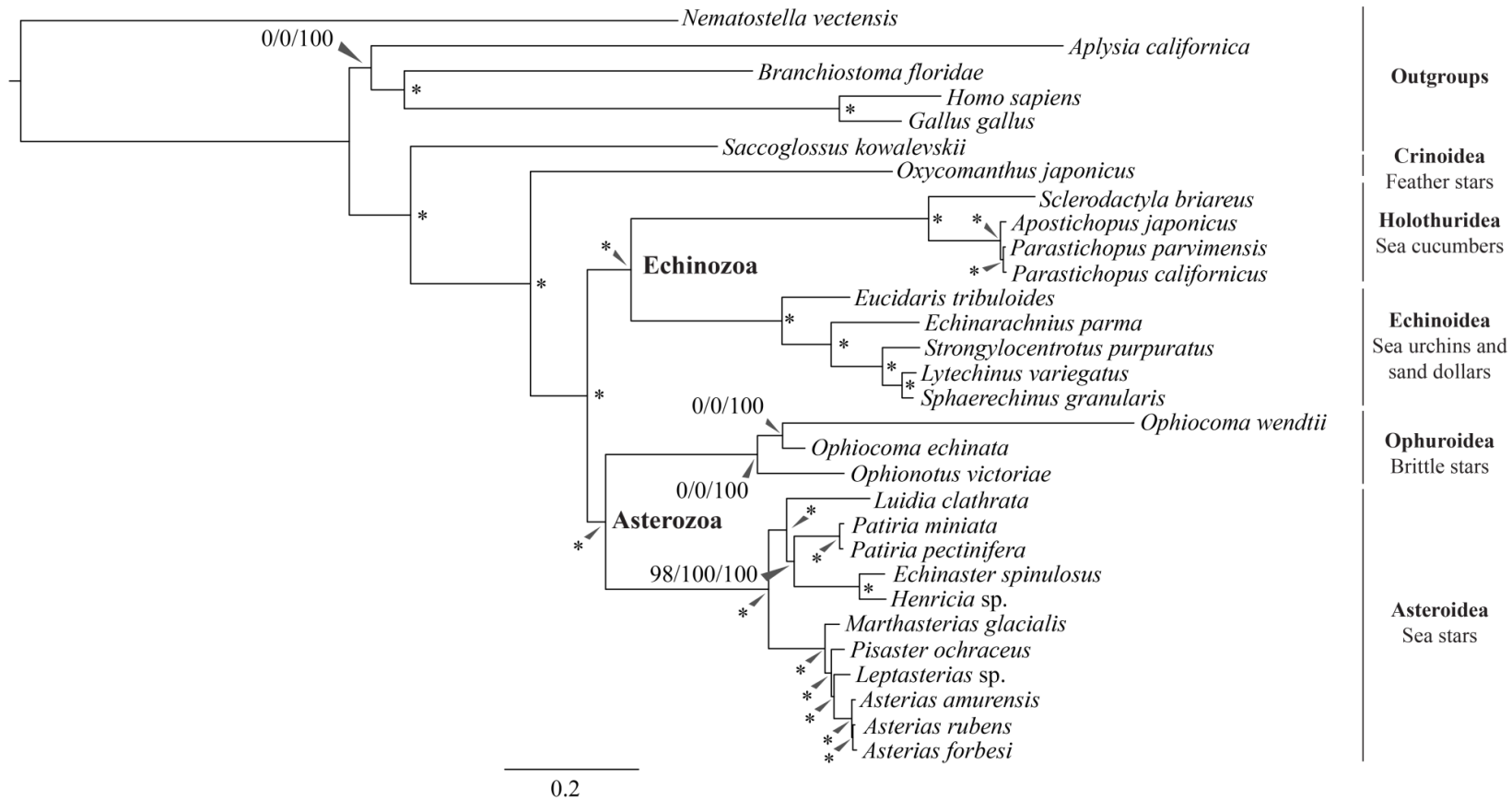
Figure 2: Phylogenetic relationship of extant echinoderms.



Support values for the phylogenetic trees using RAxML and PhyloBayes on the dense and sparse supermatrices. Nodes are scored with support values: dense supermatrix RAxML 1000 bootstraps/ sparse supermatrix RAxML 100 bootstraps/dense supermatrix PhyloBayes posterior probabilities; asterisks denote 100/100/100 support. This tree, scale bar, and subsequent branch lengths presented here are from the dense supermatrix RAxML analysis. See Supplemental Fig. 1 for the tree topology predicted by the PhyloBayes analysis.

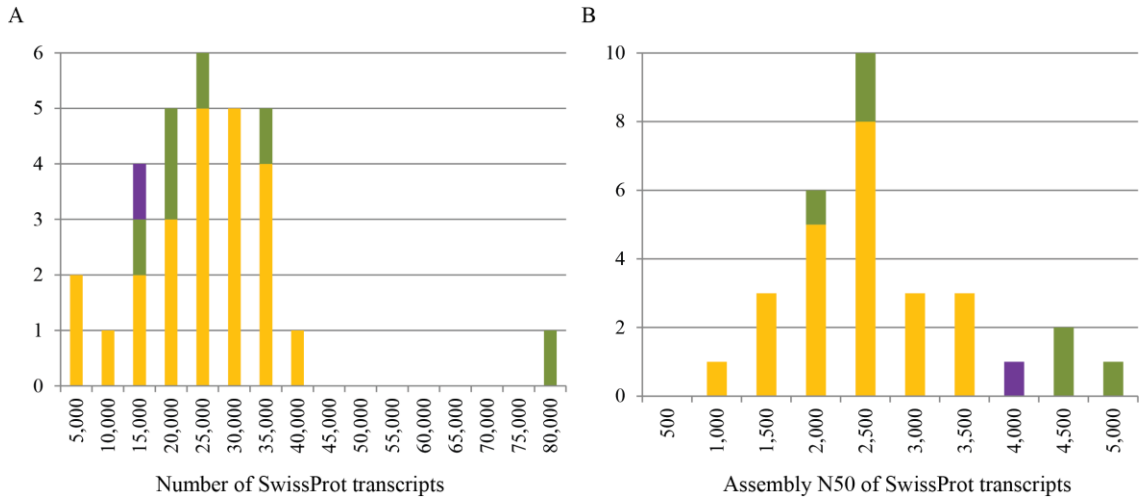
SUPPLEMENTAL INFORMATION

Supplemental Figure 1: Phylogenetic relationship of extant echinoderms.



Support values for the phylogenetic trees using RAxML and PhyloBayes on the dense and sparse supermatrices. Presentation as in Figure 2, however, this tree, scale bar, and subsequent branch lengths presented here are from the dense supermatrix PhyloBayes analysis. Nodes are scored with support values: dense supermatrix RAxML 1000 bootstraps/ sparse supermatrix RAxML 100 bootstraps/dense supermatrix PhyloBayes posterior probabilities; asterisks denote 100/100/100 support.

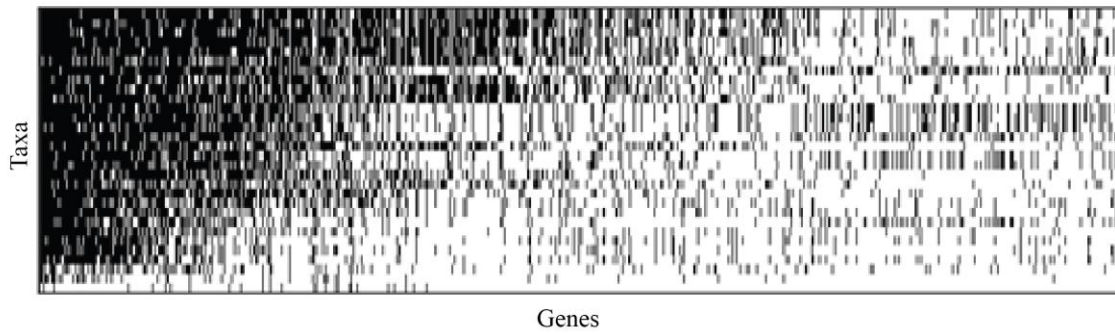
Supplemental Figure 2: Postassembly comparisons of RefSeq and *de novo* assembled datasets.



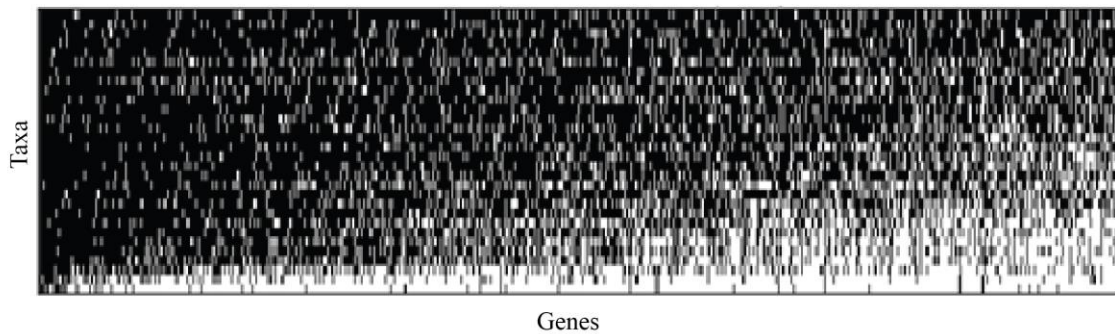
A) The numbers of SwissProt transcripts is comparable between RefSeq datasets (green and purple) and *de novo* assembled transcriptomes (orange). B) Comparing the N50 of the SwissProt transcripts, the *de novo* transcriptomes are on average only slightly smaller than the RefSeq datasets. The *S. purpuratus* RefSeq dataset is in purple, outgroup RefSeq datasets in green and *de novo* assembled transcriptomes in orange; colors as in Supplemental Table 1.

Supplemental Figure 3: Sparse and dense supermatrices including all thirty taxa.

A

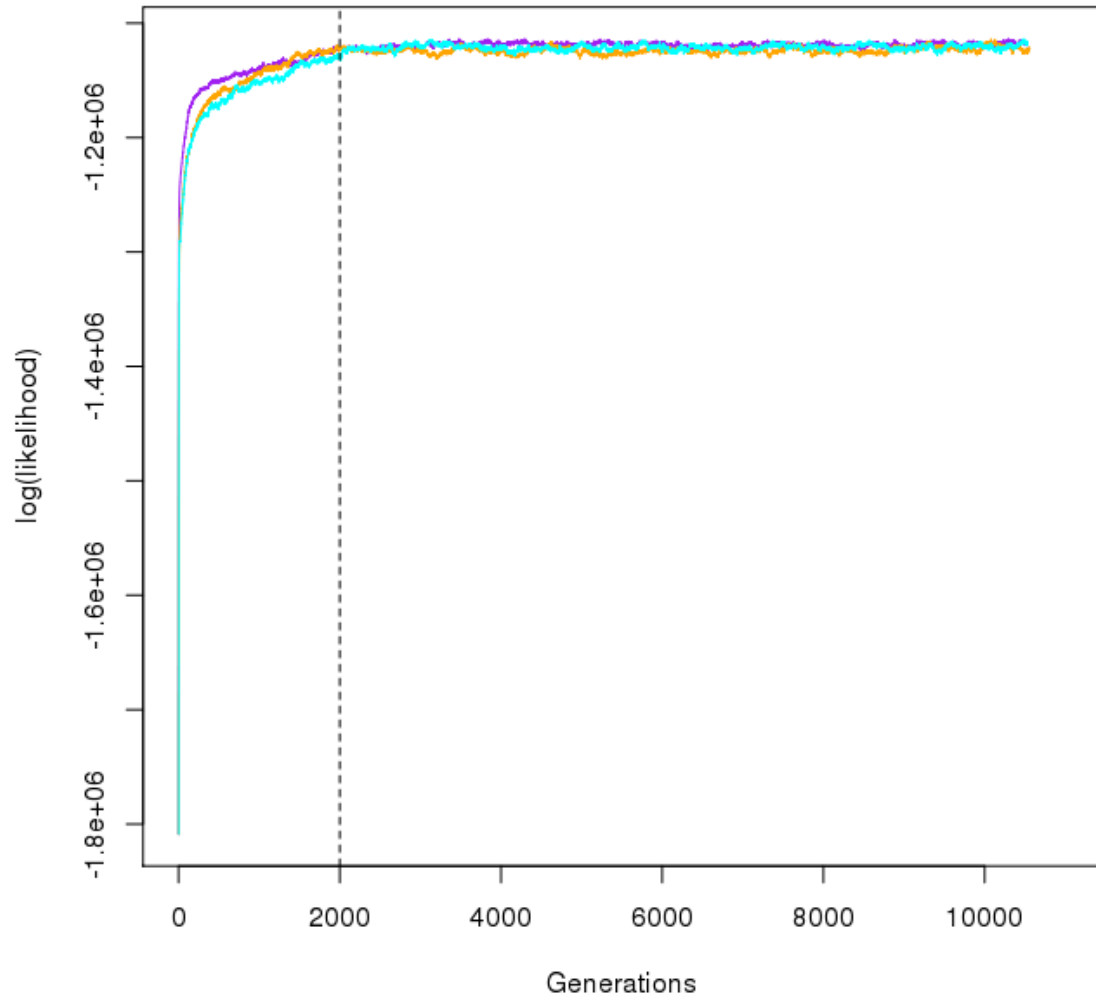


B



Visual representation of the supermatrices. Each horizontal row is a single taxa and each vertical column is a gene alignment; presence is marked in black and absence in white. The taxa are arranged from top to bottom from most genes present to least. A) The sparse supermatrix is 34% occupied, contains all 30 taxa, and contains alignments of 4,645 peptide sequences. B) The dense supermatrix is 70% occupied, contains all 30 taxa, and contains alignments of 1,125 peptide sequences.

Supplemental Figure 4: Test of convergence of PhyloBayes chains.



All three independent chains converge with a maximum difference of 1.17×10^{-3} after a burn-in of 2000 generations (dashed line).

Supplemental Table 1: Assembly and post assembly summary by taxa.

| Species | Assembled Transcripts | Sample Source | SwissProt Transcripts | SwissProt N50 |
|--------------------------------------|-----------------------|---------------|-----------------------|---------------|
| <i>Aplysia californica</i> | 26,249 | RefSeq | 18,424 | 4,099 |
| <i>Branchiostoma floridae</i> | 28,575 | RefSeq | 21,881 | 2,127 |
| <i>Gallus gallus</i> | 36,995 | RefSeq | 31,868 | 4,527 |
| <i>Homo sapiens</i> | 91,944 | RefSeq | 77,696 | 4,208 |
| <i>Nematostella vectensis</i> | 24,462 | RefSeq | 18,261 | 1,609 |
| <i>Saccoglossus kowalevskii</i> | 12,851 | RefSeq | 10,936 | 2,294 |
| <i>Strongylocentrotus purpuratus</i> | 23,078 | RefSeq | 19,554 | 4,363 |
| <i>Ophiocoma wendtii</i> | 5,025 | 454 | 1,995 | 771 |
| <i>Ophionotus victoriae</i> | 3,210 | 454 | 694 | 1,645 |
| <i>Parastichopus parvimensis</i> | 107,585 | 72bp PE | 34,809 | 3,076 |
| <i>Eucidaris tribuloides</i> | 45,385 | 80bp PE | 7,410 | 1,307 |
| <i>Patiria miniata</i> | 76,847 | 80bp PE | 24,679 | 2,494 |
| <i>Oxycomanthus japonicus</i> | 39,225 | 80bp PE | 15,868 | 2,816 |
| <i>Ophiocoma echinata</i> | 111,491 | 80bp PE | 21,485 | 2,454 |
| <i>Lytechinus variegatus</i> | 90,621 | 80bp PE | 25,994 | 2,914 |
| <i>Asterias forbesi</i> | 68,714 | 80bp PE | 22,625 | 2,784 |
| <i>Sclerodactyla briareus</i> | 58,273 | 80bp PE | 19,297 | 3,154 |
| <i>Asterias rubens</i> | 81,470 | 100bp PE | 25,733 | 2,349 |
| <i>Henricia species</i> | 137,160 | 100bp PE | 36,754 | 2,231 |
| <i>Echinaster spinulosus</i> | 119,580 | 100bp PE | 34,231 | 2,314 |
| <i>Echinarachnius parma</i> | 96,977 | 100bp PE | 26,016 | 2,017 |
| <i>Leptasterias species</i> | 108,544 | 100bp PE | 33,803 | 3,172 |
| <i>Luidia clathrata</i> | 84,380 | 100bp PE | 23,407 | 1,835 |
| <i>Marthasterias glacialis</i> | 118,847 | 100bp PE | 28,327 | 1,928 |
| <i>Pisaster ochraceus</i> | 37,111 | 100bp PE | 10,361 | 1,225 |
| <i>Parastichopus californicus</i> | 30,607 | 100bp PE | 10,379 | 1,132 |
| <i>Sphaerechinus granularis</i> | 92,460 | 100bp PE | 24,024 | 2,008 |
| <i>Apostichopus japonicus</i> | 85,061 | 100bp PE | 26,902 | 1,943 |
| <i>Patiria pectinifera</i> | 118,294 | 100bp PE | 33,009 | 1,746 |
| <i>Asterias amurensis</i> | 63,300 | 100bp PE | 19,592 | 2,088 |

Public datasets (red background) were obtained from RefSeq, the SRA and personal communication; new datasets from this study are in blue. Data from RefSeq (green and purple background) were compared with the *de novo* transcriptomes assembled in this study (orange background). Members of Echinodermata are in purple or orange and organisms serving as phylogenetic outgroups are in green.

Supplemental Table 2: Classification of sea stars.

| Taxon | Class | Subclass | Order | Family | Genus |
|-------------------------|--------------|-----------------|---------------|----------------|---------------|
| Luidia clathrata | Astroidea | Valvatacea | Paxillosida | Luidiidae | Luidia |
| Patiria pectinifera | Astroidea | Valvatacea | Valvatida | Asterinidae | Patiria |
| Patiria miniata | Astroidea | Valvatacea | Valvatida | Asterinidae | Patiria |
| Henricia sp | Astroidea | Spinulosacea | Spinulosida | Echinasteridae | Henricia |
| Echinaster spinulosus | Astroidea | Spinulosacea | Spinulosida | Echinasteridae | Echinaster |
| Marthasterias glacialis | Astroidea | Forcipulatacea | Forcipulatida | Asteriidae | Marthasterias |
| Pisaster ochraceus | Astroidea | Forcipulatacea | Forcipulatida | Asteriidae | Pisaster |
| Leptasterias sp | Astroidea | Forcipulatacea | Forcipulatida | Asteriidae | Leptasterias |
| Asterias amurensis | Astroidea | Forcipulatacea | Forcipulatida | Asteriidae | Asterias |
| Asterias rubens | Astroidea | Forcipulatacea | Forcipulatida | Asteriidae | Asterias |
| Asterias forbesi | Astroidea | Forcipulatacea | Forcipulatida | Asteriidae | Asterias |

The taxonomic breakdown of all sea star species in all three phylogenetic analyses. Monophyletic groups are labeled in blue and paraphyletic in purple. In red are relationships that are polyphyletic in all three analyses.

Supplemental Table 3: Morphological trait changes within echinoderms.

| Character Trait Number | Character Trait | State 1 | State 2 | State 3 |
|------------------------|--|---------------------|---|---------------------------------------|
| 1 | Origin of oral somatocoel in nonfeeding larvae | anterior enterocoel | posterior enterocoel | schizocoely |
| 7 | Adult mouth forms from larval left | False | True | |
| 12 | Free-living | False | True | |
| 13 | Meridional ambulacral growth | False | True | |
| 21 | Hemal system with discrete canals | False | True | |
| 22 | Multiple gonads | False | True | |
| 24 | Outer genital coelom surrounds gonad | False | True | |
| 26 | Stone canal calcified | False | True | |
| 27 | Internal hydropore | False | True | |
| 28 | Perianal coelom vis a vis main body coelom | undifferentiated | differentiated | |
| 30 | Expansion of lantern coelom | False | True | |
| 33 | Tube feet with calcified disk | False | True | |
| 37 | Internal skeleton on esophagus | False | True | |
| 38 | Anus in adult | False | True | |
| 39 | Position of anus opposite to peristome | False | True | |
| 40 | Looped and cylindrical gut | False | True | |
| 41 | Additional secretory cells in tube feet in apical tuft cells | False | True | |
| 42 | Sperm morphology in species with external fertilization | spherical | elongate | |
| 43 | Axial gland abutting left axial sinus, but not enclosed | False | True | |
| 44 | Axial complex – stone canal in axial-sinus wall | False | True | |
| 45 | Right axial sinus | absent | restricted to distal end of complex, forming dorsal sac | extends along length of axial complex |
| 48 | Adambulacral ossicles differentiated | False | True | |
| 51 | Batyl alcohol | False | True | |
| 52 | Gonopores | oral | aboral | serial |
| 53 | Mouth plates with specialized jaw | False | True | |
| 54 | Odontophore | False | True | |
| 60 | lrRNA adjacency in mitochondrial gene order | COI-5' | g tRNA-5' | |
| A | Presence of fenestration of larval skeleton | False | True | |
| B | Presence of micromere lineage | False | True | |

Documented character changes within Echinodermata. Adapted from and character numbers as in Janies (2001).

Supplemental Table 4: High copy sequences trimmed from *O. wendtii* 454 assembly

| Sequence name | Sequence trimmed |
|-------------------------|--|
| 454 adapter A1 | TCCATCTCATCCCTGCGTGTCCCATCTGTTCCCTCCCTGTCTCA |
| 454 adapter A2 | TGAGACAGGGAGGGAACAGATGGGACACGCAGGGATGAGATGGA |
| 454 adapter B1 | CCTATCCCCTGTGTGCCTTGCCTATCCCCTGTTGCGTGTCTCA |
| 454 adapter B2 | TGAGACACGCAACAGGGGAAAAGGCAAGGCACACAGGGGATAGG |
| matches found 326555 | ACGAGCGGCCA |
| matches found 57840 | GCCTCCCTCGCGCCATCAGCCGCGCAGGT |
| matches found 9664 | ATCGGTGATTAGTTGTTTTCGGTCATCAAACCTTCCTCACCCAAGGGAATTTGGGG AGCCGCACAGAAACATGGTGGGTGTTACAATCTAAGTGCTGATAGGACTTGCGC TTAGTAGCTGCAGCCTAGATGAAGCGATCTGATTCTGATAATTCACCAGTGCAGC GAAAATTACAGCGCTTTGAACCAGTGGAAAAGGCGCTATATAAATCCAAATTAT TATTATTGGCCGCTCGT |
| matches found 7122 | ACCAAACCAGTACTAATGAAGAAAAAAAA |
| matches found 6854 | GAGAGATAAATGATCTAAATGTGCGTTTATTCAAAAATAGATTTTAAGCTGTGTT GATTGGGATGCTGTTTGT |
| matches found 7591 | TCATCTGATGATGTCAATATATCATACAAGTCATTCATTGAGAAGTTTAATGTAAT GTACGAAGCGTGT |

Sequences trimmed from individual 454 reads during the assembly of the *O. wendtii* transcriptome. Some sequences are technical in nature and should be removed (e.g. adapters), though other sequences were removed simply due to their very high abundance. Highly repeated sequences greatly increase the computational burden of the assembly and were therefore removed.

Chapter III: 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

Tara Fresques*, Vanesa Zazueta-Novoa*, **Adrian Reich**, and Gary M. Wessel

* These authors contributed equally to this work

Developmental Dynamics (2013)

CONTRIBUTION

I assembled the *de novo* transcriptome and constructed a database used throughout the project. A single *P. miniata* female was used to extract RNA from ovaries. The isolated RNA was purified with the RNEasy Mini kit (Qiagen) with on column DNase. Using standard techniques, a library was constructed using the Illumina mRNA-Seq kit. It was then sequenced on a GAIIX with paired end reads with a total read length of 105bp. The reads were assembled using Velvet (1.0.09) and Oases (0.1.14) with a k-mer of 31 (Schulz *et al.*, 2012). From each collection of loci, a single exemplar sequence was selected that was the most abundant and at least 80% the length of the longest member of the locus. The exemplar sequences were annotated with BLAST2GO (Conesa *et al.*, 2005). All of the sequences were then loaded into a custom FileMaker database, and the sequences could be shared with collaborators.

ABSTRACT

Background: Echinodermata is a diverse Phylum, a sister group to chordates, and contains diverse organisms that may be useful to understand varied mechanisms of germ-line specification. **Results:** We tested 23 genes in development of the sea star *Patiria miniata* that fall into five categories: 1) Conserved germ-line factors; 2) Genes involved in the inductive mechanism of germ-line specification; 3) Germ-line associated genes; 4) Molecules involved in left-right asymmetry; and 5) Genes involved in regulation and maintenance of the genome during early embryogenesis. Overall, our results support the contention that the posterior enterocoel is a source of the germ line in the sea star *P. miniata*. **Conclusion:** The germ line in this organism appears to be specified late in embryogenesis, and in a pattern more consistent with inductive interactions amongst cells. This is distinct from the mechanism seen in sea urchins, a close relative of the sea star clade. We propose that *P. miniata* may serve as a valuable model to study inductive mechanisms of germ-cell specification and when compared to germ-line formation in the sea urchin *S. purpuratus* may reveal developmental transitions that occur in the evolution of inherited and inductive mechanisms of germ-line specification.

INTRODUCTION

Evolutionary changes have resulted in a diverse series of mechanisms to accomplish the task of germ-line specification. Three extremes in these mechanisms are widely recognized in the animal kingdom and include 1) germ-cell derivation from adult multipotent stem cells (e.g. neoblasts in planaria, I-cells in hydra), 2) inheritance of maternal factors in early embryogenesis (e.g. pole plasm in *Drosophila melanogaster*, germ plasm in *Xenopus laevis*), and 3) cell-cell communication resulting in induction of a germ-line lineage (e.g. mouse and axolotl; Extavour and Akam, 2003; Solana, 2013). Because these different mechanisms of germ-line specification are polyphyletic, transitions between these germ-line specification mechanisms appear to have occurred multiple times within animal evolution (Extavour and Akam, 2003).

When an embryo exhibits a germ-line determination mechanism that exceeds one biological threshold or another, the investigator usually classifies that mechanism as either inductive or inherited. While this is important for ease and clarity in communication it does not reflect the biological mechanism(s) effectively. Since germ-line determination is likely a result of multiple parallel pathways and activities leading to determination, the transition from one state to another in an embryo may result from a continuum of changes instead of a series of binary switches. Part of the problem with our current definitions of germ-line specification results from the fact that most of what we know comes from a small set of animals; mostly developing by an inherited mechanism. These animals include *D. melanogaster*, *Caenorhabditis elegans*, *X. laevis*, and *Danio rerio*, each with rich contributions of genetic, biochemical, and embryological analysis (e.g. Gao and Arkov, 2013; Lai and King, 2013; Seydoux and Braun, 2006; Voronina, 2013). Unfortunately, we do not have as rich a data set for understanding the inductive mechanisms of germ-line determination. The mouse is the best understood model of inductive germ line determination and in this embryo, the epigenetic changes essential to prime PGC formation occurs as a result of signaling between cells (e.g. Magnusdottir *et al.*, 2012). This epigenetic reconfiguration then diverts the fate of the cells from a somatic direction, and instead leads to a

germ line fate. Whether these features seen in mice are conserved in other organisms using inductive mechanisms is not yet clear and having a strong data set of inductive germ-line features from multiple different animals will be important in order to establish a baseline from which comparisons can be made between inherited mechanisms. Reaching this goal may then reveal the features of germ cell specification shared by inductive mechanisms, those that are unique to inherited and inductive mechanisms, and how evolutionary transitions between each mechanism may occur.

Preliminary experiments in sea stars suggest that this animal uses inductive mechanisms to specify their germ cells. The conserved germ-line factor *Vasa* becomes restricted to the posterior enterocoel (PE; see Fig. 1) only by the early larval stage (Juliano and Wessel, 2009). In addition, PE removal experiments show that the cells from this structure contribute to the future germ cell lineage (Inoue *et al.*, 1992). This is very different than what is known in the closely related taxon of sea urchins. Sea urchins specify their germ cells as early as the 32-cell stage when 4 small micromere cells (sMics) arise as the product of two unequal cell divisions (see Fig. 1). Multiple germ cell marker RNAs and proteins accumulate in the sea urchin sMic lineage (Juliano *et al.*, 2006; Voronina *et al.*, 2008; Wessel *et al.*, 2013), they express these markers cell autonomously (Yajima and Wessel, 2012), and removal of these cells results in loss of the germ lineage in adults (Yajima and Wessel, 2011). In comparison to sea urchins, sea stars and all other groups within echinoderms do not have a sMic lineage. This leads us to believe the sMic lineage and the associated mode of early germ-line specification arose independently in the echinoid (sea urchin) lineage from a common ancestor that used inductive processes for germ-line determination.

We hypothesize that sea stars use an inductive mechanism of germ cell specification that may represent the ancestral mode of germ cell specification in echinoderms (Extavour and Akam, 2003). To begin to test this hypothesis we undertook the current study with the goal of analyzing the expression of germ cell factors in sea stars based on candidate genes that are important for

both inductive and inherited modes of germ cell specification in other organisms. In this way we can compare the expression patterns of germ-line associated markers between sea stars, sea urchins, and other model organisms – especially mice - to test if sea stars specify their germ line in an inductive mode and if so, test if this mode has mechanisms conserved over long periods of evolutionary time.

The genes studied here fall into five categories: 1) Conserved germ-line factors; 2) Genes involved in the inductive mechanism of germ-line specification; 3) Germ-line associated genes; 4) Molecules involved in left-right asymmetry; and 5) Genes involved in regulation and maintenance of the genome during early embryogenesis. Overall, our results support the contention that the PE is a source of the germ line in the sea star *P. miniata* and that there is no specific accumulation of germ cell markers in any cells prior to PE formation. Our results lead us to conclude that the endomesoderm retains the expression of many pluripotency-associated genes which later give rise to the PE. In addition, we found that PE formation and *P. miniata* PGC specification is likely determined by inductive interactions amongst cells which simultaneously cause both the accumulation of germ cell determinants and the loss of somatic cell markers in the presumptive PE. We propose the sea star may serve as a valuable model for future study of the inductive mechanisms of germ-line determination, and when compared to the data sets in sea urchins already available, may serve as a useful comparative model for understanding the developmental transitions between an inductive germ-line determination mechanism and an inherited mechanism.

RESULTS AND DISCUSSION

We selected the genes used in this study by first identifying genes that are associated with PGC specification in a variety of animals that exhibit diverse mechanisms of germ-line determination. Genes involved in both inherited and inductive mechanisms were chosen as well as those involved in left-right asymmetry. This latter group was selected because the hypothesis

being tested is that the PE contributes to the primordial germ cells – the PE structure is on the left side of the midline in the larva. We obtained sequences of *D. melanogaster* and mouse proteins of all the genes tested here from NCBI. Orthologous protein sequences from sea urchins were found by BLAST analysis against the published sea urchin database (Spbase.org). *P. miniata* orthologous protein sequences were found by BLAST analysis against a nascent *P. miniata* ovary transcriptome database. The top *P. miniata* hit was used for reciprocal-BLAST analysis to the non-redundant NCBI database to test orthology. Alignments using these orthologous sequences from *Mus musculus*, *D. melanogaster*, *S. purpuratus* and *P. miniata* were performed to further test authenticity (Tables 1-5 and Supplementary Fig. 1). The lists of primers used for PCR amplification of each gene in sea star and the sizes of predicted and acquired PCR products are shown (Tables 1-5).

Conserved germ-line determinants – Select expression in the Posterior Enterocoel

The genes in the first set we explored are those most highly conserved amongst animals as being part of the germ-line determination mechanism (Table 1). Vasa, Nanos, and Piwi are a classical cluster of germ-line factors, found in all animals at some point in the construction or maintenance of a new germ line in both inductive and inherited germ-line formation mechanisms. Vasa is a dead-box helicase involved in regulating the translation of RNAs in the germ-line. We found in the sea star that Vasa gene expression (mRNA accumulation) is ubiquitous in eggs and early embryos and first becomes restricted to the vegetal pole of the blastula. During gastrulation, Vasa mRNA becomes enriched in the middle region of the archenteron and by early larval stages is restricted to the left side of the mid-archenteron where the PE buds. Vasa remains selectively expressed in the PE throughout the development of the larva (Fig. 2). We also noted less detectable amounts of Vasa transcripts are present in the perimeter of the left and right coelomic pouches in larval stages (Fig. 2; late larva; asterisks). Overall these results support the contention that the PE is at least relevant for consideration of the origin of germ-line determination in this organism. The result also speaks more generally to Vasa function – the transcript is broadly

expressed in the egg and early embryo – clearly it is not strictly a germ-line factor. Previous results in sea stars show that Vasa protein accumulates ubiquitously early in development as well. Not until the PE forms does the Vasa protein become restricted (Juliano and Wessel, 2009), much like the expression of its mRNA. This is a particularly important distinction to be made since in sea urchins the protein does not accumulate coincident with the mRNA. Protein translation is widespread in the sea urchin embryo but degradation is selective to the somatic cells. Therefore, the sMics, the PGC lineage in sea urchins, accumulate the Vasa protein selectively even with broad mRNA presence. This selection (degradation of Vasa protein outside of the presumptive PGCs) appears to be a function of at least one E3-ubiquitin ligase activity, Gustavus (Gustafson *et al.*, 2011).

Piwi is a conserved germ-line factor involved in small RNA-mediated degradation of transposons, and its accumulation is similar to Vasa in *P. miniata* development. That is, broad early expression with selective accumulation in the PE by the early larval stage (Fig. 2). Although the Piwi transcript does not follow as tight an expression domain as Vasa, it is clear that Piwi transcripts accumulate in the archenteron as soon as it is formed and later in gastrulation the Piwi mRNA is enriched in the region where the PE forms - within the mid-region (future mid-gut) of the archenteron (Fig. 2, late gastrula, early larva). By the late larval stage Piwi transcripts disappear from the gut, but are retained in the PE (Fig. 2, larval stages). In the late larval stage we also note Piwi transcripts in the posterior portions of the left and right anterior coelomic pouches.

In the sea urchin, Piwi expression is much like Vasa, that is, broad early expression with protein accumulation selective to the sMics, and only subsequently during gastrulation does the mRNA become restricted to the sMics (Juliano *et al.*, 2006; Rodriguez *et al.*, 2005). In that respect it is similar in the sea star where the vegetal plate and endodermal tissues are the last to down regulate the Piwi mRNA while being retained in the future germ line, either the sMics in sea urchins or the PE in sea stars (Fig. 2, late larva, asterisk). Clearly both Vasa and Piwi gene

expression patterns support the contention that the PE contributes to the germ line, and that their expression is not uniquely germ line.

Nanos is another conserved germ-line factor found in all animals studied. Nanos interacts with Pumilio to bind the 3'UTRs of select mRNAs bearing a Pumilio Response Element (PRE, also referred to as Nanos Response Element, NRE). Binding of this complex to the mRNA reduces translation of the encoded protein either through the translational machinery (Cho *et al.*, 2006; Vardy and Orr-Weaver, 2007b) or by degrading the mRNA by recruiting deadenylase activity (Kadyrova *et al.*, 2007; Vardy and Orr-Weaver, 2007a, b). Pumilio is usually expressed more broadly than Nanos and can interact with a variety of regulators of the bound mRNA. Nanos on the other hand is often restricted to the germ line and its expression is associated with decreasing the cell cycle of the PGC through binding of cyclin mRNAs. Mis-expression of Nanos in non-germ line cells will often lead to developmental abnormalities or death in the somatic cells (Lai and King, 2013; Lai *et al.*, 2012). In sea urchins, Nanos is tightly regulated by a variety of means to be specific to the germ line cells (e.g. Oulhen *et al.*, 2013). In the sea star, Pumilio, but not Nanos, is broadly expressed throughout much of the embryo early in development. Following gastrulation Pumilio became enriched to the gut but was largely excluded from the coelomic pouches and the PE (Fig. 2). Nanos however first appears significantly detectable in the PE when it forms (Fig. 2, late larva, asterisk). In this regard, it is like in sea urchins, where Nanos appears only once the sMics form. This is unlike other germ-line marker transcripts that accumulate in larger expression domains prior to restriction (see Vasa and Piwi, Fig. 2).

Boule/Dazl expression overlaps that of Pumilio but with distinct characters. Boule gene expression generally is present uniformly in early development with slight enrichment in the endomesoderm. In early larvae, the message is enriched in the esophagus significantly over other regions of the embryo or gut. In later larvae, expression is enriched in the esophagus in addition to a small number of cells (Fig. 2, asterisks), which may be the precursors to the dorsal ganglia in late larva (Yankura *et al.*, 2013).

Overall this gene set supports the hypothesis that the PE is a special derivation. In concert, the Piwi, Nanos and Vasa selective mRNA expression, and Vasa protein expression (Juliano and Wessel, 2009) argues that the PE contributes to the germ line. The endomesoderm however, seems to retain many pluripotency-related genes not otherwise present in the ectoderm or future mesodermal cells. Perhaps the endoderm retains broad developmental potency utilizing these various germ cell markers that only later in development become restricted to the PE.

Gene regulatory molecules involved in inductive specification of germ cells in the mouse are conserved in echinoderms

The mechanism of inductive germ-line specification is thought to be the most ancient mechanism used by animals (Extavour and Akam, 2003). Many signaling molecules involved in inductive germ-line specification are not unique to the germ line during animal development and are instead used in a variety of developmental processes. The list of genes involved in the inductive mechanisms of germ-line determination is small – most work thus far has been accomplished only by genetic approaches and only in the mouse. This set of genes is important to test in order to determine if the signaling pathways and downstream effector molecules that are required for inductive germ-line specification in the mouse are conserved in other animals.

Blimp1 (B lymphocyte-induced maturation protein) also known as Prdm1 (Positive regulatory domain I element of the β -IFN gene promoter, containing Zn-fingers) is a transcriptional regulator. Blimp1 is believed to generally repress its target genes, although some evidence indicates that Blimp1 can also serve as an activator of gene expression in certain contexts (John and Garrett-Sinha, 2009; Magnúsdóttir *et al.*, 2013). Blimp1 was first discovered in the immune system, but it is expressed and functional in many tissues during development (Vincent *et al.*, 2005), and in many animals studied (e.g. de Souza *et al.*, 1999). In mice, Blimp1 is a key factor for germ-line determination (Kurimoto *et al.*, 2008; Saitou *et al.*, 2005; Saitou and Yamaji, 2010; Seervai and Wessel, 2013). Blimp1 appears to repress genes within the presumptive germ cells of mice (including Hox genes, esp. Hoxb1) whereas these same genes

remain active in neighboring cells as they begin somatic differentiation. Therefore, Blimp1 in the mouse germ line appears to function in retaining a pluripotent fate by repressing genes important for somatic fates commitments. In contrast, the sea urchin Blimp1 is involved in endoderm gene specification; Blimp1 appears to activate Wnt8 but repress Notch and HesC expression. Knocking down Blimp1 results in a lack of endoderm morphogenesis and perhaps conversion to non-endodermal fates (Livi and Davidson, 2006; Smith and Davidson, 2008; Smith *et al.*, 2008). In the sea star oocyte, Blimp1 transcripts are in low abundance, but in the blastula they accumulate significantly in a torus of cells surrounding the vegetal pole (Fig. 3, asterisks, and Hinman and Davidson, 2003; Hinman *et al.*, 2007). In mid-gastrulae, expression of Blimp1 extends into the posterior archenteron. In late gastrula and early larval stages, expression is restricted to the midgut and hindgut regions of the archenteron, with a slight clearing in the gut region where the PE will form, followed by a marked absence in the PE. Blimp1 is enriched in the stomach, intestine and anus of the late larva, and enrichment persists later in development to the esophagus. Were Blimp1 to have a conserved function in the sea star as in the mouse, it may be in retaining some potentiality of cell fate in the endoderm. Removing its function may, as in mice, enable cells to differentiate into mesodermal lineages and thereby lose their potential for endodermal and germ-line fates.

Prdm14 is closely related to Blimp1/Prdm1 and is a key regulator of mammalian germ cell development (Yamaji *et al.*, 2008). It plays a critical role in cell fate pluripotency by suppressing the expression of differentiation marker genes. In mice, Prdm14 is expressed early in germ line determination, by 6.5dpc, and its up-regulation is likely in response to Bmp4 signaling. Recently it was shown that Prdm14 functions to ensure pluripotency through two pathways: 1) it antagonizes activation of the fibroblast growth factor receptor (Fgfr) signaling by the core pluripotency transcriptional circuitry, and 2) it represses expression of *de novo* DNA methyltransferases that would otherwise modify the epigenome to a primed (somatic) epiblast-like state. Prdm14 exerts these effects by recruiting polycomb repressive complex 2 (Prc2)

specifically to key genomic sites and repressing nearby gene activity (Yamaji *et al.*, 2013). Prdm14 in the sea star is present widespread and has a very similar expression pattern to Bmp2/4 (below) in embryos and larvae. That is, Prdm14 transcripts accumulate in the archenteron of the gastrula and larvae showed strong Prdm14 staining in the mouth opening, stomach, intestine, anus, and the preoral and postoral ciliary bands, but not in the anterior coeloms nor in the PE (Fig. 3, asterisks). In the oocyte, Prdm14 transcripts are ubiquitous (Fig. 3). Unlike in mice, Prdm14 in sea stars or sea urchins does not correlate with the accumulation of other PGC markers. Instead, Prdm14 localizes to ciliary bands in urchins and to the embryonic gut and ciliary bands of sea star larva. This leads us to hypothesize that Blimp1 and Prdm14 may be involved in retaining gut pluripotency early in sea star development, which may be an essential role, but this role does not appear to be sequestered to the germ line.

Bmp2/4 is a critical signaling component for germ-line specification in the early mouse embryo (Saitou *et al.*, 2002; Saitou and Yamaji, 2010). It is expressed by cells in the extra-embryonic tissues and is required for cells in the posterior epiblast to become PGCs. Zhou *et al.* (2010) also found Bmp4 enhances germ cell derivation *in vitro* from ES cells, although currently the link is not clear, if any, between the Bmp4 signaling and Blimp1 expression (Toyooka *et al.*, 2003). The sea star, as in sea urchin, appears to have a less diverse family of Bmp signaling molecules and the closest ortholog to the mouse Bmp4 in the sea star is a Bmp2/4 gene (Lapraz *et al.*, 2006, see Supplementary Fig. 1). Bmp2/4 in sea star is present broadly in development (Fig. 3). Remarkably, the Bmp2/4 transcripts are enriched in the nuclei of young oocytes, but not in nuclei of full-grown oocytes or embryonic cells. Some gene expression is regulated by transcript retention in the egg nucleus of the sea urchin (Angerer and Angerer, 1981; Venezky *et al.*, 1981) though this is the first example of selective transcript retention in the germinal vesicle (GV) of a sea star. In gastrula and early larval stages, Bmp2/4 transcripts accumulate very similarly to Prdm14, and show a clear accumulation in the gut of the embryo, but not in the anterior coeloms nor in the PE. In late larvae, Bmp2/4 is expressed largely around the opening of the mouth, anus

and the preoral and postoral ciliary bands (Fig. 3; asterisks). Recently, another report on Bmp2/4 in *P. miniata* showed enriched Bmp2/4 expression in the future site of mouth formation in blastula stage embryos (Yankura *et al.*, 2013). We did not see this profile here although it may have been a time point missed in our analysis – we focused on PE formation here and terminated the in situ hybridization analysis to best reveal signal information in later embryos and larvae.

The Wnt signaling pathway is thought to regulate several aspects of germ-line function. These include germ-line stem cell functions (e.g. Golestaneh *et al.*, 2009), germ-cell migration (e.g. Chawengsaksophak *et al.*, 2012), and germ cell-soma interactions in the gonad (e.g. Tanwar *et al.*, 2010). Most importantly, and in regards to mouse inductive PGC specification, Wnt3 signaling is required to prime the cells in the posterior epiblast so they are competent to respond to Bmp4 PGC specification signals (Ohinata *et al.*, 2009). In addition, Wnt8 is a somatic marker in axolotls that is expressed during gastrulation in areas of germ cell precursor formation (Bachvarova *et al.*, 2001; Johnson *et al.*, 2003). We chose to determine the expression patterns of Wnt3 and Wnt8 in the sea star to determine if their expression correlates with PE formation. Moreover, in sea urchins as in many embryos, the Wnt signaling pathway is involved in broad scale axial patterning (e.g. Kumburegama and Wikramanayake, 2009; Stamateris *et al.*, 2010). In sea urchins, the nuclearization of β -catenin (a mark of active canonical Wnt signaling) begins in the micromeres at the 16-cell stage, the precursors to the sMics. β -catenin and Blimp1 further trigger Wnt8 expression (Oliveri *et al.*, 2008; Yamazaki *et al.*, 2012). These expression profiles are common in all indirect-developing sea urchins examined so far (Nakata and Minokawa, 2009; Yamazaki *et al.*, 2010). Furthermore, in sea urchins Wnt8 is the primary component of the early endomesoderm gene regulatory network, and its role is to activate the β -catenin/Tcf signal transduction system (Oliveri *et al.*, 2008; Smith and Davidson, 2008). In sea stars, Wnt8 and Wnt3 expression patterns show a segmented, circumferential pattern with an overlapping border in the posterior third of the ectoderm (Fig. 3). As in sea urchins, Wnt8 transcripts are not

maternally expressed in the sea star. By the late blastula stage, however, Wnt8 transcript enrichment is seen as a ring around the vegetal half of the embryo (Fig. 3; asterisks). This expression pattern continued until mid-gastrula, but its enrichment decreased during early and late larval stages. The Wnt3 expression pattern in *P. miniata* follows a similar expression pattern as Wnt8, but the circumferential pattern is closer to the vegetal pole of the embryo (Fig. 3, asterisks). This pattern continued during the gastrula and larval stages and both genes appeared to have continuous expression domains. Neither of the Wnts is detectable in the invaginated endoderm (where the greatest enrichment occurs for Blimp1, Prdm14, and Bmp2/4 transcripts) nor in the PE (see also Yankura *et al.*, 2013). We conclude from this small sampling that at least Wnt 3 and 8 are not directly linked to PE formation, though indirect instruction remains possible especially if they are conferring competency to respond to later specification signals (e.g. Bmp2/4).

Lin28 plays multiple roles in regulating cellular homeostasis at least in part by binding and regulating miRNAs (West *et al.*, 2009). A well-known target of Lin28 negative regulation is let-7, a miRNA critically involved in developmental regulation in *C. elegans* (Bussing *et al.*, 2008). Lin28 expression is linked to pluripotency in a variety of animals. For example, *Lin28* is essential for proper PGC development in mice (West *et al.*, 2009). Let-7 usually represses Blimp1, therefore, the presence of Lin28 in the germ line suppresses let-7 and enables Blimp1 to accumulate and specify germ-line cells. In sea star oocytes, Lin28 transcripts accumulate ubiquitously (Fig. 3) and then become restricted to the vegetal pole of the blastula and then to the archenteron in the early gastrula stage. During late gastrulation, Lin28 transcripts are restricted to the archenteron within two enriched domains. The first one forms a ring in the upper part of the archenteron and the second domain forms a ring closer to the vegetal pole (Fig. 3, asterisks). Lin28 transcripts accumulate in the lower part of the forming esophagus of the early larva, and in the stomach, intestine, anus and PE in the late larva, but not in the coelomic pouches.

Importantly, a conserved let-7 miRNA was found in sea urchins (Kadri *et al.*, 2011; Song *et al.*, 2012), although its site of accumulation is not known, nor whether it is present in sea stars.

Assuming Lin28 functions similarly in sea stars as it does in other organisms the marked enrichment in the endoderm (along with Blimp1/Prdm1 and Prdm14) supports the hypothesis that it may be involved in retaining pluripotency in this tissue early in sea star development. Since Lin28 transcripts are not present in the PE, either they do not function in germ-line determination, or they have already accomplished their germ-line role (in the endoderm) prior to this time. We hypothesize that the endoderm (invaginated epithelium) may have significant pluripotentiality that is lost in other tissues. We do not see any direct evidence that Bmp2/4, Wnt3, or Wnt8 expression patterns correlate with PE formation. However, it is still possible that these signaling pathways may act on precursor PE cells at an earlier stage to confer the competency to respond to later PGC inducing signals. It is also possible that Bmp signals may act at a distance during PGC specification in this embryo as well.

Germ-line associated genes.

We examined the expression patterns of germ-line associated genes to test correlations with PE formation in sea stars. This cohort of genes is associated with sex determination and, potentially, regulation of germ-line factors (Tables 3-5).

CNOT6 is important for PGC development in sea urchins (Appendix IV, Swartz *et al.*, *submitted*). CNOT6 has deadenylase activity and is recruited to mRNAs for widespread mRNA degradation. CNOT6 is down regulated in the PGCs, and thus these cells retain mRNAs for prolonged times. This is especially important considering the reduced transcriptional activity of the PGCs when compared to their neighboring somatic cells. Indeed, in sea urchins, the sMics retain their maternal (as well as microinjected, exogenously generated) mRNAs for many days whereas their neighboring cells turnover the same transcripts within about one day (Gustafson and Wessel, 2010; Oulhen and Wessel, 2013). CNOT6 is present ubiquitously in the sea urchin embryo except for a marked depletion in the sMics, and this selective depletion of CNOT6

mRNA appears to result from nanos expression selectively in the sMics (Appendix IV, Swartz *et al.*, *submitted*). The CNOT6 mRNA has two nanos/pumilio response elements in its 3'UTR that causes degradation of CNOT6 selectively in the sMics in a nanos-dependent fashion. The CNOT6 present in all other cells of the embryo contributes to the egg – embryo transition e.g. a clearing of the general, maternal, and pluripotent egg mRNAs and “freeing” the somatic cells to differentiate. In contrast, the depletion of CNOT6 in the sMics contributes to the retention of these same mRNAs and their pluripotency (Appendix IV, Swartz *et al.*, *submitted*). It is not clear yet if the sea star has cells that effectively retain maternal mRNAs, so here we determined the expression of CNOT6 to test if such a mechanism may exist. CNOT6 mRNA is present uniformly in the sea star oocytes. By the blastula stage, we detected a depletion of CNOT6 transcripts around the blastopore (Fig. 4, arrows). CNOT6 transcripts are enriched in the ectoderm and archenteron of mid-gastrula stage embryos but a clear region remains within the vegetal pole (Fig. 4, arrow). In late gastrulae and early larval stages, CNOT6 is still expressed in the ectoderm and in the gut but is depleted in the PE and in the coelomic pouches. In the late larval stage, transcripts are enriched in the gut and both the preoral and postoral ciliary bands (Fig. 4, asterisks). We predict that general mRNA retention is greater in the endodermal cells through early larval stages that, once again, emphasizes the endoderm as a tissue retaining its pluripotency.

Gustavus is an E3 ubiquitin ligase that binds Vasa and leads to its degradation. Originally found in *Drosophila* (Styhler *et al.*, 2002), it is now also known to bind and regulate Vasa protein accumulation in the sea urchin. This function is particularly important in sea urchins since Vasa protein is translated in all cells from ubiquitous maternal mRNA but becomes uniquely retained in sMics as a result of Gustavus-mediated Vasa protein turnover in all somatic cells. Since Vasa protein first appears to be ubiquitous in the sea star embryo, and then clears to become enriched in the PE in late larval stages (Juliano and Wessel, 2009), we tested the location of Gustavus mRNA in the sea star. In sea star oocytes, Gustavus transcripts are present ubiquitously (Fig. 4).

Gustavus transcripts accumulate in the vegetal pole of the blastula and around the blastopore in early gastrulae (Fig. 4). In the late gastrula, Gustavus transcripts are enriched in two rings; one ring forms at the base of the forming anus (blastopore) and another ring forms at the foregut, leaving a less-dense region in the midgut, the site where the PE will form (Fig. 4). During early larval stages, Gustavus is expressed in the gut, but not in the coelomic pouches nor in the PE, the sites of greatest Vasa protein accumulation (Juliano and Wessel, 2009). In late larvae, transcripts remain in the stomach, intestine, anus and ciliary bands (Fig. 4). We hypothesize that in the sea star larval stages, Gustavus activity in the midgut may degrade Vasa protein while the lack of Gustavus activity in the PE and anterior coelomic pouches allows Vasa protein to accumulate selectively in these structures.

Echinoderm SoxE is a transcription factor of the HMG family and is the ortholog of the vertebrate member Sox 9 (Howard-Ashby *et al.*, 2006). Sox9 is involved in sex determination in all vertebrates examined; it is regulated positively by SRY in (male) mammals and is enriched in the somatic cells of the presumptive male gonad due to a positive autoregulatory feedback loop (Gilbert *et al.*, 2010; Kashimada and Koopman, 2010). In other vertebrates, including fish, reptiles, and amphibians, Sox9 may be activated by differing mechanisms, including temperature and other environmental factors and is important for both male and female gene expression leading to sexual dimorphism (Kuroiwa *et al.*, 2002; Mawaribuchi *et al.*, 2012; Muramatsu *et al.*, 2007; Seervai and Wessel, 2013; Spotila *et al.*, 1998; Uno *et al.*, 2008). In sea urchins SoxE is expressed in the left coelomic pouch approximately 50 percent of the time while it has a broader distribution the remaining time (Duboc *et al.*, 2005; Juliano *et al.*, 2006). The SoxE transcript in sea stars is present at low levels in oocytes and is difficult to detect at blastula and mid-gastrula stages. A strong enrichment is then seen in the left anterior coelom and a mild enrichment appears at the midgut in late gastrulae (Fig. 4). During early larval development, SoxE transcripts are retained in the most posterior tip of the left anterior coelom (Fig. 4, asterisk), then in the same site of both anterior pouches, and in the PE. The expression of SoxE in sea stars is similar to sea

urchins in larval stages since it accumulates similarly in the left coelomic pouch and shows a bimodal expression pattern (Juliano *et al.*, 2006).

Ovo is a Zn-finger transcription factor that is important for oogenesis in *Drosophila* and spermatogenesis in mice (Dai *et al.*, 1998; Oliver *et al.*, 1987). In sea urchins, Ovo is expressed ubiquitously in the blastula stage and is enriched in the vegetal pole at the mesenchyme blastula stage, an expression domain similar to Vasa. However, unlike Vasa, Ovo expression is not detected in gastrula and larval stages (Juliano *et al.*, 2006). In the sea star, Ovo transcripts are not detectable in the oocyte but are detectable ubiquitously at low levels in the blastula stage (Fig. 4). During gastrulation Ovo transcripts accumulate throughout the developing gut. In early larvae Ovo transcripts are enriched throughout the gut and esophagus, and in late larva Ovo transcripts are enriched in the gut and pre-oral and post-oral ciliary bands. Although we did not find any obvious correlation between the expression of Ovo and the formation of the PE, its consistent link to the invaginating epithelium may indicate its involvement in the retention of developmental potentiality leading to PE formation.

Changes in cadherin expression are intimately linked to morphogenetic processes that involve the loss of epithelial character and the delamination of cells from an epithelial sheet (Birchmeier *et al.*, 1993; Gumbiner, 1996; Takeichi, 1988), a character often seen in PGCs (McLaren, 2003). In sea urchin embryos, the cellular movements associated with the ingression of primary mesenchyme cells (PMCs) and convergent-extension of the archenteron involve the dynamic regulation of intercellular adhesion, including the loss of cell adhesion molecules (Fink and McClay, 1985; Miller and McClay, 1997). The sMics of the sea urchin appear to retain their cadherin through development until they reach the tip of the archenteron during gastrulation. Premature depletion of the G-cadherin protein appears to disrupt sMic patterning and expression of several cell-specific markers (Yajima and Wessel, 2012) and cadherin orthologs appear important for germ cell function in many animals (Blaser *et al.*, 2005; Chihara and Nance, 2012). Therefore we were interested to test the profile of the G-cadherin orthology in this sea star.

The sequence of the sea star ortholog of G-cadherin shows significant sequence identity to all classic cadherins, particularly in the cytoplasmic domain, the region predicted to be involved in catenin binding (Supplemental Fig. 1). G-cadherin mRNA accumulates ubiquitously in oocytes, but is not detectable in blastula and early gastrula stages (Fig. 4). In the late gastrula, G-cadherin transcripts are enriched in two domains; one in the foregut and another in the hindgut (Fig. 4, asterisks). This creates a midgut region with less G-cadherin transcripts, the same location in which new morphogenetic movements lead to PE formation. This profile also overlaps the Gustavus mRNA profile and perhaps this region supports the evagination of the epithelium leading to PE formation by decreased cell-cell adhesion. G-cadherin transcripts then decreased overall in the larval stages and became restricted to the mouth opening, stomach, intestine, and anus.

Overall, our analysis of the expression of germ-line associated genes in sea star embryos reveals how mechanisms of PGC biology vary between animals that use different modes of PGC specification. Yet, overall similarities are seen; the RNA degradation machinery such as the deadenylase, CNOT6, may be conserved in its down-regulation in PGCs once they are formed to preserve their inherited transcripts and retain greater developmental plasticity. The ubiquitin pathway and specific E3 ligases (such as Gustavus) may be conservatively involved in down-regulating germ cell determinant protein levels outside of the germ cells once they are specified. Transcription factors that are conserved in somatic sex determination seem to have bi-modal expression patterns and be involved in sex determination regardless of the mode of PGC specification. Finally, changes in adhesion, such as through G-cadherin seem to be linked to germ cell function in many animals.

Left/Right asymmetry molecules

The PE forms only on the left side of the *P. miniata* gut during development and we hypothesize that conserved left/right signaling pathways may be involved in PE formation (and therefore PGC specification). Consequently, we identified and tested the gene expression patterns

of signaling molecules involved in left/right asymmetry. Major mechanisms for establishment of the left/right axis in sea urchin and other organisms include complex epigenetic and genetic cascades (Lin and Xu, 2009; Table 4). The initial symmetry-breaking event is not clearly understood, and is probably different between species (Vandenberg and Levin, 2013), yet this axial organization still requires a specific, and likely conserved, set of gene activities. For example, the Tgf-beta family member Nodal is expressed on the left side of chordate embryos and is widely used to specify structures on the left side differently than on the right. Nodal function activates additional genes involved in development of the left/right asymmetry, including the additional Tgf-beta factor Lefty, and the homeobox transcription factor Pitx2 (Levin *et al.*, 1995; Speder *et al.*, 2007). In the sea urchin however, this pathway is reversed and is instead expressed on the right side (Duboc *et al.*, 2005). Overexpression of Nodal throughout the sea urchin embryo results in developmental repression of the adult rudiment on the left side, and removal of Nodal in the sea urchin results in duplicated rudiments on both the right and left sides suggesting that a main function of Nodal is repression of developmental derivatives of the right coelomic pouch (Bessodes *et al.*, 2012; Duboc *et al.*, 2005; Luo and Su, 2012; Warner *et al.*, 2012). In the sea star, Nodal is not expressed in oocytes, but accumulates strongly in early blastula stages in the ectoderm (Fig. 5). Nodal mRNA remains transiently in the ectoderm until the late gastrula stage and is absent there in larval stages. A second domain of Nodal expression, however, occurs in mid gastrula embryos when Nodal message accumulates in the midgut and on the right side of the invaginated epithelium. We also see Nodal mRNA transiently within the posterior region of the right coelomic pouch, but to a much lesser extent.

Lefty follows a similar expression pattern as Nodal in the ectoderm. However, Lefty does not accumulate in the archenteron until later in gastrulation when it accumulates in the right side of the archenteron and in the right coelomic pouch (Fig. 5, late gastrula, asterisk). Pitx2 transcripts, although ubiquitous in *P. miniata* oocytes, do not accumulate significantly in early development until the late gastrula stage. Pitx2 transcripts follow a similar expression domains as

Nodal and Lefty, albeit much delayed (Fig. 5, late gastrula, asterisk) and this is similar in the closely related sea star *Asterina pectinifera* (Hibino *et al.*, 2006). Pitx2 is expressed in the posterior portion of the right coelomic pouch which persists from the late gastrula stage until late larval stages, although its accumulation in the right pouch of *A. pectinifera* is much broader than in *P. miniata*. We noticed that in the sea star *P. miniata*, the initial break in left/right asymmetry within the archenteron (as seen with the right-sided expression of Nodal) occurs at the same time that Vasa expression is restricted to the left side of the archenteron.

Therefore, we hypothesize that left/right signaling and Nodal are required for the initial break in left/right asymmetry of Vasa expression and therefore of PE specification. Overall the sea star left/right asymmetry program appears to closely mirror the program in the sea urchin, even with the significant expression in the right side of the archenteron. Thus, the reversal of the left/right program in echinoderms likely occurred prior to the sea urchin-sea star split; phylogenetic analyses suggests Nodal on the right may be ancestral and that reversal may have occurred in the chordate lineage (Bessodes *et al.*, 2012; Duboc *et al.*, 2005; Luo and Su, 2012; Warner *et al.*, 2012). This also may be related to the reversal of the dorsal/ventral program in chordates relative to non-chordates (Grande and Patel, 2009).

Genomic maintenance during morphogenesis and early embryogenesis

The last group of genes we studied encodes molecules related to genomic and epigenomic regulation during morphogenesis to test if these genes are associated selectively with sea star PE cells (Table 5, Fig. 6). Baf250 is an E3 ubiquitin ligase that functions in selective histone (H2B) turnover. Since Baf250 associates with the mammalian SWI/SNF complex it is thought that this histone-turnover machinery regulates epigenetic modifications that lead to selective gene activity (Li *et al.*, 2010). Such modifications are of particular interest in mammalian embryos with the identification of selective epigenetic reprogramming during germ cell formation (Magnusdottir *et al.*, 2012). In the sea star *P. miniata*, Baf250 transcripts accumulate ubiquitously in oocytes. In blastula stage embryos, Baf250 transcripts become

enriched in the blastopore (Fig. 6, asterisk), and in gastrula embryos transcripts accumulate in the archenteron. During larval stages, Baf250 accumulates in the esophagus, stomach, intestine, anus, and in the PE but not in the coelomic pouches.

Recent results in zebrafish suggest that the DNA repair elements Brca1 and Brca2 are involved in germ line development (Shive *et al.*, 2010). We find that Brca1 and 2 in the sea star have overlapping expression profiles: Brca1 transcripts are expressed ubiquitously throughout early development until the late gastrula stage when they become enriched in the midgut (Fig. 6). Brca2 transcripts are also present in oocytes and they accumulate in the gut of late gastrula stage embryos (Fig. 6, asterisk). The accumulation of both Brca transcripts in oocytes suggests there might be a role for the maternal message in provisioning early blastomeres with Brca proteins that could affect DNA repair during the rapid cleavage divisions that occur.

Traffic jam is an atypical basic leucine zipper transcription factor that regulates somatic-germ cell interactions and its loss results in male and female infertility in *Drosophila* (Li *et al.*, 2003). Here we found that transcripts of the ortholog of Traffic Jam, called Maf, are ubiquitously distributed in oocytes but are not enriched in blastula and mid-gastrula stage embryos. Maf transcripts are enriched slightly in the ciliary bands of the larva (Fig. 6, asterisks but not in the PE).

CONCLUSIONS

Our results support the contention that the PE is a source of germ line cells in the sea star *P. miniata*. Several gene expression patterns reflect a broad initial expression of germ line markers in the embryo followed by a restriction to the PE. These include Vasa, Nanos, and Piwi (see Fig. 7). While these results on their own do not prove that the PE contains the germ line, they are complementary to other studies that suggest that these cells give rise to sea star germ cells. The PE is most likely the site of germ cell formation in this animal based on four criteria: 1) PE removal experiments result in larvae with significantly less germ cells (Inoue *et al.*, 1992), 2)

Vasa immunolabeling experiments show the PE is the first restricted site of Vasa protein localization (Juliano and Wessel, 2009), 3) conserved germ line factors accumulate selectively in the PE *e.g.* *nanos* and *piwi* in addition to *vasa*, and 4) the PE exhibits rapid depletion of mRNAs encoding factors involved in somatic cell fates *e.g.* *Blimp1*. These observations and the fact that the selective expression in the PE is relatively late in development – following gastrulation - leads us to reason that the germ line in this organism is determined by inductive interactions amongst cells.

Blimp1 is a somatic factor in echinoderms (as it is necessary for endomesoderm gene regulatory networks, *e.g.* *Spbase.org*) and the selective loss of somatic markers in the PGCs represents a broad theme in the inductive mode of germ cell specification (see also Fig. 7). In the mouse inductive mode of germ cell specification, markers of mesoderm, such as *Hoxb1* and *Hoxa1*, are lost from PGCs as soon as they receive PGC inducing signals (Saitou *et al.*, 2002). The loss of *Blimp1* transcripts in the PE of sea stars contributes to the hypothesis that when an animal uses the inductive mode for germ cell specification it is conservatively induced from a pluripotential mesodermal lineage.

A consensus revealed in this study was that the gut appears to harbor a gene set making it retain pluripotency and germ line potential. Many factors, including *Vasa*, *Piwi*, *Blimp1*, and *Prdm14*, are enriched in the endomesoderm of the gut and suggest the gut retains developmental potential whereas the ectoderm is devoid of many of these same factors. This result may simply mean that the endomesoderm forms later in its differentiation program when compared to the ectoderm, or that the cells within this tissue give rise to many more cell types later in development. The additional possibility is that the endomesoderm is the site of germ line formation and retains developmental potency. The gene expression profile on its own is not convincing and will certainly require functional analysis with metrics of germ line success to understand functionality of the genes tested here. However, when the precursor cells to the PGCs are disrupted in sea urchins *Vasa* is up-regulated throughout the endoderm of the remaining

embryo (Voronina *et al.*, 2008). In this experimental case within the sea urchin, Vasa mimics the same broad endodermal expression profile that we see with many pluripotency-related genes in sea stars. Thus, these two echinodermal representatives may have diverged in their germ line determination by transposing the germ line program earlier in sea urchins, and to the sMics. Sea urchins represent a unique model in that this animal retains the ability to use multiple mechanisms for PGC specification upon its disruption (compensatory Vasa upregulation and recovery of germ line cells; Voronina *et al.*, 2008; Yajima and Wessel, 2011). We favor the conclusion that the mechanism in sea urchins is derived and perhaps tending towards an inherited mechanism of germ line determination, especially when compared to the sea star.

Two extreme mechanisms of germ-line determination appear in animal development - inherited vs inductive (Ewen-Campen *et al.*, 2010; Extavour and Akam, 2003; Juliano and Wessel, 2010; Seervai and Wessel, 2013). Embryos using inherited mechanisms usually establish their germ line early in development by acquisition of a specific region of egg cytoplasm - maternally deposited in the oocyte. This is the best known mechanism and is used by many model organisms e.g. fly, worm, frog, and zebrafish. Mice are the best studied organism that uses inductive mechanisms of germ-line determination. In this mechanism, cell interactions are responsible, usually later in development, to establish a germ line lineage. Results presented here support the contention that the sea star PE is a site of germ line formation and the evidence suggests this structure may fit better with the criteria of an inductive mechanism. This means functional studies that determine signaling networks required for PGC specification in this organism may complement the genetic and tissue culture approaches used in mice to reveal inductive germ-line determination mechanisms. Furthermore, comparisons between the differing mechanisms of PGC specification between sea star and sea urchins will be useful to understand transitions that occur in the evolution from one mode to another.

MATERIALS AND METHODS

Animals and embryo culture

Patiria miniata were collected from several sites in southern California [www.scbiomarine.com; phalmay@earthlink.net] and embryos were grown basically as described (Foltz *et al.*, 2004). Briefly, sperm were collected from a gonad biopsy and placed into a microfuge tube on ice. Oocytes were collected from a gonad biopsy and matured in vitro with 2 μ M 1-Methyl-Adenine. Resultant eggs were fertilized with a dilute sperm suspension and embryos were cultured as previously described (Hinman *et al.*, 2003). Samples from different developmental stages (oocytes; hatched blastula, 18.5 hours post-fertilization (hpf); mid-gastrula, 27.5 hpf; late gastrula, 47 hpf; early larva, 3 days post-fertilization (dpf); late larva, 4-7 dpf) were collected, fixed and stored in 70% ethanol at -20 as described (Arenas-Mena *et al.*, 2000).

RNA analysis

Whole mount *in situ* RNA hybridizations were performed using digoxigenin-labeled RNA probes as previously described (Arenas-Mena *et al.*, 2000). cDNAs from oocytes and early development stages were used as templates for PCR reactions. Primers designed to amplify each gene of interest included a T7 RNA polymerase sequence in the 5' end of reverse primers. The resultant PCR products were used as templates for transcription by T7 RNA polymerase to yield an antisense RNA probe with the DIG RNA Labeling Kit (SP6/T7) (Roche Applied Science, IN, USA). Oocytes and embryos were fixed, hybridized, and the signals were detected essentially as described (Arenas-Mena *et al.*, 2000). Negative controls for these experiments included the use of a non-relevant transcript probe (Neomycin). Oocytes and embryos were visualized on a Zeiss Axioplan microscope, and the specimens are oriented with their left side to the left i.e. we position the larva ventral side down, left side of the larva to the left of the image. This is the opposite of the classic human orientation scheme but we think it makes the structures easier to interpret.

REFERENCES

- Angerer, L.M., and Angerer, R.C. (1981). Detection of poly A+ RNA in sea urchin eggs and embryos by quantitative in situ hybridization. *Nucleic Acids Res* 9, 2819-2840.
- Arenas-Mena, C., Cameron, A.R., and Davidson, E.H. (2000). Spatial expression of Hox cluster genes in the ontogeny of a sea urchin. *Development* 127, 4631-4643.
- Bachvarova, R.F., Masi, T., Hall, L., and Johnson, A.D. (2001). Expression of Axwnt-8 and Axsxl in the urodele, axolotl: comparison with Xenopus. *Dev Genes Evol* 211, 501-505.
- Bessodes, N., Haillet, E., Duboc, V., Rottinger, E., Lahaye, F., and Lepage, T. (2012). Reciprocal signaling between the ectoderm and a mesendodermal left-right organizer directs left-right determination in the sea urchin embryo. *PLoS Genet* 8.
- Birchmeier, W., Weidner, K.M., Hulsken, J., and Behrens, J. (1993). Molecular mechanisms leading to cell junction (cadherin) deficiency in invasive carcinomas. *Semin Cancer Biol* 4, 231-239.
- Blaser, H., Eisenbeiss, S., Neumann, M., Reichman-Fried, M., Thisse, B., Thisse, C., and Raz, E. (2005). Transition from non-motile behaviour to directed migration during early PGC development in zebrafish. *J Cell Sci* 118, 4027-4038.
- Bussing, I., Slack, F.J., and Grosshans, H. (2008). let-7 microRNAs in development, stem cells and cancer. *Trends Mol Med* 14, 400-409.
- Chawengsaksophak, K., Svingen, T., Ng, E.T., Epp, T., Spiller, C.M., Clark, C., Cooper, H., and Koopman, P. (2012). Loss of Wnt5a disrupts primordial germ cell migration and male sexual development in mice. *Biol Reprod* 86, 1-12.
- Chihara, D., and Nance, J. (2012). An E-cadherin-mediated hitchhiking mechanism for *C. elegans* germ cell internalization during gastrulation. *Development* 139, 2547-2556.
- Cho, P.F., Gamberi, C., Cho-Park, Y.A., Cho-Park, I.B., Lasko, P., and Sonenberg, N. (2006). Cap-dependent translational inhibition establishes two opposing morphogen gradients in *Drosophila* embryos. *Curr Biol* 16, 2035-2041.
- Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674-3676.
- Dai, X., Schonbaum, C., Degenstein, L., Bai, W., Mahowald, A., and Fuchs, E. (1998). The ovo gene required for cuticle formation and oogenesis in flies is involved in hair formation and spermatogenesis in mice. *Genes Dev* 12, 3452-3463.
- de Souza, F.S., Gawantka, V., Gomez, A.P., Delius, H., Ang, S.L., and Niehrs, C. (1999). The zinc finger gene Xblimp1 controls anterior endomesodermal cell fate in Spemann's organizer. *EMBO J* 18, 6062-6072.
- Duboc, V., Rottinger, E., Lapraz, F., Besnardeau, L., and Lepage, T. (2005). Left-right asymmetry in the sea urchin embryo is regulated by nodal signaling on the right side. *Dev Cell* 9, 147-158.
- Ewen-Campen, B., Schwager, E.E., and Extavour, C.G.M. (2010). The molecular machinery of germ line specification. *Mol Reprod Dev* 77, 3-18.
- Extavour, C.G., and Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* 130, 5869-5884.
- Fink, R.D., and McClay, D.R. (1985). Three cell recognition changes accompany the ingression of sea urchin primary mesenchyme cells. *Dev Biol* 107, 66-74.
- Foltz, K.R., Adams, N.L., and Runft, L.L. (2004). Echinoderm eggs and embryos: procurement and culture. *Methods Cell Biol* 74, 39-74.
- Gao, M., and Arkov, A.L. (2013). Next generation organelles: structure and role of germ granules in the germline. *Mol Reprod Dev* 80, 610-623.

- Gilbert, S.F., McDonald, E., Boyle, N., Buttino, N., Gyi, L., Mai, M., Prakash, N., and Robinson, J. (2010). Symbiosis as a source of selectable epigenetic variation: taking the heat for the big guy. *Philos Trans R Soc Lond B Biol Sci* 365, 671-678.
- Golestaneh, N., Beauchamp, E., Fallen, S., Kokkinaki, M., Uren, A., and Dym, M. (2009). Wnt signaling promotes proliferation and stemness regulation of spermatogonial stem/progenitor cells. *Reproduction* 138, 151-162.
- Grande, C., and Patel, N.H. (2009). Nodal signalling is involved in left-right asymmetry in snails. *Nature* 457, 1007-1011.
- Gumbiner, B.M. (1996). Cell adhesion: the molecular basis of tissue architecture and morphogenesis. *Cell* 84, 345-357.
- Gustafson, E.A., and Wessel, G.M. (2010). Exogenous RNA is selectively retained in the small micromeres during sea urchin embryogenesis. *Mol Reprod Dev* 77, 836-836.
- Gustafson, E.A., Yajima, M., Juliano, C.E., and Wessel, G.M. (2011). Post-translational regulation by gustavus contributes to selective Vasa protein accumulation in multipotent cells during embryogenesis. *Dev Biol* 349, 440-450.
- Hibino, T., Nishino, A., and Amemiya, S. (2006). Phylogenetic correspondence of the body axes in bilaterians is revealed by the right-sided expression of Pitx genes in echinoderm larvae. *Dev Growth Differ* 48, 587-595.
- Hinman, V.F., and Davidson, E.H. (2003). Expression of AmKrox, a starfish ortholog of a sea urchin transcription factor essential for endomesodermal specification. *Gene Expr Patterns* 3, 423-426.
- Hinman, V.F., Nguyen, A., and Davidson, E.H. (2007). Caught in the evolutionary act: precise cis-regulatory basis of difference in the organization of gene networks of sea stars and sea urchins. *Dev Biol* 312, 584-595.
- Hinman, V.F., Nguyen, A.T., and Davidson, E.H. (2003). Expression and function of a starfish Otx ortholog, AmOtx: a conserved role for Otx proteins in endoderm development that predates divergence of the eleutherozoa. *Mech Dev* 120, 1165-1176.
- Howard-Ashby, M., Materna, S.C., Brown, C.T., Chen, L., Cameron, R.A., and Davidson, E.H. (2006). Identification and characterization of homeobox transcription factor genes in *Strongylocentrotus purpuratus*, and their expression in embryonic development. *Dev Biol* 300, 74-89.
- Inoue, C., Kiyomoto, M., and Shirai, H. (1992). Germ cell differentiation in starfish: the posterior enterocoel as the origin of germ cells in *Asterina pectinifera*. *Dev Growth Differ* 34, 413-418.
- John, S.A., and Garrett-Sinha, L.A. (2009). Blimp1: a conserved transcriptional repressor critical for differentiation of many tissues. *Exp Cell Res* 315, 1077-1084.
- Johnson, A.D., Crother, B., White, M.E., Patient, R., Bachvarova, R.F., Drum, M., and 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., and Wessel, G. (2010). Developmental biology. Versatile germline genes. *Science* 329, 640-641.
- Juliano, C.E., Voronina, E., Stack, C., Aldrich, M., Cameron, A.R., and 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., and Wessel, G.M. (2009). An evolutionary transition of Vasa regulation in echinoderms. *Evol Dev* 11, 560-573.
- Kadri, S., Hinman, V.F., and Benos, P.V. (2011). RNA deep sequencing reveals differential microRNA expression during development of sea urchin and sea star. *PLoS One* 6.
- Kadyrova, L.Y., Habara, Y., Lee, T.H., and Wharton, R.P. (2007). Translational control of maternal Cyclin B mRNA by Nanos in the *Drosophila* germline. *Development* 134, 1519-1527.

- Kashimada, K., and Koopman, P. (2010). Sry: the master switch in mammalian sex determination. *Development* *137*, 3921-3930.
- Kumburegama, S., and Wikramanayake, A.H. (2009). Wnt signaling in the early sea urchin embryo. In *Wnt Signaling* (Springer), pp. 187-199.
- Kurimoto, K., Yamaji, M., Seki, Y., and Saitou, M. (2008). Specification of the germ cell lineage in mice: a process orchestrated by the PR-domain proteins, Blimp1 and Prdm14. *Cell Cycle* *7*, 3514-3518.
- Kuroiwa, A., Uchikawa, M., Kamachi, Y., Kondoh, H., Nishida-Umehara, C., Masabanda, J., Griffin, D.K., and Matsuda, Y. (2002). Chromosome assignment of eight SOX family genes in chicken. *Cytogenet Genome Res* *98*, 189-193.
- Lai, F., and King, M.L. (2013). Repressive translational control in germ cells. *Mol Reprod Dev* *80*, 665-676.
- Lai, F., Singh, A., and King, M.L. (2012). Xenopus Nanos1 is required to prevent endoderm gene expression and apoptosis in primordial germ cells. *Development* *139*, 1476-1486.
- Lapraz, F., Rottinger, E., Duboc, V., Range, R., Duloquin, L., Walton, K., Wu, S.-Y., Bradham, C., Loza, M.A., Hibino, T., *et al.* (2006). RTK and TGF-beta signaling pathways genes in the sea urchin genome. *Dev Biol* *300*, 132-152.
- Levin, M., Johnson, R.L., Stern, C.D., Kuehn, M., and Tabin, C. (1995). A molecular pathway determining left-right asymmetry in chick embryogenesis. *Cell* *82*, 803-814.
- Li, M.A., Alls, J.D., Avancini, R.M., Koo, K., and Godt, D. (2003). The large Maf factor Traffic Jam controls gonad morphogenesis in Drosophila. *Nat Cell Biol* *5*, 994-991000.
- Li, X.S., Trojer, P., Matsumura, T., Treisman, J.E., and Tanese, N. (2010). Mammalian SWI/SNF--a subunit BAF250/ARID1 is an E3 ubiquitin ligase that targets histone H2B. *Mol Cell Biol* *30*, 1673-1688.
- Lin, X., and Xu, X. (2009). Distinct functions of Wnt/beta-catenin signaling in KV development and cardiac asymmetry. *Development* *136*, 207-217.
- Livi, C.B., and Davidson, E.H. (2006). Expression and function of blimp1/krox, an alternatively transcribed regulatory gene of the sea urchin endomesoderm network. *Dev Biol* *293*, 513-525.
- Luo, Y.-J., and Su, Y.-H. (2012). Opposing nodal and BMP signals regulate left-right asymmetry in the sea urchin larva. *PLoS Biol* *10*.
- Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B., and Surani, M.A. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nature Cell Biology* *15*, 905-915.
- Magnúsdóttir, E., Gillich, A., Grabole, N., and Surani, M.A. (2012). Combinatorial control of cell fate and reprogramming in the mammalian germline. *Curr Opin Genet Dev* *22*, 466-474.
- Mawaribuchi, S., Yoshimoto, S., Ohashi, S., Takamatsu, N., and Ito, M. (2012). Molecular evolution of vertebrate sex-determining genes. *Chromosome Res* *20*, 139-151.
- McLaren, A. (2003). Primordial germ cells in the mouse. *Dev Biol* *262*, 1-15.
- Miller, J.R., and McClay, D.R. (1997). Characterization of the role of cadherin in regulating cell adhesion during sea urchin development. *Dev Biol* *192*, 323-339.
- Muramatsu, S., Wakabayashi, M., Ohno, T., Amano, K., Ooishi, R., Sugahara, T., Shiojiri, S., Tashiro, K., Suzuki, Y., Nishimura, R., *et al.* (2007). Functional gene screening system identified TRPV4 as a regulator of chondrogenic differentiation. *J Biol Chem* *282*, 32158-32167.
- Nakata, H., and Minokawa, T. (2009). Expression patterns of wnt8 orthologs in two sand dollar species with different developmental modes. *Gene Expr Patterns* *9*, 152-157.
- Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. (2009). A signaling principle for the specification of the germ cell lineage in mice. *Cell* *137*, 571-584.
- Oliver, B., Perrimon, N., and Mahowald, A.P. (1987). The ovo locus is required for sex-specific germ line maintenance in Drosophila. *Genes Dev* *1*, 913-923.
- Oliveri, P., Tu, Q., and Davidson, E.H. (2008). Global regulatory logic for specification of an embryonic cell lineage. *Proc Natl Acad Sci U S A* *105*, 5955-5962.

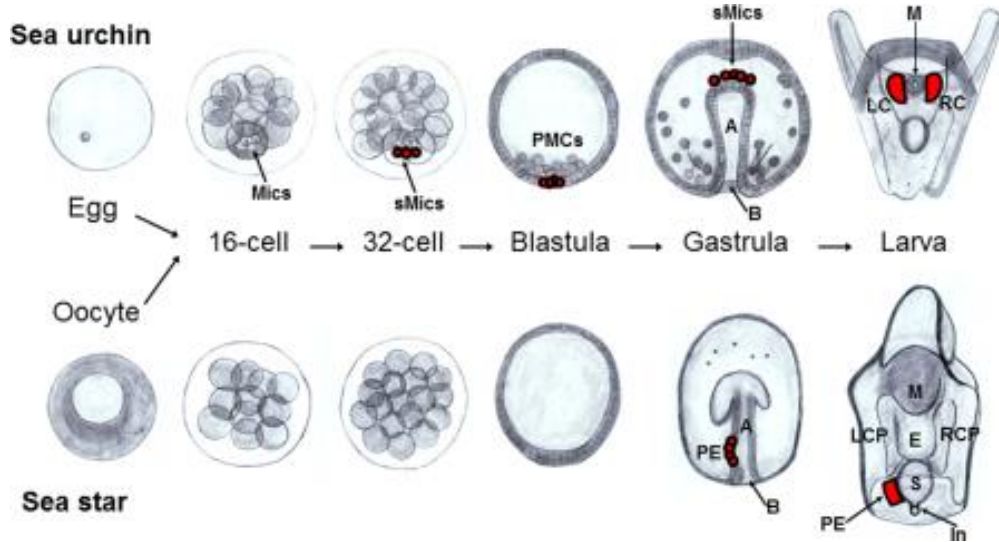
- Oulhen, N., and Wessel, G.M. (2013). Retention of exogenous mRNAs selectively in the germ cells of the sea urchin requires only a 5'-cap and a 3'-UTR. *Mol Reprod Dev* 80, 561-569.
- Oulhen, N., Yoshida, T., Yajima, M., Song, J.L., Sakuma, T., Sakamoto, N., Yamamoto, T., and Wessel, G.M. (2013). The 3'UTR of *nanos2* directs enrichment in the germ cell lineage of the sea urchin. *Dev Biol* 377, 275-283.
- Rodriguez, A.J., Seipel, S.A., Hamill, D.R., Romancino, D.P., Di Carlo, M., Suprenant, K.A., and 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., Barton, S.C., and Surani, M.A. (2002). A molecular programme for the specification of germ cell fate in mice. *Nature* 418, 293-300.
- Saitou, M., Payer, B., O'Carroll, D., Ohinata, Y., and Surani, M.A. (2005). *Blimp1* and the emergence of the germ line during development in the mouse. *Cell Cycle* 4, 1736-1740.
- Saitou, M., and Yamaji, M. (2010). Germ cell specification in mice: signaling, transcription regulation, and epigenetic consequences. *Reproduction* 139, 931-942.
- Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086-1092.
- Seervai, R.N.H., and Wessel, G.M. (2013). Lessons for inductive germline determination. *Mol Reprod Dev* 80, 590-609.
- Seydoux, G., and Braun, R.E. (2006). Pathway to totipotency: lessons from germ cells. *Cell* 127, 891-904.
- Shive, H.R., West, R.R., Embree, L.J., Azuma, M., Sood, R., Liu, P., and Hickstein, D.D. (2010). *brca2* in zebrafish ovarian development, spermatogenesis, and tumorigenesis. *Proceedings of the National Academy of Sciences* 107, 19350-19355.
- Smith, J., and Davidson, E.H. (2008). Gene regulatory network subcircuit controlling a dynamic spatial pattern of signaling in the sea urchin embryo. *Proc Natl Acad Sci U S A* 105, 20089-20094.
- Smith, J., Kraemer, E., Liu, H., Theodoris, C., and Davidson, E. (2008). A spatially dynamic cohort of regulatory genes in the endomesodermal gene network of the sea urchin embryo. *Dev Biol* 313, 863-875.
- Solana, J. (2013). Closing the circle of germline and stem cells: the Primordial Stem Cell hypothesis. *Evodevo* 4, 1-17.
- Song, J.L., Stoeckius, M., Maaskola, J., Friedlander, M., Stepicheva, N., Juliano, C., Lebedeva, S., Thompson, W., Rajewsky, N., and Wessel, G.M. (2012). Select microRNAs are essential for early development in the sea urchin. *Dev Biol* 362, 104-113.
- Speder, P., Petzoldt, A., Suzanne, M., and Noselli, S. (2007). Strategies to establish left/right asymmetry in vertebrates and invertebrates. *Curr Opin Genet Dev* 17, 351-358.
- Spotila, L.D., Spotila, J.R., and Hall, S.E. (1998). Sequence and expression analysis of *WT1* and *Sox9* in the red-eared slider turtle, *Trachemys scripta*. *Journal of Experimental Zoology* 281, 417-427.
- Stamateris, R.E., Rafiq, K., and Etensohn, C.A. (2010). The expression and distribution of *Wnt* and *Wnt* receptor mRNAs during early sea urchin development. *Gene Expr Patterns* 10, 60-64.
- Styhler, S., Nakamura, A., and Lasko, P. (2002). *VASA* localization requires the *SPRY*-domain and *SOCS*-box containing protein, *GUSTAVUS*. *Developmental Cell* 3, 865-876.
- Takeichi, M. (1988). The cadherins: cell-cell adhesion molecules controlling animal morphogenesis. *Development* 102, 639-655.
- Tanwar, P.S., Kaneko-Tarui, T., Zhang, L., Rani, P., Taketo, M.M., and Teixeira, J. (2010). Constitutive *WNT*/*beta*-catenin signaling in murine Sertoli cells disrupts their differentiation and ability to support spermatogenesis. *Biol Reprod* 82, 422-432.
- Toyooka, Y., Tsunekawa, N., Akasu, R., and Noce, T. (2003). Embryonic stem cells can form germ cells in vitro. *Proc Natl Acad Sci U S A* 100, 11457-11462.

- Uno, Y., Nishida, C., Yoshimoto, S., Ito, M., Oshima, Y., Yokoyama, S., Nakamura, M., and Matsuda, Y. (2008). Diversity in the origins of sex chromosomes in anurans inferred from comparative mapping of sexual differentiation genes for three species of the Raninae and Xenopodinae. *Chromosome Res* 16, 999-991011.
- Vandenberg, L.N., and Levin, M. (2013). A unified model for left-right asymmetry? Comparison and synthesis of molecular models of embryonic laterality. *Dev Biol* 379, 1-15.
- Vardy, L., and Orr-Weaver, T.L. (2007a). The Drosophila PNG kinase complex regulates the translation of cyclin B. *Dev Cell* 12, 157-166.
- Vardy, L., and Orr-Weaver, T.L. (2007b). Regulating translation of maternal messages: multiple repression mechanisms. *Trends Cell Biol* 17, 547-554.
- Venezky, D.L., Angerer, L.M., and Angerer, R.C. (1981). Accumulation of histone repeat transcripts in the sea urchin egg pronucleus. *Cell* 24, 385-391.
- Vincent, S.D., Dunn, N.R., Sciammas, R., Shapiro-Shalef, M., Davis, M.M., Calame, K., Bikoff, E.K., and Robertson, E.J. (2005). The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse. *Development* 132, 1315-1325.
- Voronina, E. (2013). The diverse functions of germline P-granules in *Caenorhabditis elegans*. *Mol Reprod Dev* 80, 624-631.
- Voronina, E., Lopez, M., Juliano, C.E., Gustafson, E., Song, J.L., Extavour, C., George, S., Oliveri, P., McClay, D., and Wessel, G. (2008). Vasa protein expression is restricted to the small micromeres of the sea urchin, but is inducible in other lineages early in development. *Developmental biology* 314, 276-286.
- Warner, J.F., Lyons, D.C., and McClay, D.R. (2012). Left-right asymmetry in the sea urchin embryo: BMP and the asymmetrical origins of the adult. *PLoS Biol* 10.
- Wessel, G.M., Brayboy, L., Fresques, T., Gustafson, E.A., Oulhen, N., Ramos, I., Reich, A., Swartz, S.Z., Yajima, M., and Zazueta, V. (2013). The biology of the germ line in echinoderms. *Mol Reprod Dev*.
- West, J.A., Viswanathan, S.R., Yabuuchi, A., Cunniff, K., Takeuchi, A., Park, I.-H., Sero, J.E., Zhu, H., Perez-Atayde, A., and Frazier, A.L. (2009). A role for Lin28 in primordial germ-cell development and germ-cell malignancy. *Nature* 460, 909-913.
- Yajima, M., and Wessel, G.M. (2011). The DEAD-box RNA helicase Vasa functions in embryonic mitotic progression in the sea urchin. *Development* 138, 2217-2222.
- Yajima, M., and Wessel, G.M. (2012). Autonomy in specification of primordial germ cells and their passive translocation in the sea urchin. *Development* 139, 3786-3794.
- Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y., and Saitou, M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat Genet* 40, 1016-1022.
- Yamaji, M., Ueda, J., Hayashi, K., Ohta, H., Yabuta, Y., Kurimoto, K., Nakato, R., Yamada, Y., Shirahige, K., and Saitou, M. (2013). PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. *Cell Stem Cell* 12, 368-382.
- Yamazaki, A., Furuzawa, Y., and Yamaguchi, M. (2010). Conserved early expression patterns of micromere specification genes in two echinoid species belonging to the orders clypeasteroidea and echinoidea. *Dev Dyn* 239, 3391-3403.
- Yamazaki, A., Kidachi, Y., and Minokawa, T. (2012). "Micromere" formation and expression of endomesoderm regulatory genes during embryogenesis of the primitive echinoid *Prionocidaris baculosa*. *Dev Growth Differ* 54, 566-578.
- Yankura, K.A., Koechlein, C.S., Cryan, A.F., Cheattle, A., and Hinman, V.F. (2013). Gene regulatory network for neurogenesis in a sea star embryo connects broad neural specification and localized patterning. *Proc Natl Acad Sci U S A* 110, 8591-8596.

- Zhou, G.-B., Meng, Q.-G., and Li, N. (2010). In vitro derivation of germ cells from embryonic stem cells in mammals. *Mol Reprod Dev* 77, 586-594.

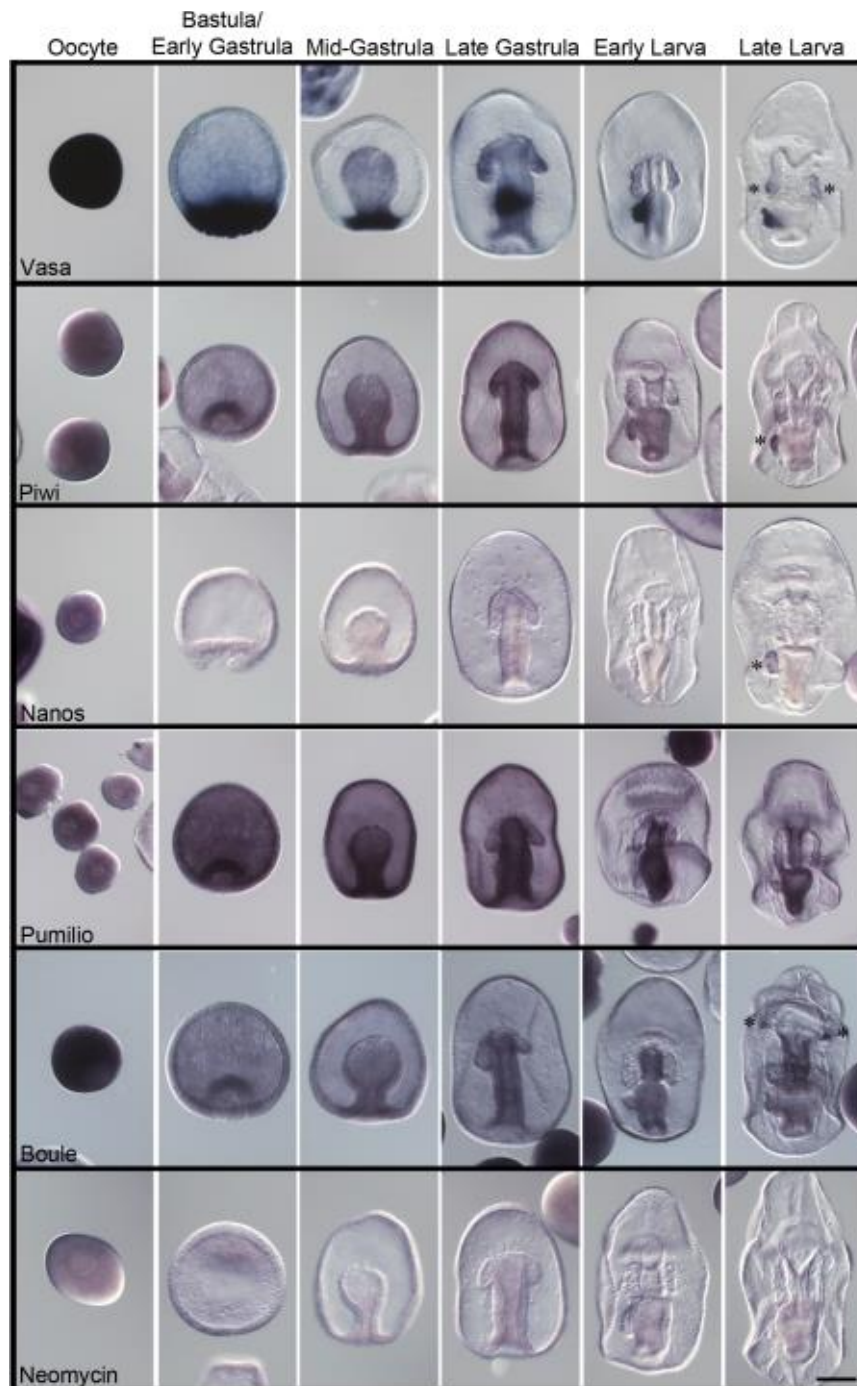
FIGURES AND TABLES

Figure 1. Schematic representation of the developmental stages in sea star and sea urchin.



Sea urchins exhibit two asymmetric cleavage events within the cells at the vegetal pole (bottom). These asymmetric divisions result in 4 micromeres forming at the 16-cell stage, followed by 4 large and 4 small micromeres (sMics, in red) forming at the 32-cell stage. The PMC (descendants of the large micromeres) ingress to form the skeleton, whereas descendants of the sMics remain relatively quiescent through embryogenesis, divide only once before gastrulation, and then integrate into the larval coelomic pouches where the adult rudiment forms. Unlike in sea urchins, the sea star has symmetrical cell divisions and does not segregate its germ-line cells during early development. Instead, a PE (in red) projects from the dorsal wall of the archenteron into the blastocoel and then moves to the left side in late gastrula-early larval stages. Left and right anterior (top) coelomic pouches are present on both sides of the esophagus in larvae, and they subsequently extend posteriorly. Later in development, the left coelomic pouch integrates cells of the PE as it extends posteriorly. The blastopore is located at the lower opening of each embryo in the gastrula stage. Mics=Micromeres, sMics=Small micromeres, PMCs=Primary mesenchyme cells, LCP=Left coelomic pouch, RCP=Right coelomic pouch, PE=Posterior enterocoel, LC=Left coelom, RC=Right coelom, M=Mouth, E=Esophagus, S=Stomach, In=Intestine, B=Blastopore, A=Archenteron.

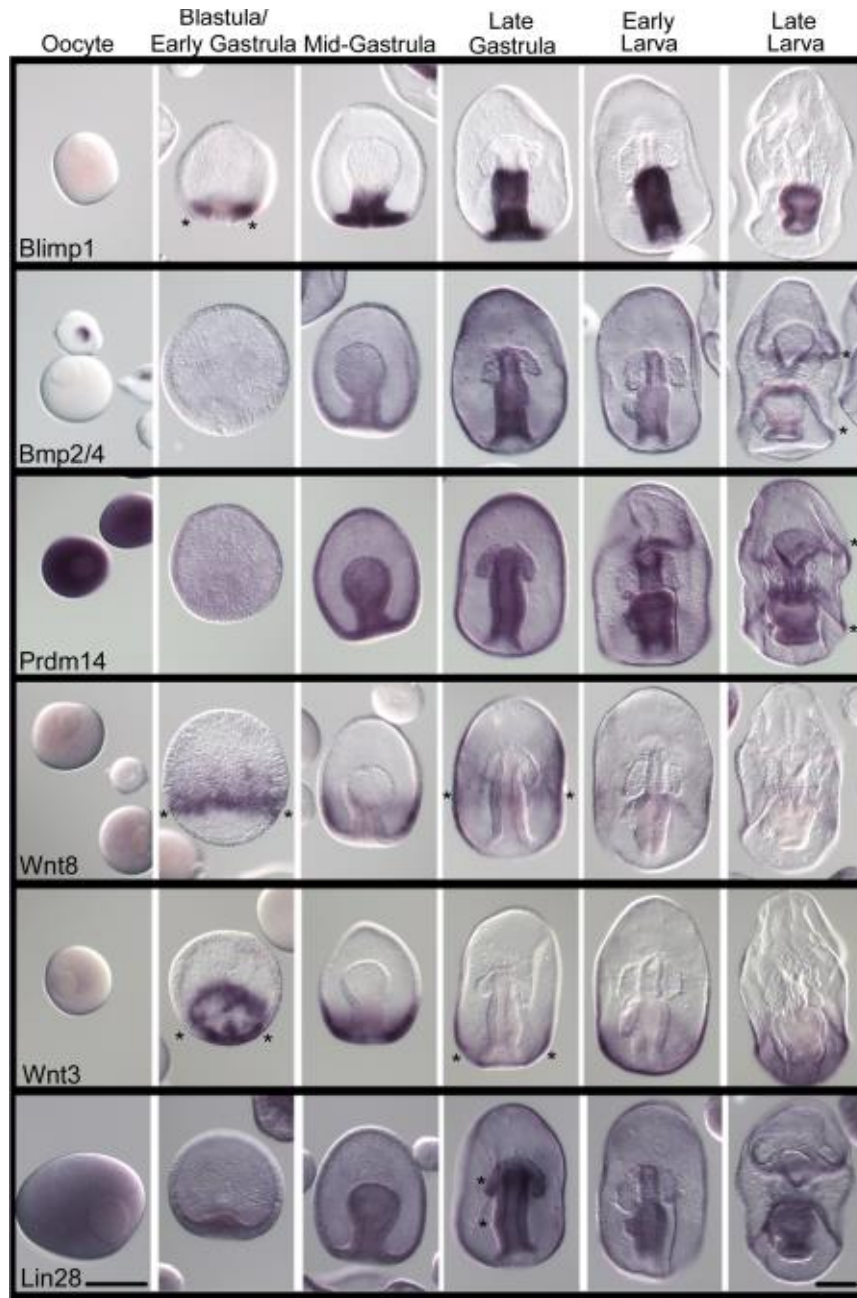
Figure 2. Expression of conserved germ-line determinants during *P. miniata* embryonic development.



Line 1) Vasa transcripts are widespread in oocytes and become restricted to the vegetal pole in blastula stage embryos. During gastrulation Vasa is then restricted to the center of the archenteron and later to the left side of the archenteron. As soon as the PE is formed Vasa persists in the PE. **Line 2) Piwi** transcripts are widespread in oocytes, become restricted to the archenteron during gastrulation, then to the center of the archenteron during late gastrulation, and persist in the PE of

larva as soon as it is formed. **Line 3) Nanos** expression is only detected broadly in immature oocytes and in the PE of older larva. **Line 4) Pumilio** transcripts accumulate broadly throughout the embryo during early development and become enriched in the esophagus and stomach during larval stages. **Line 5) Boule** transcripts are widespread in oocytes but are not present during development until larval stages. Boule transcripts accumulate in the oral ectoderm/ciliary band and in the esophagus. A sequence for the **Neomycin (Neo) resistance gene (Line 6)** was used as a negative control for the hybridization procedure. Asterisks (*) represent areas of emphasis for mRNA accumulation. Dorsal views of the larva. Scale bar represents 100 μ m.

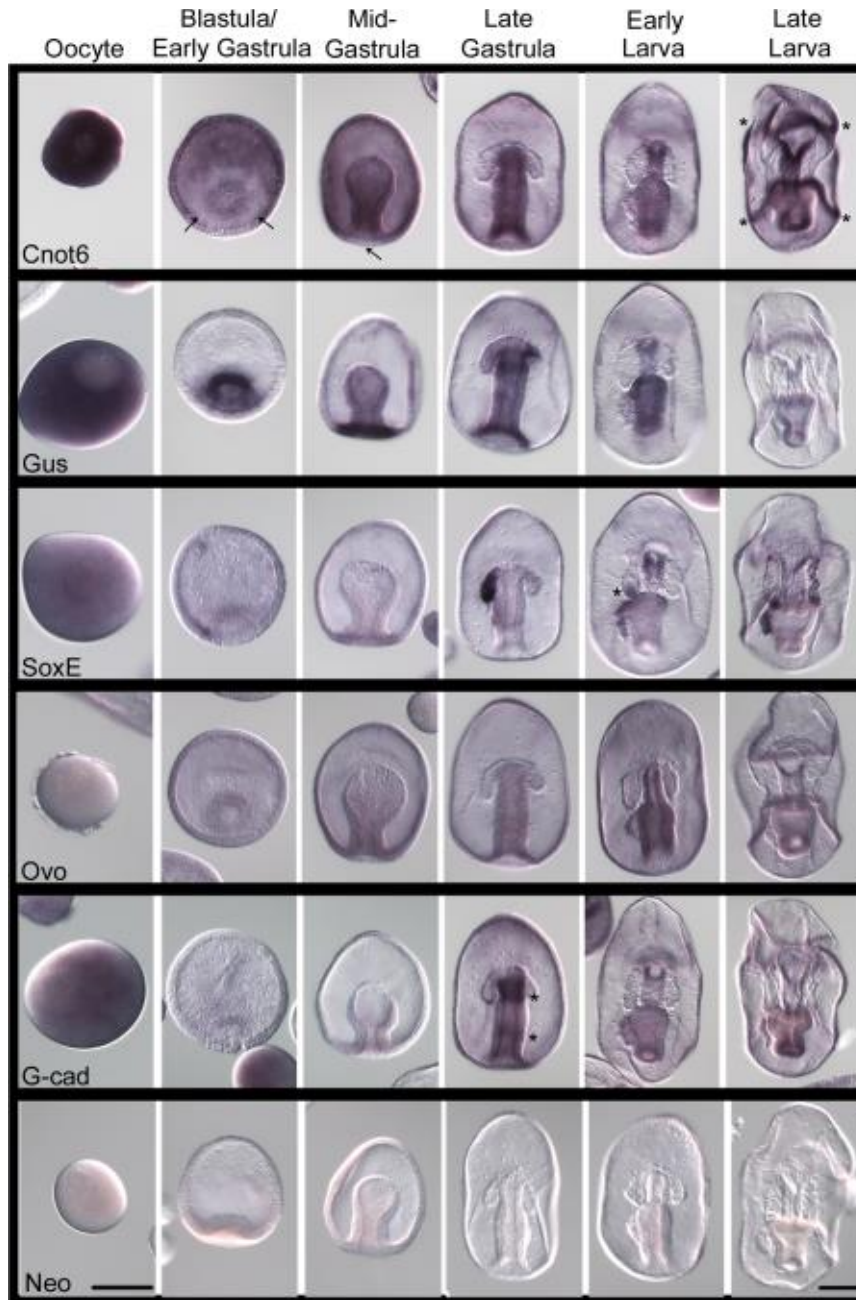
Figure 3. Expression of genes involved in inductive germ-line specification during *P. miniata* embryonic development.



In situ hybridization showing **Line 1) Blimp1** transcripts, which first accumulate in the vegetal pole of the blastula, and remain largely endodermal through development. Of note is the rapid loss of Blimp1 in the newly formed PE (see also Figure 7). **Line 2) Bmp2/4** transcripts are seen in young oocyte germinal vesicles (oocyte nucleus). During development Bmp2/4 is expressed throughout the embryo except for in the coelomic pouches and PE. **Line 3) Prdm14** is present throughout development and largely throughout the embryo with the exception of the coelomic pouches. **Lines 4 and 5) Wnt 8 and Wnt 3** (respectively) form a complementary pattern with

Wnt3 being more vegetal and Wnt 8 more equatorial in the embryo and larva. No selective accumulation is seen in the coelomic pouches or PE in larval stages. **Line 6) Lin28** is expressed broadly during early development until the late gastrula stage when it becomes enriched in the foregut. In larval stages, Lin28 is enriched in the gut and PE and it is distinctly absent from the coelomic pouches. Asterisks (*) show areas of emphasis for transcript detection. Dorsal views of the larva. Scale bar represents 100µm.

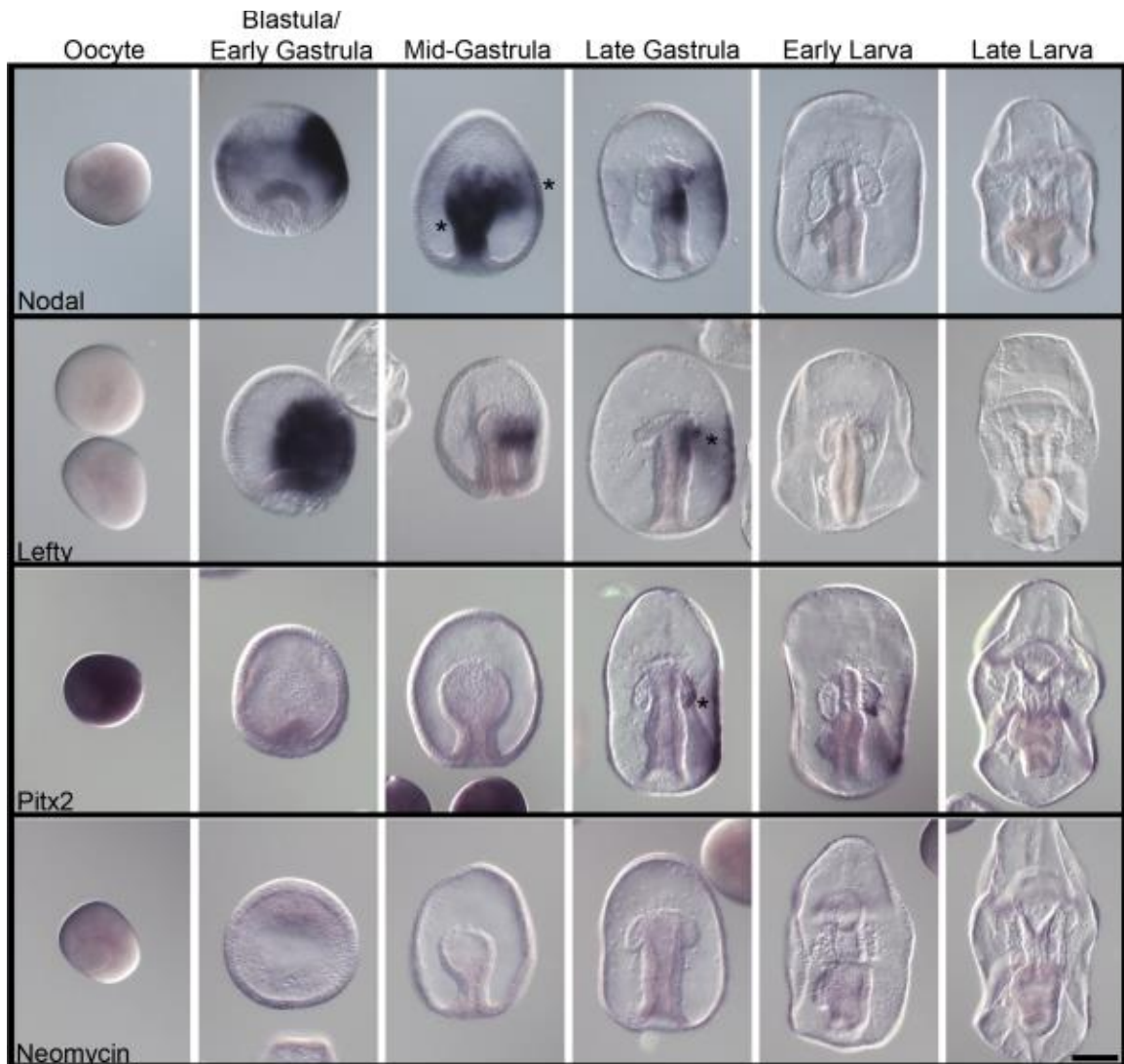
Figure 4. Expression of germ-line associated genes during *P. miniata* embryonic development.



Line 1) Cnot6 is nearly uniform throughout development and is absent from the vegetal-most region of the embryo, the coelomic pouches, and the PE as well as the ciliary band in larvae. **Line 2) Gustavus (Gus)** is enriched in the developing mesoderm in early development. Gus transcripts become enriched in the mouth and stomach in larval stages. **Line 3) SoxE** (the Sox 9/10 ortholog) is largely absent from the early embryo but accumulates significantly in the left coelomic pouch following gastrulation. SoxE transcripts then accumulate in the PE and the tips of both coelomic pouches during larval stages. **Line 4) Ovo** is present in the gut of the embryo and larva, with no significant enrichment in the pouches or the PE. **Line 5) G-cadherin** accumulates in the gut

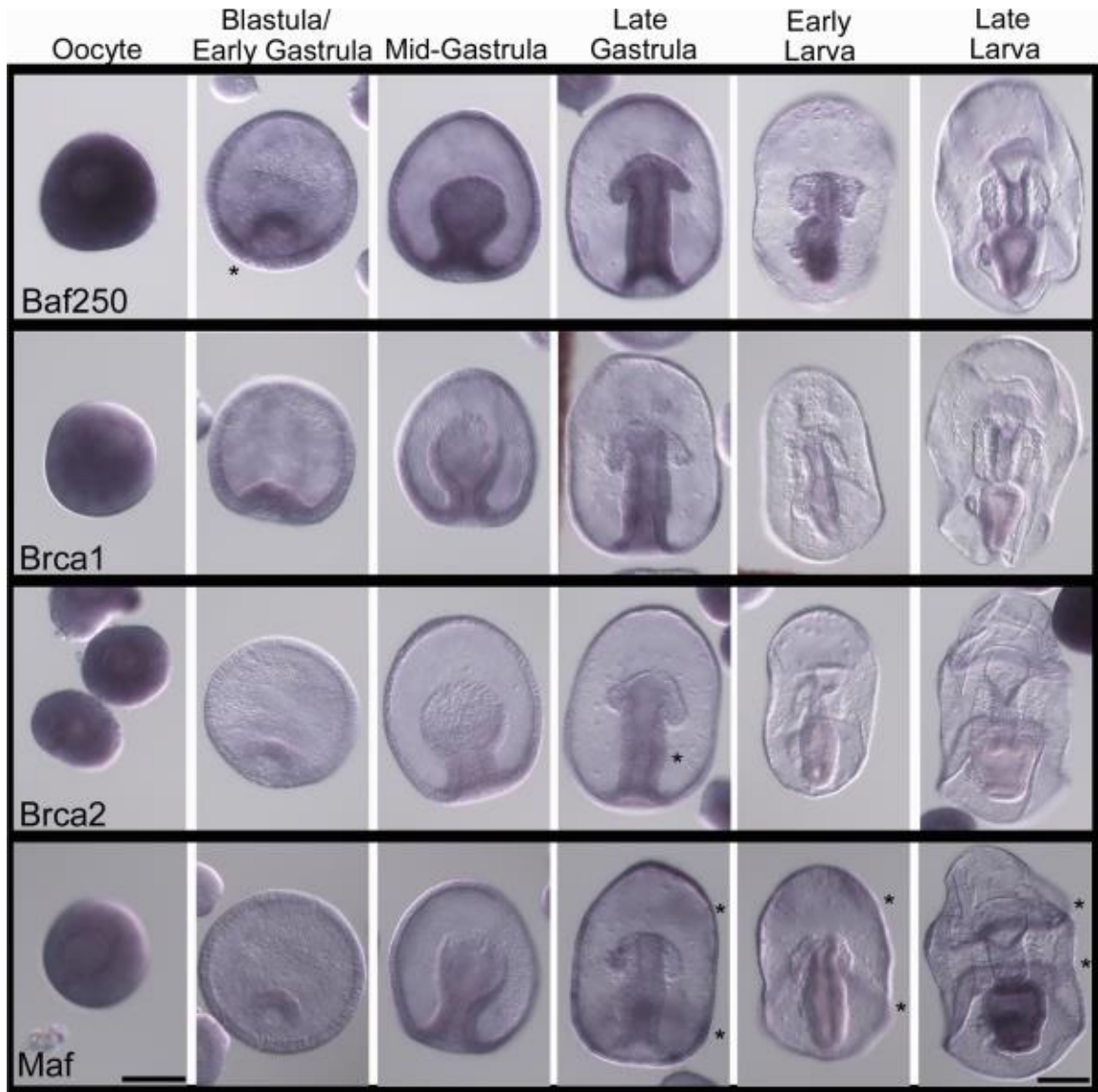
during the late gastrula stage in distinct bands in the foregut and midgut regions. In larval stages G-cadherin becomes enriched in the mouth and stomach. A sequence for the **Neomycin (Neo) resistance gene (Line 6)** was used as a negative control for the hybridization procedure. Asterisks (*) show areas show areas of emphasis for transcript detection. Arrows emphasize notable areas of transcript depletion. Dorsal views of the larva. Scale bar represents 100 μm .

Figure 5. Expression of left/right asymmetry markers during *P. miniata* embryonic development.



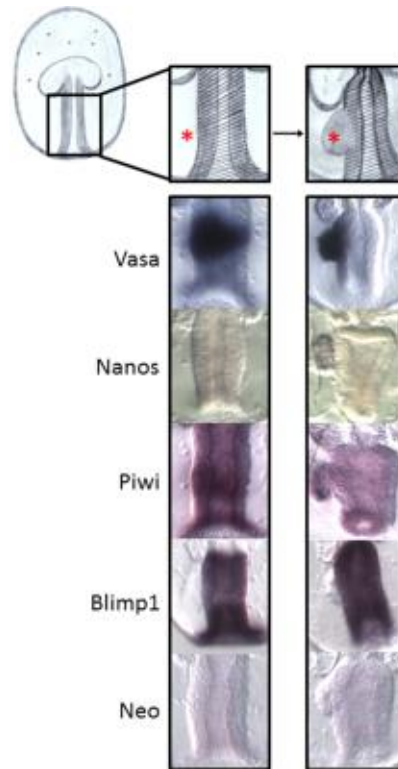
Line 1) Nodal is first expressed in the ectoderm at the blastula stage. A second expression domain of Nodal appears in mid-gastrula stage embryos when it is expressed symmetrically in the developing archenteron. Nodal expression becomes restricted to the right side of the archenteron in late-gastrula stage embryos. **Line 2) Lefty** is first expressed in the ectoderm at the blastula stage. A second expression domain of Lefty appears in late-gastrula stage embryos when it is expressed asymmetrically in the right side of the archenteron. **Line 3) Pitx2** is first expressed ubiquitously in oocytes. Pitx2 expression localizes to the right coelomic pouch and right ectoderm in late gastrula staged embryos stage. Pitx2 expression in the right coelomic pouch and right ectoderm persists through late larval stages. A sequence for the **Neomycin (Neo) resistance gene (Line 4)** was used as a negative control for the hybridization procedure. Asterisks (*) show areas of emphasis for transcript detection. Dorsal views of the larva. Scale bar represents 100 μm .

Figure 6. Molecules involved in genomic regulation and maintenance during *P. miniata* embryonic development.



Line 1) Baf250 is expressed ubiquitously in the embryo throughout development with some enrichment in the gut and depletion in the coelomic pouches. **Lines 2 and 3) Brca1 and 2** are enriched in the gut of late gastrula embryos and we note no significant accumulation in the PE. **Line 4) Maf** is most apparent in the stomach of the larva and no specific enrichment in the PE. Asterisks (*) show areas of emphasis for transcript detection. Dorsal views of the larva. Scale bar represents 100 μ m.

Figure 7. Transcript dynamics during posterior enterocoel formation.



The conserved germ-line determinant, Nanos, accumulates specifically in the PE (red asterisk) after formation. During gastrulation, the conserved germ-line determinants Vasa and Piwi, become enriched in the midgut (shaded area). As soon as the PE is formed Vasa and Piwi transcripts start to become restricted to the PE and to clear from the nearby stomach. This is in stark contrast to the somatic cell marker, Blimp1. During gastrulation, Blimp1 transcripts similarly become enriched in the midgut, however, as soon as the PE is formed, Blimp1 transcripts are restricted to the stomach and clear from the nearby PE.

Table 1. Conserved germ-line determinants.

| Genes | Domains and Function | Reference | Orthologs (organism) | %Identities, % Similarities | NCBI or Spbase Reference number | Pm transcript number | Primers sequences | Size of amplify product |
|----------------|---|---|--|------------------------------------|--|-----------------------------|--|--------------------------------|
| Piwi | PAZ (piwi, argonaute, zwiille) domain, required for negative regulation of transposable elements in germ cells | Juliano <i>et al.</i> , 2006 | Mm-Piwi1L Dm-Argonaute 3, isoform G Sp-Seawi | 51%, 74% 40%, 59% 50%, 71% | NP_067286.1 NP_001163498.1 AAG42534.1 | Pm_33095 | F:CGACGGCAGCCA GATCACCTA R:taatacgactcactataggg CCAGGCAGCAGTA CTTCTTGA | 728 bp |
| Nanos | CCHC Zn- finger domain, With partmer, Pumilio, serves as a negative regulator of translation in germ cells | Juliano <i>et al.</i> , 2006 Lai <i>et al.</i> , 2013. | Mm-Nanos2 Dm-Nanos Sp-Nanos2 | 54%, 65% 62%, 73% 39%, 50% | NP_918953.2 NP_476658.1 NP_001073023.1 | Pm_4079 | F:GGAGATTGAGAG CGAAGAT R:taatacgactcactataggg TGTTGAATTTTCATG AGGCAAA | 971 bp |
| Vasa | CCHC Zn-finger and DEAD-box domains, ATP-dependent RNA helicase involved in germ-line specification | Juliano <i>et al.</i> , 2006 | Mm-Vasa Dm-Vasa Sp-Vasa | 61%, 76% 53%, 69% 69%, 82% | NP_034159.1 NP_723899.1 NP_001139665.1 | Pm_1519 | F:CGGTCCAGAAGT ACGGGATA R:taatacgactcactataggg GTAGAAGCTGGTT GCCTTGC | 992 bp |
| Pumilio | Puf domain, With partner Nanos, RNA-binding protein that negatively regulates translation of target RNA in germ cells | Lai <i>et al.</i> , 2013. | Mm-Pum2 Dm-PumF Sp-Pum | 55%, 65% 79%, 88% 65%, 71% | NP_109648.2 NP_001247002.1 SPU_006847 | Pm_2787 | F:GGTAGTAACATG GGGGACCAG R:taatacgactcactataggg GGCCTTGTGTTGA CCTTGCT | 805 bp |

| | | | | | | | | |
|------------------------|--|--|----------------------------------|----------------------------------|--|----------|--|--------|
| Boule/ Dazl | RNA binding protein involved in germ-line determination in mammals | Xu <i>et al.</i> , 2001. Shah <i>et al.</i> , 2010. | Mm-Boule Dm-Boule Sp-Boule | 59%, 71% 51%, 62% 37%, 46% | NP_083543.2 NP_729457.1 SPU_008194.1 | Pm_22341 | F:TCGGTTCATAAC TGCCATCA R:taatagactcactataggg TTATGGCACCTG GTGAGAG | 925 bp |
|------------------------|--|--|----------------------------------|----------------------------------|--|----------|--|--------|

Table 2. Genes involved in the inductive mechanisms of germ-line specification.

| Genes | Domains and Function | Reference | Orthologs (organism) | %Identities, % Similarities | NCBI or Spbase Reference number | Pm transcript number | Primers sequence | Size of amplify product |
|----------------------|--|---|---|------------------------------------|---|----------------------------------|--|-------------------------|
| Blimp1/ Prdm1 | Transcription factor involved in germ line determination in mice | Saitou <i>et al.</i> , 2010. Kurimoto <i>et al.</i> , 2008. | Mm-Prdm1 Dm-Blimp1 Sp-Blimp1/Krox1b | 78%, 87% 67%, 79% 43%, 55% | AAI29802.1 NP_647982.1 NP_001073021.1 | Pm_43022 (AAP35029.1) | F:CCATTCTCCGT ACTCGTGGT R:taatacgactcactatag ggCGGAAGTCTG TGCATGAGAA | 912 bp |
| Prdm14 | PR domain Zn-finger, transcriptional regulator involved in germ line determination in mice | Saitou <i>et al.</i> , 2010. Yamaji <i>et al.</i> , 2008 | Mm-Prdm14 Dm-Prdm14 Sp-Prdm14 | 87%, 94% 37%, 55%** 94%, 97% | NP_001074678.2 XP_794184.3 | Pm_65577 | F:AACCGCTCTTC CGATCTGT R:taatacgactcactatag ggGTGTGACGCG AAGGCTTTT | 402 bp |
| BMP2/4 | TGF-beta ligand signaling molecule. Critical for induction of PGCs in mice | Saitou <i>et al.</i> , 2010. Saitou, 2009. | Mm-BMP2 Dm-Decapentaplegic, isoformA Sp-BMP 2/4 | 45%, 61% 40%, 55% 50%, 65% | NP_031579.2 NP_477311.1 SPU_021497 | Pm_4348 | F:CGTGCCACAGT ACATGCTGGA R:taatacgactcactatag ggGCTCGCTGAC AGACCGAGCTA | 590 bp |
| Lin28 | Cold shock and a cluster of two CCHC Zn-finger domains, RNA binding protein that is required for PGC development | Bussing <i>et al.</i> 2008. West <i>et al.</i> 2009. | Mm-Lin28B Dm-Lin28 Sp-Lin28 | 53%, 61% 55%, 75% 64%, 80% | NP_001026942.1 NP_647983.1 SPU_027195 | Pm_90489 | F:GGCCGACGAG GGCAAGCTGTG R:taatacgactcactatag ggGGCCAGTCAC CGACTCCGCT | 215 bp |
| Wnt3 | Wnt ligand signaling molecule. Involved in germ line competency in mice. | Saitou <i>et al.</i> , 2010. Ohinata <i>et al.</i> , 2009. | Mm-Wnt3 Dm-Wnt2 Sp-Wnt3 | 53%, 69% 43%, 57% 61%, 73% | NP_033547.1 NP_476810.1 XP_790595.2 | Combo of: HP125189.1 and cloning | F:TAAATTCATCA GCCCCAAGG R:taatacgactcactatag ggATGGCTTCGTT CTTGAATGC | 975 bp |

| | | | | | | | | |
|-------------|--|--|---|----------------------------------|---|----------|---|--------|
| Wnt8 | Wnt ligand signaling molecule. Correlates with localization of PGC's in axolotl. | Bachvarov <i>a et al.</i> , 2001. Johnson <i>et al.</i> , 2003. | Mm-Wnt8b Dm-Wingless, isoform B Sp-Wnt8 | 59%, 73% 40%, 58% 48%, 67% | NP_035850.2 NP_723268.1 NP_999832.1 | Pm_82262 | F:GCAGCGACAA CATCAAATTCG R:taatac gactcactatag ggGCTCTTCCGAT CTGACGGCTG | 365 bp |
|-------------|--|--|---|----------------------------------|---|----------|---|--------|

Table 3. Germ-line associated genes.

| Genes | Domains and Function | Reference | Orthologs (organism) | %Identities, % Similarities | NCBI or Spbase Reference number | Pm transcript number | Primers sequence | Size of amplify product |
|-----------------|---|--|---|----------------------------------|---|----------------------|---|-------------------------|
| SoxE | HMG box, transcription factor involved in sex determination. | Juliano <i>et al.</i> , 2006 | Mm-Sox9 Dm-Sox100B Sp-SoxE | 77%, 90% 61%, 79% 82%, 89% | NP_035578.3 NP_651839.1 SPU_016881 | Pm_60849 | F:CGTCTATTCGCGAC GCCGTGT R:taatacgaactcactatagggG CTTCTTCCTTGGGTG TGGTC | 496 bp |
| Ovo | C2H2 Zn-finger domain, transcription factor involved in Drosophila oogenesis and mouse spermatogenesis | Dai <i>et al.</i> , 1998 | Mm-Ovo2L, isoform A Dm-Ovo, isoform A Sp-Ovo | 59%, 71% 69%, 72% 53%, 66% | NP_081200.2 NP_525077.2 SPU_012448 | Pm_17822 | F:CCACAGACGACAC ACATCTCA R:taatacgaactcactatagggG TGAAGGCCTTGCTGC AATAC | 503 bp |
| Gustavus | SPRY and SOCS box domains, E3 ubiquitin ligase specific receptor involved in the regulatory balance of Vasa ubiquitylation. | Styhler <i>et al.</i> , 2002 Gustafson <i>et al.</i> , 2011 | Mm-Gustavus Dm-Gustavus isoform G Sp-Gustavus | 62%, 78% 65%, 79% 74%, 90% | NP_083311.1 NP_00124614 0.1 SPU_004717 | Pm_19970 | F:GGAGGATCTTCGG AGCGGTAG TGCC R:taatacgaactcactatagggG GCAGTGGTAGCTGG TAGATGTCT T | 804 bp |
| Cnot6 | EEP domain, deadenylase involved in mRNA decay | Swartz <i>et al.</i> , 2013. Wahle <i>et al.</i> , 2013. | Mm-Cnot6 Dm-TwinB Sp-Cnot6 | 60%, 77% 59%, 76% 66%, 78% | NP_997649.1 NP_732967.1 XP_779942.3 | Pm_19426 | F:CTGGACCTATCGG CGAATAA R:taatacgaactcactatagggT GATGGTCTGGATCA GCTTG | 877 bp |

| | | | | | | | | |
|-------------------|--|--------------------------|---|----------------------------------|--|---------|---|--------|
| G-Cadherin | Laminin G, Ca ²⁺ -binding EGF-like, and Cadherin tandem repeat domains, cell-cell adhesion molecule | Yajima and Wessel, 2012. | Mm-Fat tumor suppressor 1 Dm-NCad isoform G Sp-GCad | 26%, 43% 34%, 51% 39%, 54% | NP_00107475 5.2 AAN10997.1 SPU_010840 | Pm_6651 | F:CGACAAGTTCAGG CTAGACTC R:taatacgaactactataggG TGACGACAATGTCG ATGGTG | 690 bp |
|-------------------|--|--------------------------|---|----------------------------------|--|---------|---|--------|

Table 4. Left-Right asymmetry molecules.

| Genes | Domains and Function | Reference | Orthologs (organism) | %Identities, % Similarities | NCBI or Spbase Reference number | Pm transcript number | Primers sequence | Size of amplify product |
|--------------|--|------------------------------|---------------------------------|----------------------------------|--|--|---|-------------------------|
| Nodal | Tgf-beta ligand Involved in left-right asymmetry | Grande <i>et al.</i> , 2009. | Mm-Nodal Dm-Dpp Sp-Nodal | 51%, 74% 42%, 62% 53%, 69% | NP_038639 NP_477311.1 ABK33664.1 | gi 307052073 gb HP126404.1 isotig21602.Pminagast | F:CGGTGGATCGTCTACCCTAA R:taatacgactcactataggCCCGATCAAATTGTAAAAATGC | 966 bp |
| Lefty | Tgf-beta signaling molecule, involved in left-right asymmetry | Grande <i>et al.</i> , 2009. | Mm-Lefty Dm-Dpp Sp-Lefty | 24%, 40% 21%, 36% 28%, 46% | NP_034224.1 NP_477311.1 NP_001123281.1 | lcl scaffold51185675.8 | F:ATGGAGTCTCGCGTAGCTGT R:taatacgactcactataggCATGTTTGTGACGGGTCTG | 537 bp |
| Pitx2 | Homeobox domain, transcription factor involved in left-right asymmetry | Grande <i>et al.</i> , 2009. | Mm-Pitx2 Dm-Ptx1 Sp-Pitx2 | 74%, 82% 60%, 65% 73%, 82% | NP_001035969.1 NP_733410.2 SPU_004599 | Pm_22862 | F:GCGTCAGGGTG TGGTTTAAAG R:taatacgactcactataggGTTCAAGTTCTGTGGCTCA | 329 |

Table 5. Regulation and genomic maintenance during morphogenesis and early embryogenesis.

| Genes | Domains and Function | Reference | Orthologs (organism) | %Identities, % Similarities | NCBI or Spbse Reference number | Pm trascript number | Primers sequence | Size of amplify product |
|---------------|--|-------------------------------------|-----------------------------------|----------------------------------|---|---------------------|--|-------------------------|
| Brca1 | N-terminus RING finger domain, two nuclear localization signals and an acidic C-terminus. It is involved in DNA repair, transcriptional regulation, chromatin remodeling, cellular growth control and genome stability | Hoshino <i>et al.</i> , 2007 | Mm-Brca1 Sp-Brca1 | 34%, 55% 62%, 75% | NP_033894.3 SPU_011027.3a | Pm_3175 | F:GGATCTTCCCAG AGTACGACT R:taatacactactataggg GGAGCCAAGAGTT GTCAGAGT | 822 bp |
| Brca2 | N-terminus acidic transcriptional activation domain and a C-terminus DNA binding domain. It is involved in maintenance of genomic stability in response to DNA damaging agents | Rodríguez-Marí <i>et al.</i> , 2011 | Mm-Brca2 Sp-Brca2 | 36%, 57% 42%, 62% | NP_001074470.1 SPU_013435 | Pm_34269 | F:GGAGAAGCACA GCGAGGGAGG R:taatacactactataggg GGCTGGAACCCAG CCTGAAGA | 729 bp |
| Baf250 | It has a DNA binding domain called ARID. It is a component of catalytic cores that regulate the expression of Homeobox genes early in development | Li <i>et al.</i> , 2010 | Mm-Arid1B Dm-OsaA Sp-Baf250 | 43%, 69% 38%, 52% 75%, 88% | NP_001078824.1 NP_732263.1 SPU_023530 | Pm_47085 | F:CCAGTGTGCATG CCCTCAGTA R:taatacactactataggg CCTCTCTCCTTAAC TGCATGG | 506 bp |

| | | | | | | | | |
|-------------------|--|----------------------------------|---------------------------------------|---|---|-----------------|---|---------------|
| <p>Maf</p> | <p>Traffic Jam/Maf is a transcription factor that controls gonad morphogenesis. TJ protein activates piwi expression and tj gene is a piRNA cluster which define the Piwi targets for silencing, in Drosophila</p> | <p>Saito <i>et al.</i>, 2009</p> | <p>Mm-MafB Dm-Tjam Sp-Maf</p> | <p>51%, 72% 57%, 72% 61%, 80%</p> | <p>NP_034788.1 NP_609969.2 SPU_025888</p> | <p>Pm_71778</p> | <p>F:CCAAGCCTTGAT GAGCTCTAT R:taatacgactcactataggg GGACTACTCGGCA AACTAACG</p> | <p>629 bp</p> |
|-------------------|--|----------------------------------|---------------------------------------|---|---|-----------------|---|---------------|

Chapter IV: Diversity in the fertilization envelopes of echinoderms

Nathalie Oulhen, **Adrian Reich**, Julian L. Wong, Isabela Ramos, Gary M. Wessel

Evolution & Development 15, no. 1 (2013): 28-40.

CONTRIBUTION

I assembled and annotated the four *de novo* transcriptomes used in this study, and identified putative orthologous genes in the echinoderms by cluster analysis. I also constructed the *P. miniata* peptide database used for all the mass spec analyses.

RNA was extracted from the ovaries of all echinoderms in this study using the RNEasy Mini kit (Qiagen) with on column DNA digestion. The isolated RNA was processed with the Illumina mRNA-Seq kit, using standard procedures. Each sample was sequenced on a GAIIX using a single lane per organism with a paired end read length of 105bp. The individual transcriptomes were assembled using Velvet (1.0.09) and Oases (0.1.14) with a k-mer of 31 (Schulz *et al.*, 2012). From each collection of loci, a single exemplar sequence was selected that was the most abundant and at least 80% the length of the longest member of the locus. The exemplar sequences were compared with every other exemplar and with the *S. purpuratus* SPU gene predictions (Sea Urchin Genome Sequencing *et al.*, 2006) using pair wise BLASTx with a minimum score of $1e^{-5}$. All of the sequences were then clustered using MCL (Enright *et al.*, 2002) to identify putative orthologous sequences. Given an annotated SPU designation from *S. purpuratus* I extracted orthologs for all organisms in this study and using a custom FileMaker database, the sequences could be shared with collaborators. Putative orthologs from the MCL clustering analysis were tested further with CLUSTALW alignments and BLAST.

Exemplar sequences were also annotated with BLAST2GO (Conesa *et al.*, 2005) and these annotated sequences were translated in all six frames. The longest open reading frames from each of the six frames were used as a peptide library for mass spectrometry analysis.

ABSTRACT

Cell surface changes in an egg at fertilization are essential to begin development and for protecting the zygote. Most fertilized eggs construct a barrier around themselves by modifying their original extracellular matrix. This construction usually results from calcium induced exocytosis of cortical granules, the contents of which in sea urchins function to form the fertilization envelope (FE), an extracellular matrix of cortical granule contents built upon a vitelline layer scaffold. Here we examined the molecular mechanism of this process in sea stars, a close relative of the sea urchins, and analyze the evolutionary changes that likely occurred in the functionality of this structure between these two organisms. We find that the FE of sea stars is more permeable than in sea urchins, allowing diffusion of molecules in excess of 2 megadaltons. Through a proteomic and transcriptomic approach, we find that most, but not all of the proteins present in the sea urchin envelope are present in sea stars, including SFE9, proteoliasin, rendezvin, and ovoperoxidase. The mRNAs encoding these FE proteins accumulated most densely in early oocytes, and then beginning with vitellogenesis, these mRNAs decreased in abundance to levels nearly undetectable in eggs. Antibodies to the SFE9 protein of sea stars showed that the cortical granules in sea star also accumulated most significantly in early oocytes, and different from sea urchins, they translocated to the cortex of the oocytes well before meiotic initiation. These results suggest that the preparation of the cell surface changes in sea urchins has been shifted to later in oogenesis and perhaps reflects the meiotic differences among the species – sea star oocytes are stored in prophase of meiosis and fertilized during the meiotic divisions, as in most animals, whereas sea urchins are one of the few taxa in which eggs have completed meiosis prior to fertilization.

INTRODUCTION

Reproductive strategies differ amongst organisms based on their evolutionary history and the niche within which they compete. The reproductive strategy for most marine invertebrates includes broadcast spawning of their gametes, and if successful in fertilization, the embryos often utilize the water column as a food source for development before metamorphosing into an adult. Echinoderms are paradigmatic for this reproductive strategy, and have served as important research organisms for understanding mechanisms of sperm activation (Lee *et al.*, 1983), chemoattraction of sperm to the egg (Ward *et al.*, 1985), sperm-egg binding mechanisms (Vacquier and Moy, 1977), egg activation (Steinhardt *et al.*, 1977), and the diverse evolutionary basis for sperm-egg interactions (Vacquier, 1998).

Animal fertilization was first observed in sea urchins, where an envelope forms promptly after sperm fusion with the egg and thus provides a rapid metric for successful sperm-egg interaction (Briggs and Wessel, 2006; Derbes, 1847). Fusion of the male and female pronuclei, when first seen in sea urchins by Hertwig (1886) and Fol (1877), closed the chapter on the important role of sperm in the process of reproduction.

The extracellular matrix of the egg, while called many different names, e.g. vitelline layer, zona pellucida, serves two essential jobs. First, it interacts with sperm in a species-specific manner. While this function occurs in almost all animals, it is particularly striking in broadcast spawners, such as abalone and sea urchins, which can inhabit the same niches and often spawn in overlapping times. In such cases, species specificity in sperm-egg interactions relies heavily on the extracellular matrix. Following successful sperm-egg fusion, the egg's extracellular matrix quickly reveals its second job as it is transformed to minimize the chances of additional sperm from reaching the egg. This physical block to polyspermy is highly selected for because fusion of more than one sperm with an egg is lethal to the embryo. The block to polyspermy in some animals, such as sea urchins, is remarkable since sperm:egg ratios may reach the millions.

The fertilization envelope in sea urchins establishes a physical and biochemical barrier that protects the zygote from supernumerary sperm, as well as environmental and microbial agents (Wong and Wessel, 2006a). Cortical granules are the major source of proteins used to construct the fertilization envelope, (Wessel *et al.*, 2001; Wong and Wessel, 2006a). These abundant organelles, ranging to 15,000 per egg in sea urchins, are synthesized during oogenesis and released following gamete fusion (Laidlaw and Wessel, 1994). In the sea urchin, contents of the cortical granules are secreted within 30 sec of insemination and mix with the egg's vitelline layer. Hydrostatic pressure and addition of glycoproteins from the cortical granules to the vitelline layer lift the nascent fertilization envelope off the egg surface, and associated enzymes transform the envelope into an effective barrier for early embryogenesis. Sea urchin cortical granules harbor the major structural proteins of the envelope as well as enzymes essential to stabilize the envelope until hatching (Wong and Wessel, 2008).

The cortical granules contain several structural proteins and enzymes that give the fertilization envelope its distinct properties of stability yet permeability in the ocean environment. These proteins include the Soft Fertilization Envelope proteins SFE1 and SFE9, proteoliain, and rendezvin; their cognate transcripts are specifically expressed in oocytes (Laidlaw and Wessel, 1994; Wong and Wessel, 2004, 2006b). SFE1, SFE9 and proteoliain are proteins rich in low-density lipoprotein receptor type A (LDLR_A) repeats involved in protein interaction (Wessel, 1995; Wessel *et al.*, 2000; Wong and Wessel, 2004). Rendezvin (RDZ) is enriched in CUB domains, also involved in protein interaction. One RDZ gene is present in the sea urchin genome, but several transcripts are produced after alternative splicing. The full-length *rdz* transcript is alternatively spliced into at least three forms, encoding its majority proteins RDZ⁶⁰, RDZ⁹⁰, and RDZ⁴⁰. Two significantly less-abundant transcripts are also created, encoding RDZ¹²⁰ and RDZ⁷⁰. At the protein level, the different isoforms are differentially localized. RDZ⁶⁰, RDZ⁹⁰, RDZ⁴⁰, RDZ⁷⁰ only accumulate in the cortical granules, whereas RDV¹²⁰ is found in the vitelline layer

(Wong and Wessel, 2006a). After fertilization, these segregated siblings reunite within the fertilization envelope, likely via heterologous CUB interactions.

Four major enzymatic activities are essential for the proper assembly of the sea urchin fertilization envelope: proteolysis, transamidation, hydrogen peroxide synthesis, and peroxidase-dependent dityrosine crosslinking. Serine protease activity from CGSP1 (cortical granule serine protease) is the only detectable class of protease activity of the cortical granules necessary for the formation of the fertilization envelope (Carroll and Epel, 1975; Haley and Wessel, 1999; Vacquier *et al.*, 1972). Full-length CGSP1 is enzymatically quiescent in the cortical granules, inactive at pH6.5 or below. Exposure of the protease to the pH of the seawater (pH8) at exocytosis immediately activates the protease through autocatalysis (Haley and Wessel, 2004b). CGSP1 cleaves a subpopulation of the granule content proteins, such as the enzyme ovoperoxidase to limit its activity and the β -1,3 glucanase to increase its activity. Another substrate targeted by CGSP1 is p160, a protein thought to link the vitelline layer to the plasma membrane (Haley and Wessel, 2004a). At fertilization, p160 cleavage allows for the separation of the fertilization envelope from the fertilized egg.

Transamidation is mediated by transglutaminases that crosslink glutamine and lysine residues to form N-epsilon (gamma glutamyl) lysyl isopeptide bonds (Greenberg *et al.*, 1991). Two transglutaminases were found in the *Strongylocentrotus purpuratus* genome (Wong and Wessel, 2009). These two isoforms, derived from different genes, are differentially localized and were described as the extracellular transglutaminase (eTG), and the nuclear transglutaminase (nTG). Both transcripts are expressed in the oocyte. Whereas eTG mRNA persists in eggs, nTG mRNA is largely degraded during meiotic maturation (Wong and Wessel, 2009). These transglutaminases are activated by local acidification and act on fertilization envelope proteins such as SFE9, rendezvin, and ovoperoxidase.

Hydrogen peroxide is quickly synthesized at fertilization for ovoperoxidase cross-linking activity, and is synthesized by the dual oxidase homolog, Udx1 in the classically described

respiratory burst (Warburg 1926). This calcium-dependent, pH sensitive enzyme is essential for completing the physical block to polyspermy (Wong *et al.*, 2004). Unlike genes utilized exclusively for the formation of the fertilization envelope and expressed exclusively during oogenesis, such as the structural matrix proteins SFE1, SFE9, proteoliasin, rendezvin, and the enzyme ovoperoxidase, (Wessel *et al.*, 2001; Wong *et al.*, 2004), *Udx1* transcripts are present in eggs and later in development (Wong *et al.*, 2004). Interestingly, *Udx1* also plays a role in the early development as its specific inhibition induces a delay in cytokinesis (Wong and Wessel, 2005). In the egg, this hydrogen peroxide synthesis is necessary for the activity of the ovoperoxidase, a tyrosine crosslinking enzyme derived from the egg cortical granules (Foerder and Shapiro, 1977; LaFleur *et al.*, 1998). In the sea urchin *S. purpuratus*, the ovoperoxidase mRNA is present exclusively in oocytes and is turned over rapidly following germinal vesicle breakdown (LaFleur *et al.*, 1998). Under normal conditions, ovoperoxidase is specifically targeted to the FE via a calcium-dependent interaction with proteoliasin (Weidman *et al.*, 1987). The ovoperoxidase activity is sensitive to transglutaminase (Wong and Wessel, 2009), CGSP1 (Haley and Wessel, 2004b), and *Udx1* (Wong *et al.*, 2004). Semi-*in vivo* crosslinking assay identifies four major targets of ovoperoxidase (Wong and Wessel, 2008): RDZ¹²⁰, proteoliasin, SFE1, and SFE9.

The vast majority of what is known about the fertilization envelope is from the study of a few sea urchin species, yet similar fertilization envelopes are utilized by other echinoderms. Here we explore the proteome of the fertilization envelope in sea stars, and compare its sequences to those in the pencil urchin, thought to be reflective of the ancient sea urchins within the fossil record, and to the well-known sea urchins *S. purpuratus* and *Lytechinus variegatus*, for which most work on the cortical granules and fertilization envelopes have been accomplished. The sea star family, the Asteroids, contains an estimated 1,600 species worldwide (Blake, 1989). Their eggs are generally stored in prophase of meiosis I, and spawning activates release of the inducer for meiotic progression, 1-methyl adenine. Upon germinal vesicle breakdown, the oocyte

becomes fertilization-competent, and following sperm-egg fusion, a robust fertilization envelope forms. Many sea stars rely on the fertilization envelope to limit exposure to harmful elements in the marine environment; some species also rely on the envelope to constrain the blastomeres (Dan-Sohkawa, 1976), (Matsunaga *et al.*, 2002). Removal of the fertilization envelope in many sea star species leads to blastomeres dissociating from each other and subsequent death, likely because of the absence of a distinct hyaline layer, an embryonic extracellular matrix found in sea urchins. Here we determine the genes responsible for formation of the fertilization envelope in the sea star *Patiria miniata* (the common batstar) by proteomic, genomic, and functional criteria.

RESULTS

Sea star and sea urchin fertilization envelopes show differential permeability

Fluorophore-conjugated dextrans were used to compare the permeability of the fertilization envelope in the sea urchin *S. purpuratus* (*Sp*) and the sea star *P. miniata* (*Pm*; Fig. 1). Twenty minutes after fertilization, de-jellied zygotes were incubated with two different sized compounds, fluorescein-conjugated 10kDa-dex and rhodamine-conjugated 2000kDa-dex. The permeability of the fertilization envelope in sea urchins is known to be sensitive to 3-aminotriazole (3-AT), which is an inhibitor of ovoperoxidase activity (Showman and Foerder, 1979). We used this reagent to compare the di-tyrosine crosslinking in both species. Only 52% of the fluorescein 10kDa-dex diffused through the fertilization envelope in sea urchin zygotes (Fig. 1A.a), whereas this diffusion increases to 92% in the presence of 3-AT (Fig. 1A.c). In sea stars, the perivitelline level increased to 66% for the 10kD-dex Fig. 1A.e), and addition of the 3-AT increased this diffusion to 81%. (Fig. 1A.g). Sibling zygotes were simultaneously exposed to the 2000kDa-dex. Sea urchin zygotes show a low permeability for this reagent: 1% of the fluorescence was found in the perivitelline space (Fig. 1A.b), whereas the diffusion through the fertilization envelope increased to 51% in presence of 3-AT (Fig. 1A.d). In sea stars, 30% of the rhodamine present in the media was found in the perivitelline space in normal conditions, (Fig.

1A.f) while a 3-AT pre-treatment increased the diffusion to 56% (Fig. 1A.h). Interestingly, the 3-AT increased the permeability of the sea urchin fertilization envelope by 1.8 times for the 10kDa-dex and by 50 times for the 2000kDa-dex, whereas 3-AT increased the diffusion of the 10kDa-dex by 3.1 times, and only increased the diffusion of the 2000kDa-dex by 1.8 in the sea star. Altogether, these results suggest that the fertilization envelope is more permeable in sea stars than in sea urchins, and that the sea star has di-tyrosine activity, which influences the functionality of the envelope. Due to its more porous nature, however, this barrier is only evident with the larger diffusion reagents. The observation that 3-AT significantly increased the permeability of the sea star envelope also demonstrates that the perivitelline space is not in itself restrictive to the diffusion of dyes.

Only three of the five proteins found in the sea urchin fertilization envelope are present in the sea star

Purification of fertilization envelope proteins from the sea urchin *S. purpuratus* resulted in the identification of SFE9, rendezvin, ovoperoxidase, SFE1 and proteoliasin (Wong and Wessel, 2006a). To identify the components of the fertilization envelope in the sea star *P. miniata*, fertilization envelopes were purified and subjected to SDS-PAGE electrophoresis. After Coomassie blue staining, eight main protein bands were visualized and cut out of the gel for mass spectrometry analysis (Fig 2). Three main proteins were identified: SFE9, rendezvin, and proteoliasin. Except for bands 4 and 8, which were identified as SFE9 and rendezvin, respectively, the other bands contained either SFE9 and proteoliasin or SFE9 and rendezvin, or all three proteins together. The combinatorial results suggest that these proteins might be cross-linked. According to the transcriptome data, the molecular weight of SFE9 and rendezvin were predicted to be 101kDa and 201kDa respectively. Proteoliasin was expected at a molecular weight higher than 79kDa. The identification of these proteins in bands with a higher molecular weight than expected supports the hypothesis of crosslinking activity. To address the possibility that some fertilization envelope components might be in low abundance and not visualized by

Coomassie blue staining, another purification of fertilization envelope was performed and subjected to direct *in solution* trypsin digestion before mass spectrometry analysis. The same proteins, previously obtained after *in gel* trypsin digestion, were identified (data not shown). These results indicate that in the sea star *P. miniata*, the fertilization envelope is primarily composed of SFE9, rendezvin, and proteoliasin.

The proteins involved in fertilization envelope formation differs among Echinoderms

The fertilization envelope formation is well described in the sea urchin *S. purpuratus*. To address the evolution of the fertilization envelope within Echinoderms, we considered another sea urchin species, *L. variegatus* (*Lv*), the pencil urchin *Eucidaris tribuloides* (*Et*), and two sea stars: *P. miniata* (*Pm*), and *Asterias forbesi* (*Af*). *S. purpuratus* diverged from *L. variegatus* between 30 and 50 million years ago (Smith *et al.*, 2006). Sea urchins and pencil urchins diverged around 250 million years ago (Smith *et al.*, 2006). Sea urchins and sea stars diverged approximately 500 million years ago (Hinman *et al.*, 2003). The transcriptomes of *Lv*, *Et*, *Pm*, and *Af*, were obtained from ovary (Adrian Reich, unpublished data). We first looked for the transcripts encoding the three proteins found in both *Sp* and *Pm* fertilization envelopes. SFE9, proteoliasin, and rendezvin were present in all five species (Fig. 3 and Supplemental Table 2). We found that, in sea star, the permeability of the fertilization envelope for the high molecular weight is sensitive to 3-AT (Fig. 1), but an ovoperoxidase ortholog was not detected (Fig. 2). Interestingly, the transcript encoding an ovoperoxidase was found in both sea stars, as well as in the sea urchin *Lv* and in the pencil urchin transcriptomes (Fig. 3). Altogether, these results indicate that some proteins involved in the formation of the fertilization — SFE9, proteoliasin, and rendezvin — are conserved among sea urchins, pencil urchin, and sea stars.

***Rendezvin*, *SFE9*, and *proteoliasin* transcripts are specifically expressed during the early oogenesis**

To determine when the genes that encode the major fertilization envelope proteins are active in sea stars, *rendezvin*, *SFE9*, and *proteoliasin* mRNA probes were synthesized for *in situ*

hybridization (Fig. 4). A probe against *neomycin* was used as a negative control. The overall results show similar mRNA accumulation profiles for *rendezvin*, *SFE9*, and *proteoliasin*. The mRNAs accumulate uniformly throughout the oocyte, and at highest levels in young oocytes. Interestingly, the transcript levels are barely detectable in the full-grown, immature oocytes and in embryonic stages. Quantitative PCR was used to measure the relative RNA levels of *Pm-SFE9*, *proteoliasin*, and *rendezvin* in young oocytes (100 μ m diameter), full-grown immature oocytes, mature oocytes, and fertilized eggs (Fig. 5). All values were normalized against 18S RNA and the corresponding Ct values are presented in Supplemental Table 3. These qPCR data confirm the RNA expression results obtained by *in situ* hybridization. For the three transcripts, the level of mRNA decreases during later oogenesis, reaching its lowest level in the full-grown immature oocytes, mature oocytes, and fertilized eggs. These results indicate that the transcript level of the proteins found in the sea star fertilization envelope uniformly accumulate in the early oocytes.

Cortical granules translocate to the cell periphery during early oogenesis

A polyclonal antibody was generated against *Pm-SFE9*, and was used to determine the pattern of synthesis, location and fate of the major fertilization envelope proteins after fertilization. The transcript found in the *Pm* transcriptome contains 2775 nucleotides, leading to a protein sequence of 924 amino acids, with an expected size of 101kDa. The antibody was first tested by immunoblot on purified fertilization envelopes (Fig. 6). One protein was detected at the estimated molecular weight of 365kDa (arrow), a higher relative size than predicted by primary sequence alone. This difference in molecular weight could be explained by the crosslinking of SFE9 to other proteins present in the fertilization envelope, as was also found in the sea urchin (Wong and Wessel, 2008) and/or by post-translational modifications such as glycosylation. The preimmune serum did not detect this band, and demonstrates the specificity of the antiserum.

By immunofluorescence, *Pm-SFE9* was detected in the cells from early oocytes to fertilized eggs. During oogenesis, especially in 110 μ m oocytes (Fig. 7b) to mature oocytes (Fig. 7e), SFE9 is highly enriched at the periphery of the cytoplasm. Consistent with other cortical

granule content proteins, SFE9 was exocytosed at fertilization and it incorporated into the fertilization envelope (Fig. 7 and Supplemental Fig. 2). The preimmune serum did not detect any fluorescence using the same conditions (Supplemental Fig. 1). Moreover, *Pm*-SFE9 antibody did not detect any signal in embryos post-hatching, indicating its specificity for construction of the fertilization envelope (Supplemental Fig. 3). The cortical granules of young oocytes smaller than 100 μ m (Fig. 7a and Fig. 8) were distributed throughout the entire cytoplasm, and this signal became restricted to the cortex in oocytes larger than 100 μ m. These results suggest that the major period of cortical granule protein synthesis and cortical granule construction occurs early in oogenesis. After image quantification, we found that young oocytes, smaller than 100 μ m, contain approximately 11,400 \pm 2519 (n=5) cortical granules per oocyte. Electron microscopic immunolabeling shows the ultrastructure of the cortical granules of this species, and that the majority of immunolabeling is associated with the electron dense substructures of the granules (Fig. 9).

DISCUSSION

In the sea urchin *S. purpuratus*, cortical granules accumulate throughout the cytoplasm until germinal vesicle breakdown, and then translocate to the cell periphery (Wessel, 1995) and (Laidlaw and Wessel, 1994). Our results suggest that, in contrast to the sea urchin, sea star cortical granules translocate to the cortex as they are synthesized. This early translocation seems more similar to the mechanism described in mice, in which the density of cortical granules present in the cortex increases continually during oogenesis (Ducibella *et al.*, 1994). The production and migration of sea star cortical granules are continuous processes. Since the cortical granules are already at the oocyte surface prior to meiosis, what happens to them during meiosis, especially during the formation of the polar bodies and meiotic spindles? Does the meiotic spindle displace the cortical granules prior to polar body formation, or do they exocytose prematurely, as in mice? We found no evidence of precocious fertilization envelope formation in

the sea star in the area of the meiotic spindle and polar body so we anticipate a cortical granule displacement is made at meiosis.

Cortical granules were previously analyzed in the sea star *Pisaster ochraceus* (Reimer and Crawford, 1995). Using a monoclonal antibody against a 120kDa protein, it was shown that in immature oocytes, cortical granules were concentrated in the periphery of the cytoplasm, but were also found throughout the cytoplasm. In mature oocytes, a larger number of granules were located at the periphery of the cytoplasm, but some granules were still present throughout the cytoplasm. After fertilization, the staining was predominantly found in the perivitelline space, although several brightly stained granules remained in the cell cytoplasm. Later in development at the blastula, the fertilization envelope was not stained by this monoclonal antibody, but labeled granules were present in blastomeres (Reimer and Crawford, 1995). Thus, it is not clear how selectively this antibody identifies cortical granules, or whether it includes recognition of other secretory organelles. To follow the cortical granule biogenesis in the sea star *P. miniata*, we used an antibody against SFE9 to learn that in this species they move to the cell periphery during early oogenesis. The contrasting results between these two species might be explained by the different target proteins studied as well as biological trafficking of different proteins. This may also be simply a matter of species difference in the cortical granule strategy. The granules found at the oocyte periphery might contain both SFE9 and the 120-kDa protein, whereas the granules persisting in the *P. ochraceus* embryos might contain only the 120-kDa protein and/or could play a different role in the development, such as the deposition of a more general extracellular matrix protein (Wong and Wessel, 2006a), e.g. decapod oocyte granules.

Although an ovoperoxidase protein was not directly captured during proteomic analysis, we have lines of evidence to suggest that it is present. First, we found the sequence encoding the ovoperoxidase enzyme within each oocyte transcriptome from five echinoderm species analyzed, including two sea star species. Second, we indirectly observed its enzymatic activity: 3-AT is a specific inhibitor of ovoperoxidase, a myeloperoxidase-type enzyme (Daiyasu and Toh, 2000),

and exposure to 3-AT resulted in significantly increased dextran diffusion in the sea star, similar to that documented in the sea urchin (Wong and Wessel, 2008; see Fig. 1). Thus, we believe ovoperoxidase is one of the conserved fertilization envelope proteins of echinoderms, although unlike in sea urchins, in sea stars, its abundance may be limiting or it may diffuse away from the structure when its crosslinking activity is complete.

In *P. miniata*, the transcripts encoding the major cortical granule proteins (SFE9, proteoliasin and rendezvin) are synchronously regulated. Their RNA is highly expressed in early oocytes, and is rapidly lost in later oogenesis. The timing of this degradation coincides with the translocation of the majority of cortical granules to the cell periphery. In sea urchins, the RNA levels of most of fertilization envelope protein transcripts also decrease during oocyte maturation, particularly when the cortical granules move to the cell periphery (Laidlaw and Wessel, 1994). These two events occur at different phases of sea urchin and sea star oogenesis, but the parallels in relative timing suggest a common mechanism linking the reduction in RNA with the translocation of the cortical granules.

This observation opens two important considerations: Are mRNAs degraded by a shared mechanism, such as miRNAs or specific 3'UTR degradation elements or are the genes regulated by the same transcription factors to synchronize timing and protein stoichiometry? Further, cortical granule mRNA degradation begins as the oocytes rapidly increase in size, a phenomenon consistent with vitellogenesis. In echinoderms, the vitellogenin appears to be made in the digestive tract of the adult and is transported to the ovary where it is taken up into yolk granules (Brooks and Wessel, 2003). That uptake begins with a vitellogenic phase of oogenesis, a transitional period in development of this cell. Although we do not know how this transition is activated, this period may include a transition that involves reallocation of energy and resources, repressing cortical granule assembly, and the associated expression of genes that encode their content, in favor of processes that will enhance embryo viability.

Our results demonstrate that the fertilization envelope in sea urchins is a much more selective barrier than in the sea star. Three similar structural proteins rich in LDLrA repeats: proteoliainin, SFE9, SFE1 compose the fertilization envelope in sea urchin, but no SFE1 ortholog was identified in the sea star. This suggests that SFE1 is not required to form a fertilization envelope, but might be key to efficient packing of the envelope proteins to reduce permeability. SFE1 may have appeared in sea urchins by duplication of SFE9 or proteoliainin to multiply the level of LDLrA rich proteins in the fertilization envelope and thus to form a structure more efficiently protective of the egg. One way to increase protein levels of an envelope protein may be to duplicate a gene and regulate its expression in a manner similar to other genes encoding envelope proteins. This may have been an evolutionary transition that occurred between sea stars and sea urchins. Yet, the diversification in sequence motifs between the two species was otherwise minimal. Perhaps this conservation is a result of compatibility within the complex – several proteins must rapidly and effectively self-assemble, and if they are delayed or compromised in their sperm-blocking, or pathogen blocking ability, the embryos may rapidly die. Even though separated from a common ancestor by approximately 0.5 billion years, the envelope proteins remained largely similar in terms of composition, motif, and function.

The human renal glomerulus filters particles on the order of <50kDa, somewhat smaller than serum albumin, and this extracellular matrix filter takes a few weeks to develop. In contrast, the *S. purpuratus* fertilization envelope filters materials of <~40kDa (Wong and Wessel, 2008) and takes ~30 seconds to form. Based on the morphology of the fertilization envelope in sea stars, i.e. forms more slowly (several minutes) and is significantly thicker than the sea urchin fertilization envelope, we anticipated it would be relatively impermeant. Remarkably, it was far more permeable, allowing dextrans of 2000kDa in size to diffuse through. Although still an effective barrier to sperm, it is clearly more passive to large molecules. This changes what we think about its role in the sea star environment – nutrients would be far more accessible to the developing embryo, and perhaps the embryo is able to endocytose larger nutrient particles for

growth. On the other hand, it would be less capable of blocking toxins in the environment, perhaps even allowing small viral particles to diffuse through. These embryos may be more susceptible to environmental insults, especially in areas close to human effluents at the exact time in development that is most sensitive to the insulting agents.

MATERIALS AND METHODS

Animals

P. miniata were housed in aquaria with artificial seawater (ASW) at 16°C (Coral Life Scientific Grade Marine Salt; Carson, CA). Gametes were acquired by opening up the animals. Immature and full-grown oocytes were collected in filtered seawater and sperm was collected dry. Oocytes were separated by size using Nytex filters, and size separation was improved by manual sorting under the microscope. To obtain mature oocytes, the full-grown, immature oocytes were incubated for an hour in filtered sea water containing 2 µM 1-methyladenine. After addition of sperm, fertilized eggs were cultured in filtered seawater at 16°C (Wessel *et al.*, 2010).

S. purpuratus were housed in aquaria with artificial seawater (ASW) at 16°C (Coral Life Scientific Grade Marine Salt; Carson, CA). Gametes were acquired by either 0.5M KCl injection or by shaking. Eggs were collected in filtered seawater and sperm was collected dry. To obtain embryos, fertilized eggs were cultured in filtered seawater at 16°C.

Permeability assays

Fertilization envelope permeability was tested by measuring the diffusion of fluorophore-conjugated dextrans into the perivitelline space (Wong and Wessel, 2008). As appropriate, eggs were fertilized in filtered seawater or in filtered seawater containing 1 mM 3-aminotriazole (3-AT) and dejellied by acidic treatment (Foltz *et al.*, 2004). Twenty minutes after fertilization, zygotes were incubated with 5 µM fluorescein dextran 10,000 Daltons, (10-kDa-dex) and 50 nM Rhodamine dextran 2,000,000 Daltons, (2,000-kDa-dex) diluted in filtered sea water with or without 1 mM 3-AT. Ten minutes after exposure, zygotes were imaged for both fluorescein and

rhodamine using a LSM 510 laser scanning confocal microscope (Carl Zeiss, Inc.; Thornwood, NY). Average fluorescence intensity was measured using regions within the perivitelline space or the surrounding media using Metamorph software (Universal Imaging Corporation, Downingtown, PA). For each condition, measurements were made on 10 embryos.

Mass spectrometry analysis

Fertilization envelopes were separated from the cells by manual sorting under a dissecting microscope. Fifteen hundred fertilization envelopes were purified and loaded on a SDS PAGE gel for Coomassie staining. The proteins obtained were processed for *in gel* digestion using the *In gel* tryptic digestion kit (Pierce, Rockford, IL). Three hundred additional fertilization envelopes were purified for *in solution* digestion. Briefly, envelopes were resuspended in 100 mM NH₄HCO₃, pH 8, and denatured for 5 minutes at 95°C. After addition of 20 mM DTT, the solution was incubated at 56°C for 45 minutes. The sample was alkylated during a 30-minute incubation at room temperature with 55mM iodoacetamide. Proteins were digested overnight at 37°C in the presence of 10 ng/μl trypsin. Samples were identified using a Thermo-Finnigan LTQ linear ion trap mass spectrometer.

Phylogenetic analysis

Phylogenetic trees were made using the program PhyML available on the website phylogeny.fr (Dereeper *et al.*, 2008).

Whole mount RNA *in situ* hybridization (WMISH)

Sequences used to make antisense WMISH probes for *Pm-rendezvin*, *Pm-SFE9*, and *Pm-proteoliasin* were amplified from *Pm* ovary cDNA and cloned into pGEM T-Easy (Promega). The corresponding primers are presented in Supplemental Table 1. The pGEM T-Easy plasmids were linearized using either Sall (T7 transcription) or ApaI (SP6 transcription) (Promega; Madison, WI). Antisense, DIG-labeled RNA probes were constructed using a DIG RNA labeling kit (Roche; Indianapolis, IN).

WMISH experiments were performed as described previously (Minokawa *et al.*, 2004), and the alkaline phosphatase reaction was carried out for 1h. A non-specific DIG-labeled RNA probe complementary to *neomycin*, obtained from the pSport 18 (Roche; Indianapolis, IN) was used as a negative control. Samples were imaged on a Zeiss Axiovert 200M microscope equipped with a Zeiss color AxioCam MRc5 camera (Carl Zeiss, Inc.; Thornwood, NY).

Real-time quantitative PCR (QPCR)

RNA was extracted from young (100 μ m diameter), full-grown immature, and mature oocytes, or 30 min after fertilization using the RNeasy Micro Kit (Qiagen; Valencia, CA). cDNA was prepared using the TaqMan [®] Reverse Transcription Reagents kit (Applied Biosystems; Foster City, CA). QPCR was performed on a 7300 Real-Time PCR system (Applied Biosystems; Foster City, CA) with the SYBR Green PCR Master Mix Kit (Applied Biosystems; Foster City, CA). Experiments were run in triplicate, and the data were normalized to 18S RNA levels. The primers used to amplify *Pm-rendezvin*, *Pm-SFE9*, *Pm-proteoliasin* and *18S* are indicated in Supplemental Table 1.

Antibody production

A region of *Pm-SFE9* was cloned using the primers F (5'-CCCAGACCTTGGTATGCAATG-3') and R (5'-CCCAGTCGAGCAATCTCTGTAC-3'). This sequence was inserted in the pNO-TAT vector in frame with a 6xHis tag (Nagahara *et al.*, 1998). Recombinant protein was expressed in BL21 bacteria, purified on a ProBond nickel column (Invitrogen; Carlsbad, CA), and used to raise antiserum in rabbit as previously described (Wong and Wessel, 2004)

Western blot

Western blot analyses were performed following electrophoretic transfer of proteins from SDS-PAGE onto 0.22- μ m nitrocellulose membranes (Towbin *et al.*, 1979). Membranes were incubated with antibodies directed against *Pm-SFE9* (1:1000) in 20 mM Tris-HCl (pH 7.6), 1% BSA, and 0.1% Tween-20, overnight at 4°C. The antigen-antibody complex was measured by

chemiluminescence using horseradish peroxidase-coupled secondary antibodies according to the manufacturer's instructions (ECL; Amersham Pharmacia Biotech). Preimmune serum from the same rabbit was used as a control. Three hundred purified fertilization envelopes were loaded per lane.

Immunofluorescence

Oocytes and embryos were cultured as described above, and samples were collected at indicated stages for whole-mount antibody labeling. Cells were fixed overnight in 4% paraformaldehyde in ASW, washed 3 times with PBS-Tween, and stored at 4°C. Oocytes and embryos were blocked for an hour at room temperature in 4% sheep serum (Sigma; St. Louis, MO) /PBS-Tween (blocking buffer). For labeling, the cells were incubated overnight at 4°C with the anti *Pm*-SFE9 serum diluted 1:1000 in blocking buffer. The preimmune serum, also diluted by 1:1000, was used as a control. The cells were washed 3 times with PBS-Tween, and then incubated with anti-rabbit Alexa Fluor 488 conjugated antibody (Invitrogen), diluted 1:500 in blocking buffer, for two hours at room temperature. Oocytes and embryos were then washed 3 times with PBS-Tween. Pictures were taken on a LSM 510 laser scanning confocal microscope (Carl Zeiss, Inc.; Thornwood, NY). These pictures were used to define the number of cortical granules in young oocytes. Five young oocytes, with an average diameter of 72.1µm, were used for the quantification. The number of cortical granules was manually counted in the optical slice obtained using a pinhole of 0.99. The volume of each optical slice was defined by the formula $V = \pi r^2 h$. An approximate number of cortical granules per µm³ per oocyte analyzed was calculated, which was then multiplied by the volume of the corresponding oocyte to obtain the number of cortical granules per oocyte.

REFERENCES

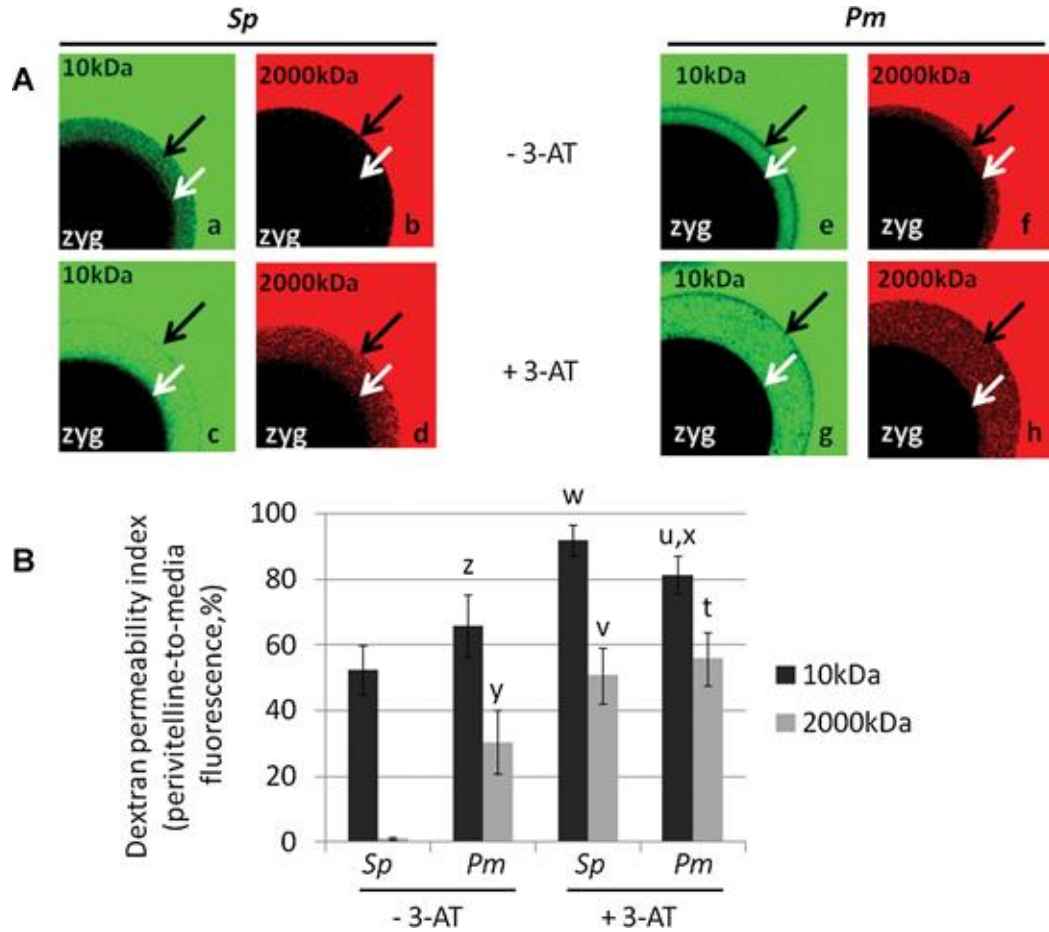
- Blake, D. (1989). Asteroidea: Functional morphology, classification and phylogeny. *Echinoderm studies* 3.
- Briggs, E., and Wessel, G.M. (2006). In the beginning...animal fertilization and sea urchin development. *Dev Biol* 300, 15-26.
- Brooks, J.M., and Wessel, G.M. (2003). Selective transport and packaging of the major yolk protein in the sea urchin. *Dev Biol* 261, 353-370.
- Carroll, E.J., Jr., and Epel, D. (1975). Isolation and biological activity of the proteases released by sea urchin eggs following fertilization. *Dev Biol* 44, 22-32.
- Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674-3676.
- Daiyasu, H., and Toh, H. (2000). Molecular evolution of the myeloperoxidase family. *J Mol Evol* 51, 433-445.
- Dan-Sohkawa, M. (1976). A 'normal' development of denuded eggs of the starfish, *Asterina pectinifera*. *Develop Growth Diff* 18, 439-445.
- Derbes, A.A. (1847). Observations sur le mecanisme et les phenomenes qui accompagnent la formation de l'embryonchez l'oursin comestible. *Ann Sci Nat Zool* 8, 80-98.
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., *et al.* (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36, W465-469.
- Ducibella, T., Duffy, P., and Buetow, J. (1994). Quantification and localization of cortical granules during oogenesis in the mouse. *Biol Reprod* 50, 467-473.
- Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30, 1575-1584.
- Foerder, C.A., and Shapiro, B.M. (1977). Release of ovoperoxidase from sea urchin eggs hardens the fertilization membrane with tyrosine crosslinks. *Proc Natl Acad Sci U S A* 74, 4214-4218.
- Fol, H. (1877). Sur le commencement de l'henogenie chez divers animaux. *Arch Zool Exp Gen T-6*, 145-169.
- Foltz, K.R., Adams, N.L., and Runft, L.L. (2004). Echinoderm eggs and embryos: procurement and culture. *Methods Cell Biol* 74, 39-74.
- Greenberg, C.S., Birckbichler, P.J., and Rice, R.H. (1991). Transglutaminases: multifunctional cross-linking enzymes that stabilize tissues. *FASEB J* 5, 3071-3077.
- Haley, S.A., and Wessel, G.M. (1999). The cortical granule serine protease CGSP1 of the sea urchin, *Strongylocentrotus purpuratus*, is autocatalytic and contains a low-density lipoprotein receptor-like domain. *Dev Biol* 211, 1-10.
- Haley, S.A., and Wessel, G.M. (2004a). Proteolytic cleavage of the cell surface protein p160 is required for detachment of the fertilization envelope in the sea urchin. *Dev Biol* 272, 191-202.
- Haley, S.A., and Wessel, G.M. (2004b). Regulated proteolysis by cortical granule serine protease 1 at fertilization. *Mol Biol Cell* 15, 2084-2092.
- Hinman, V.F., Nguyen, A.T., Cameron, R.A., and Davidson, E.H. (2003). Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. *Proc Natl Acad Sci U S A* 100, 13356-13361.
- LaFleur, G.J., Jr., Horiuchi, Y., and Wessel, G.M. (1998). Sea urchin ovoperoxidase: oocyte-specific member of a heme-dependent peroxidase superfamily that functions in the block to polyspermy. *Mech Dev* 70, 77-89.
- Laidlaw, M., and Wessel, G.M. (1994). Cortical granule biogenesis is active throughout oogenesis in sea urchins. *Development* 120, 1325-1333.

- Lee, H.C., Johnson, C., and Epel, D. (1983). Changes in internal pH associated with initiation of motility and acrosome reaction of sea urchin sperm. *Dev Biol* 95, 31-45.
- Matsunaga, M., Uemura, I., Tamura, M., and Nemoto, S.I. (2002). Role of specialized microvilli and the fertilization envelope in the spatial positioning of blastomeres in early development of embryos of the starfish *Astropecten scoparius*. *Biological Bulletin* 202, 213-222.
- Minokawa, T., Rast, J.P., Arenas-Mena, C., Franco, C.B., and Davidson, E.H. (2004). Expression patterns of four different regulatory genes that function during sea urchin development. *Gene Expr Patterns* 4, 449-456.
- Nagahara, H., Vocero-Akbani, A.M., Snyder, E.L., Ho, A., Latham, D.G., Lissy, N.A., Becker-Hapak, M., Ezhevsky, S.A., and Dowdy, S.F. (1998). Transduction of full-length TAT fusion proteins into mammalian cells: TAT-p27Kip1 induces cell migration. *Nat Med* 4, 1449-1452.
- Reimer, C.L., and Crawford, B.J. (1995). Identification and partial characterization of yolk and cortical granule proteins in eggs and embryos of the starfish, *Pisaster ochraceus*. *Dev Biol* 167, 439-457.
- Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086-1092.
- Sea Urchin Genome Sequencing, C., Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess, D.R., *et al.* (2006). The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314, 941-952.
- Showman, R.M., and Foerder, C.A. (1979). Removal of the fertilization membrane of sea urchin embryos employing aminotriazole. *Exp Cell Res* 120, 253-255.
- Smith, A.B., Pisani, D., Mackenzie-Dodds, J.A., Stockley, B., Webster, B.L., and 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.
- Steinhardt, R., Zucker, R., and Schatten, G. (1977). Intracellular calcium release at fertilization in the sea urchin egg. *Dev Biol* 58, 185-196.
- Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A* 76, 4350-4354.
- Vacquier, V.D. (1998). Evolution of gamete recognition proteins. *Science* 281, 1995-1998.
- Vacquier, V.D., Epel, D., and Douglas, L.A. (1972). Sea urchin eggs release protease activity at fertilization. *Nature* 237, 34-36.
- Vacquier, V.D., and Moy, G.W. (1977). Isolation of bindin: the protein responsible for adhesion of sperm to sea urchin eggs. *Proc Natl Acad Sci U S A* 74, 2456-2460.
- Ward, G.E., Brokaw, C.J., Garbers, D.L., and Vacquier, V.D. (1985). Chemotaxis of *Arbacia punctulata* spermatozoa to resact, a peptide from the egg jelly layer. *J Cell Biol* 101, 2324-2329.
- Weidman, P.J., Teller, D.C., and Shapiro, B.M. (1987). Purification and characterization of proteolisin, a coordinating protein in fertilization envelope assembly. *J Biol Chem* 262, 15076-15084.
- Wessel, G.M. (1995). A protein of the sea urchin cortical granules is targeted to the fertilization envelope and contains an LDL-receptor-like motif. *Dev Biol* 167, 388-397.
- Wessel, G.M., Brooks, J.M., Green, E., Haley, S., Voronina, E., Wong, J., Zaydfudim, V., and Conner, S. (2001). The biology of cortical granules. *Int Rev Cytol* 209, 117-206.
- Wessel, G.M., Conner, S., Laidlaw, M., Harrison, J., and LaFleur, G.J., Jr. (2000). SFE1, a constituent of the fertilization envelope in the sea urchin is made by oocytes and contains low-density lipoprotein-receptor-like repeats. *Biol Reprod* 63, 1706-1712.
- Wessel, G.M., Reich, A.M., and Klatsky, P.C. (2010). Use of sea stars to study basic reproductive processes. *Syst Biol Reprod Med* 56, 236-245.
- Wong, J.L., Creton, R., and Wessel, G.M. (2004). The oxidative burst at fertilization is dependent upon activation of the dual oxidase Udx1. *Dev Cell* 7, 801-814.

- Wong, J.L., and Wessel, G.M. (2004). Major components of a sea urchin block to polyspermy are structurally and functionally conserved. *Evol Dev* 6, 134-153.
- Wong, J.L., and Wessel, G.M. (2005). Reactive oxygen species and Udx1 during early sea urchin development. *Dev Biol* 288, 317-333.
- Wong, J.L., and Wessel, G.M. (2006a). Defending the zygote: search for the ancestral animal block to polyspermy. *Curr Top Dev Biol* 72, 1-151.
- Wong, J.L., and Wessel, G.M. (2006b). Rendezvin: An essential gene encoding independent, differentially secreted egg proteins that organize the fertilization envelope proteome after self-association. *Mol Biol Cell* 17, 5241-5252.
- Wong, J.L., and Wessel, G.M. (2008). Free-radical crosslinking of specific proteins alters the function of the egg extracellular matrix at fertilization. *Development* 135, 431-440.
- Wong, J.L., and Wessel, G.M. (2009). Extracellular matrix modifications at fertilization: regulation of dityrosine crosslinking by transamidation. *Development* 136, 1835-1847.

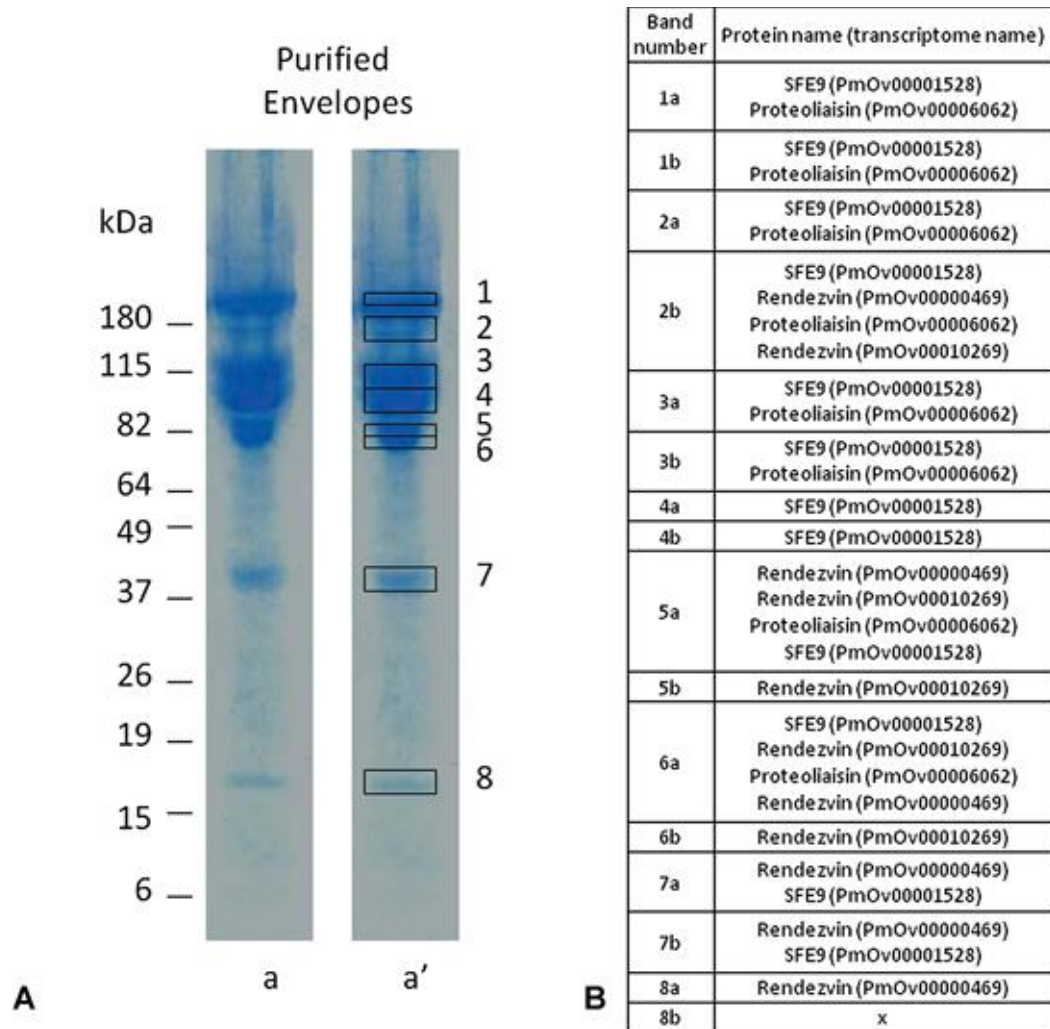
FIGURES

Figure 1. The fertilization envelope is more permeable in sea star than in sea urchin.



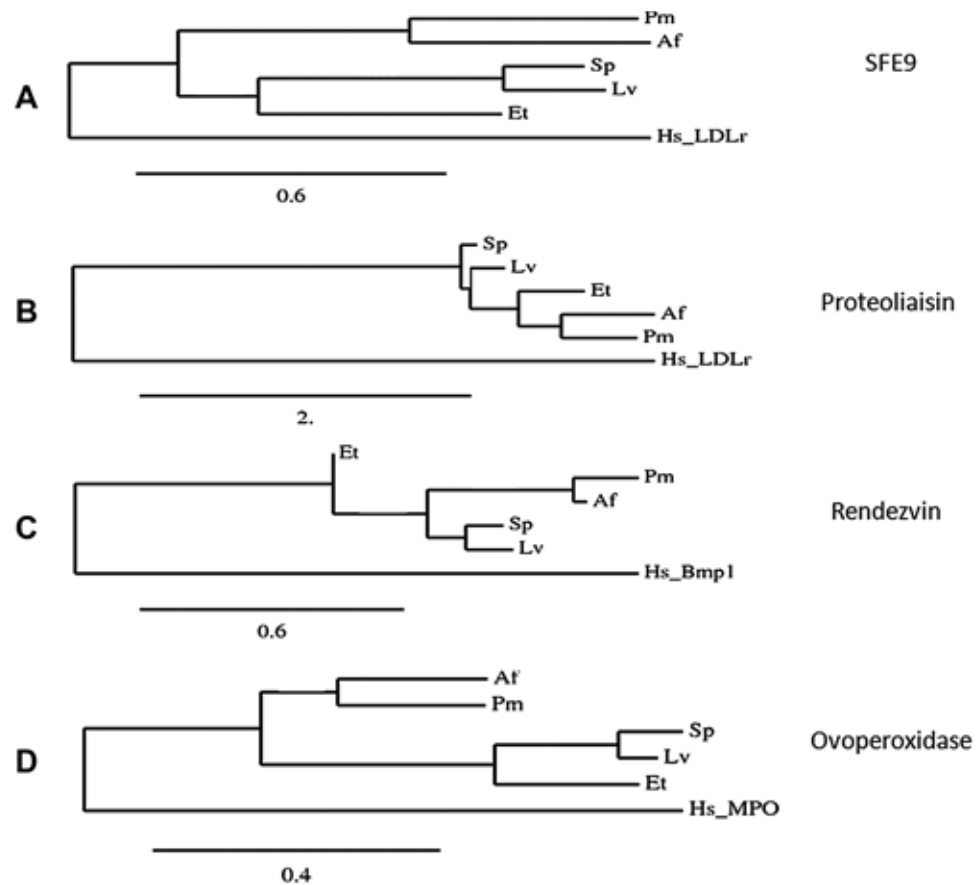
(A) After fertilization, dejellied zygotes from sea urchin *Sp* (a, b, c, d) and sea star *Pm* (e, f, g, h) were simultaneously exposed to a 10-kDa fluorescein dextran (a, c, e, g), and a 2000-kDa rhodamine dextran (b, d, f, h). In both species, the diffusion of the fluorophore dextran into the perivitelline space was analyzed in normal conditions (a, b, e, f) or in presence of 3-AT, (c, d, g, h). Perivitelline space is found between the zygote (zyg, white arrow), and the fertilization envelope (black arrow). (B) Determination of the dextran permeability index, corresponding to the ratio of fluorescence present in the perivitelline space to the media, results are shown as percentages. Ten individuals were measured in each condition. Significance was assessed for each condition between *Sp* and *Pm* using Student's *t*-test $P < 0.05$. Significant differences were obtained in normal condition between *Sp* and *Pm* for the diffusion of the 10-kDa (a) and 2000-kDa (b) dextran, and between *Sp* and *Pm* in the presence of 3-AT for the diffusion of the 10-kDa dextran (c). Significant differences were also obtained for *Sp* (d,e) and *Pm* (f,g) between normal condition without 3-AT and the addition of 3-AT for both conjugated dextrans.

Figure 2. In the sea star *Pm*, the fertilization envelope is composed of three major proteins: SFE9, rendezvin, and proteoliaisin.



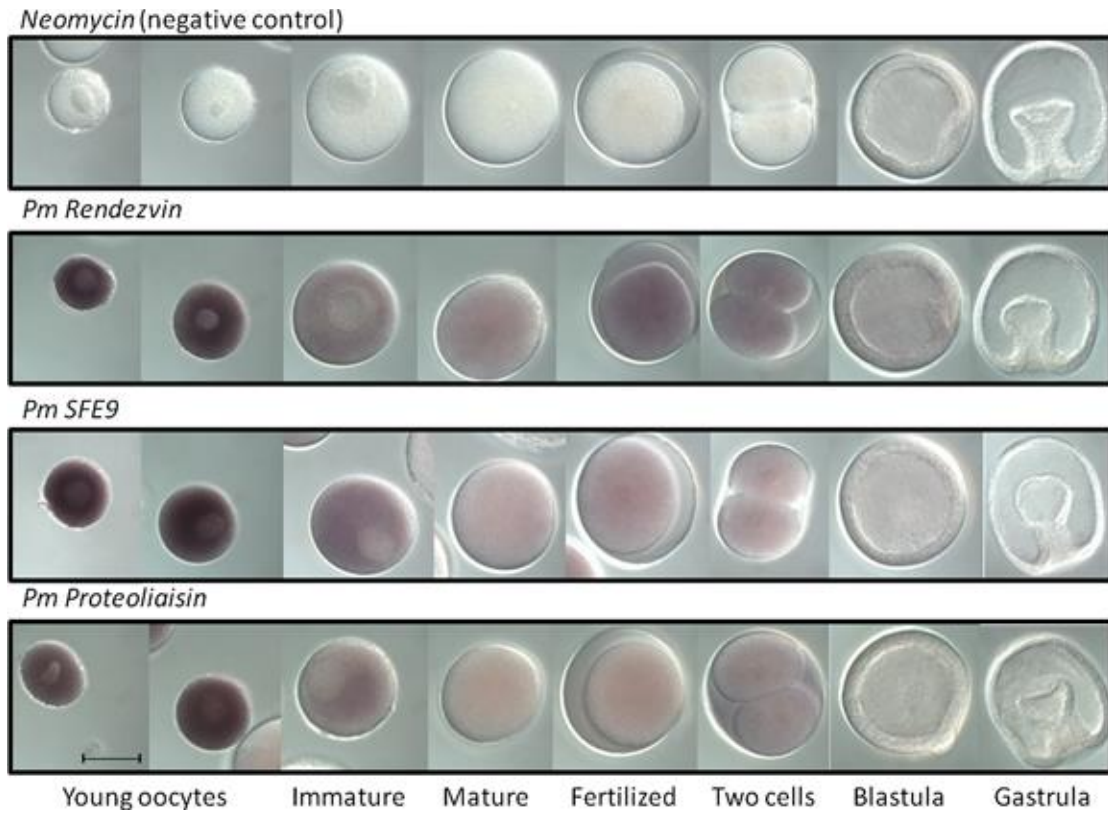
(A) Fertilization envelopes were isolated from zygotes and loaded on a SDS-PAGE gel. After Coomassie blue staining, eight main bands were obtained (a). Each band, represented by a black box (b), was cut out of the gel for mass spectrometry analysis. (B) Results of the mass spectrometry. The purification was done twice using different *Pm* sea stars for the fertilization. The same bands were obtained in both experiments and analyzed by mass spectrometry. The letters a and b following the band numbers correspond to the duplicates for each band. The table lists the name of the protein identified in each band followed by the corresponding transcriptome number found in the *Pm* transcriptome.

Figure 3. Phylogenetic trees representing the proteins involved in the formation of the fertilization envelope.



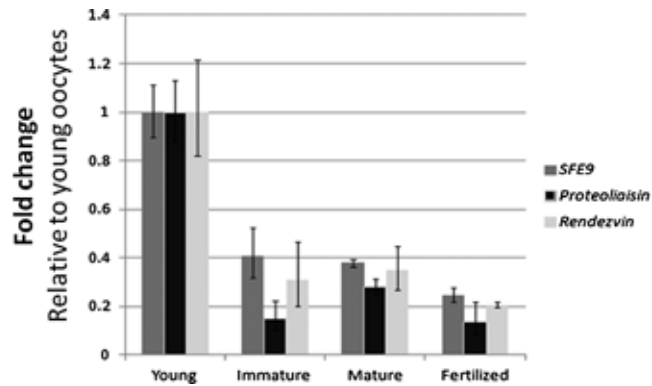
The amino acid sequences of SFE9, proteoliasin, rendezvin and ovoperoxidase, identified in the five species: *S. purpuratus* (Sp), *L. variegatus* (Lv), *E. tribuloides* (Et), *P. miniata* (Pm), *A. forbesi* (Af); were used to design phylogenetic trees. (A) Tree for SFE9. Homo sapiens LDLr (NCBI: NP000518) was used as an outgroup. (B) Tree for proteoliasin. Homo sapiens LDLr (NCBI: NP000518) was used as an outgroup. (C) Tree for rendezvin. Homo sapiens Bmp1 (NCBI: AAI01764) was used as an outgroup. (D) Tree for ovoperoxidase. Homo sapiens proteins myeloperoxidase, MPO (NCBI: AAA59863) was used as an outgroup. For each tree, the scale bar indicates an evolutionary distance of amino acid substitutions per position.

Figure 4. *Pm rendezvin*, *SFE9*, and *proteolaisin* mRNAs are highly and uniformly expressed during early oogenesis.



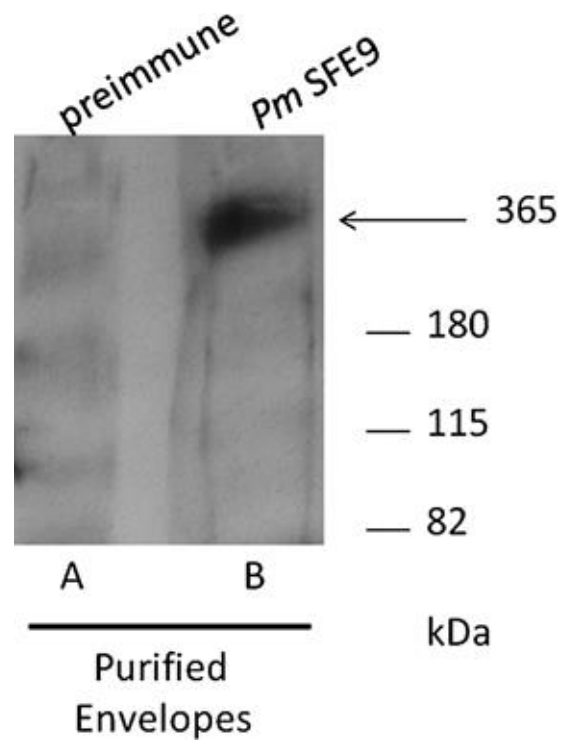
Whole mount *in situ* hybridization in the sea star *P. miniata*, using probes against *Pm-rendezvin*, *SFE9*, and *proteolaisin*, in young oocytes, immature full-grown oocytes, mature oocytes, 30 minutes after fertilization, at the two-cell stage, in blastula, and in gastrula. *Neomycin* is used as a negative control.

Figure 5. *Pm-SFE9*, *proteoliasin*, and *rendezvin* RNA levels decrease during oogenesis.



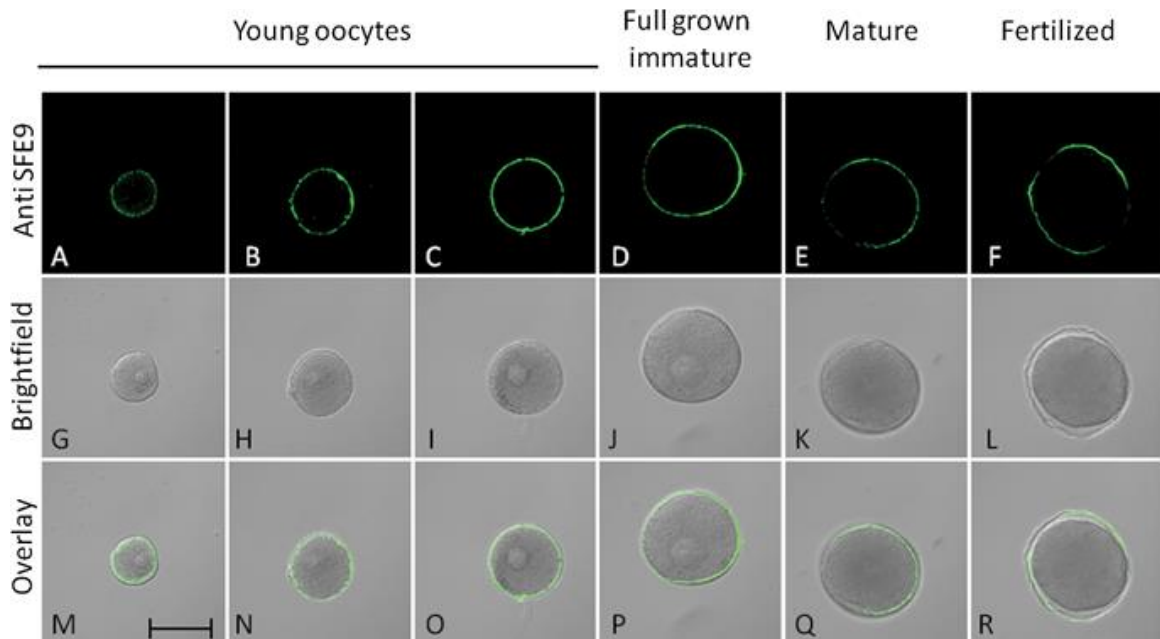
QPCR was used to measure the RNA levels of *Pm-SFE9*, *proteoliasin*, and *rendezvin*, at the indicated developmental stages: young oocytes (100 μm diameter), full-grown immature oocytes, mature oocytes, and 30 minutes after fertilization. All values were normalized against 18S RNA and represented as a fold-change relative the amount of RNA present in the young oocytes. Significance was assessed for each transcript between young oocytes and each other developmental stage using Student's *t*-test, $P < 0.05$. Significant differences were obtained between the young oocytes and all the other stages: immature oocytes, mature oocytes and fertilized eggs for each transcript: *SFE9* (a), *proteoliasin* (b), and *rendezvin* (c).

Figure 6. *Pm*-SFE9 antibody specifically recognizes one high molecular weight bands.



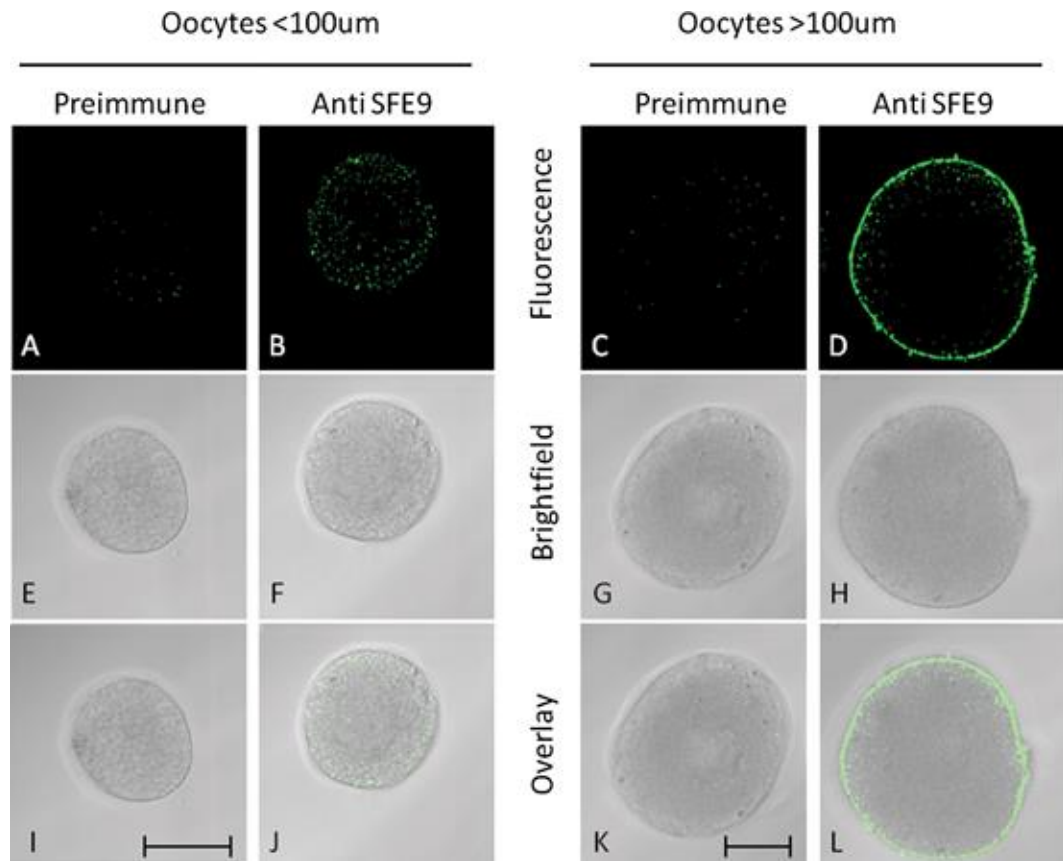
Western blot using the antiserum raised against *Pm*-SFE9 (B) or its preimmune serum (A) on proteins obtained from purified fertilization envelopes (A,B). 300 fertilization envelopes were loaded in each well.

Figure 7. The protein *Pm*-SFE9 is present throughout oogenesis, maturation, and fertilization.



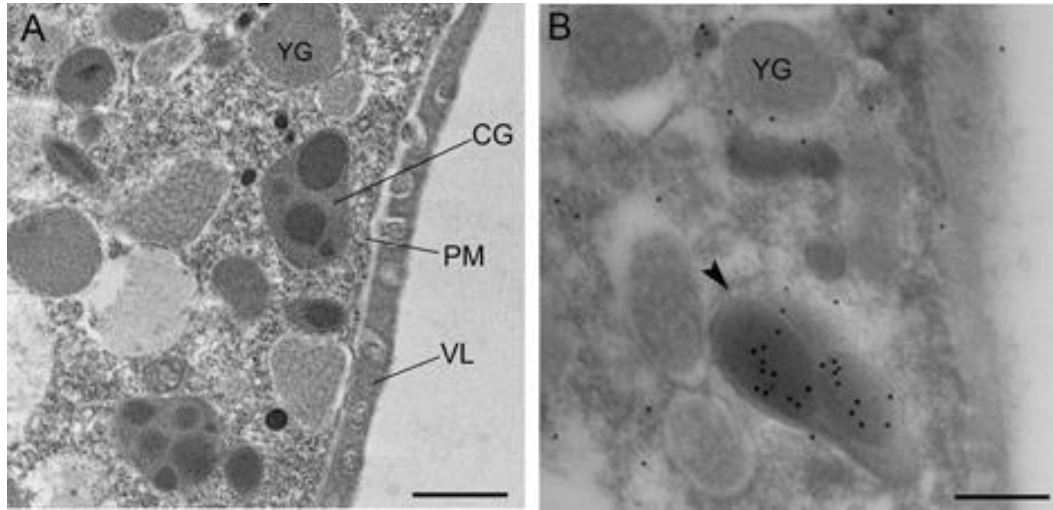
Immunofluorescence using the antiserum against *Pm*-SFE9 (A,B,C,D,E,F) on 85- μ m diameter oocytes (A), 110- μ m diameter oocytes (B), 130- μ m diameter oocytes (C), full-grown oocytes (D), mature oocytes (E), and fertilized eggs (F). The corresponding differential interference contrast images are respectively shown in G to L. For each developmental stage, the overlay of the fluorescence and the DIC image is represented in M to R. Scale bar, 100 μ m. Pictures were taken using the same microscope settings (laser intensity, pin-hole opening) at 200x magnification.

Figure 8. Sea star cortical granules move to the periphery of the cell during early oogenesis.



Immunofluorescence using the antiserum against *Pm*-SFE9 (B,D) or its preimmune serum (A,C) on oocytes having a diameter smaller (A,B) or greater than 100 μm (C,D). The corresponding differential interference contrast images are respectively shown from E to H. For each developmental stage, the overlay of the fluorescence and the DIC image is represented in I to L. Scale bar = 100 μm . Pictures were taken using the same microscope settings (laser intensity, pin-hole opening) at 400x magnification.

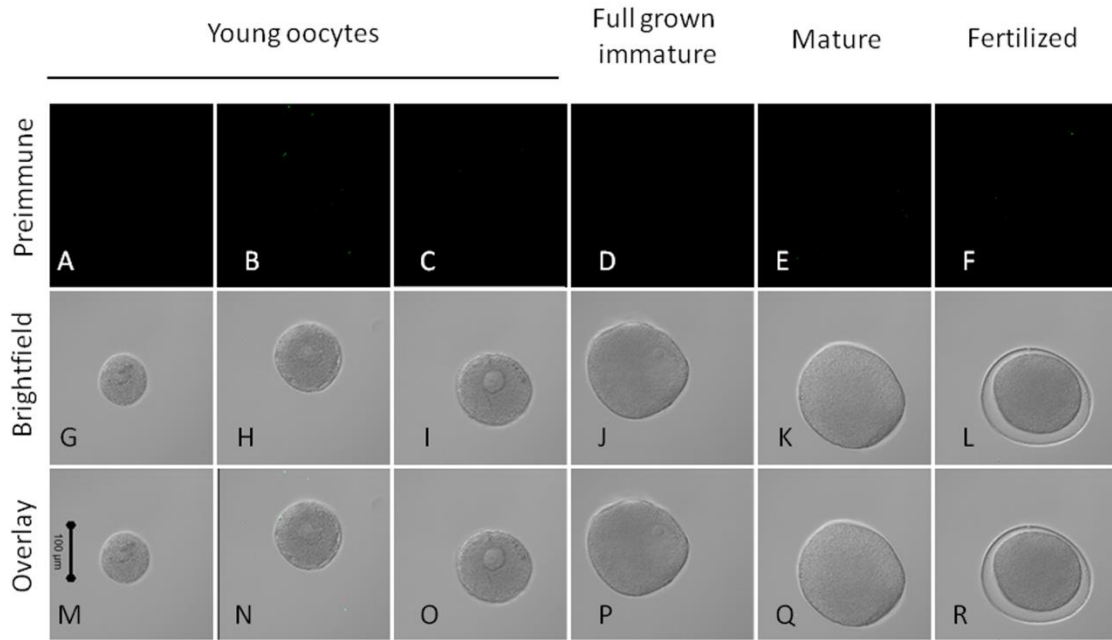
Figure 9. The cortex of immature oocytes and ultrastructural immunolocalization of SFE9 in cortical granules.



(A) Electron micrograph of the cortex of a full grown immature oocyte. Cortical granules (CG), vitelline layer (VL), plasma membrane (PM), yolk granule (YG). **(B)** Immunogold electron microscopy showing SFE9 accumulated in the cortical granules (arrowhead) in immature oocytes. Scale bar, 0.5 μm .

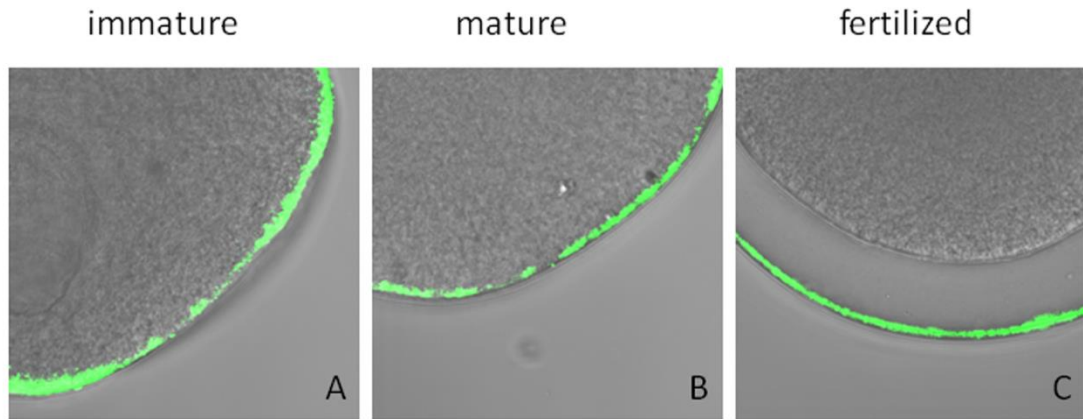
SUPPLEMENTAL INFORMATION

Supplemental Figure 1. *Pm*-SFE9 antibody specificity.



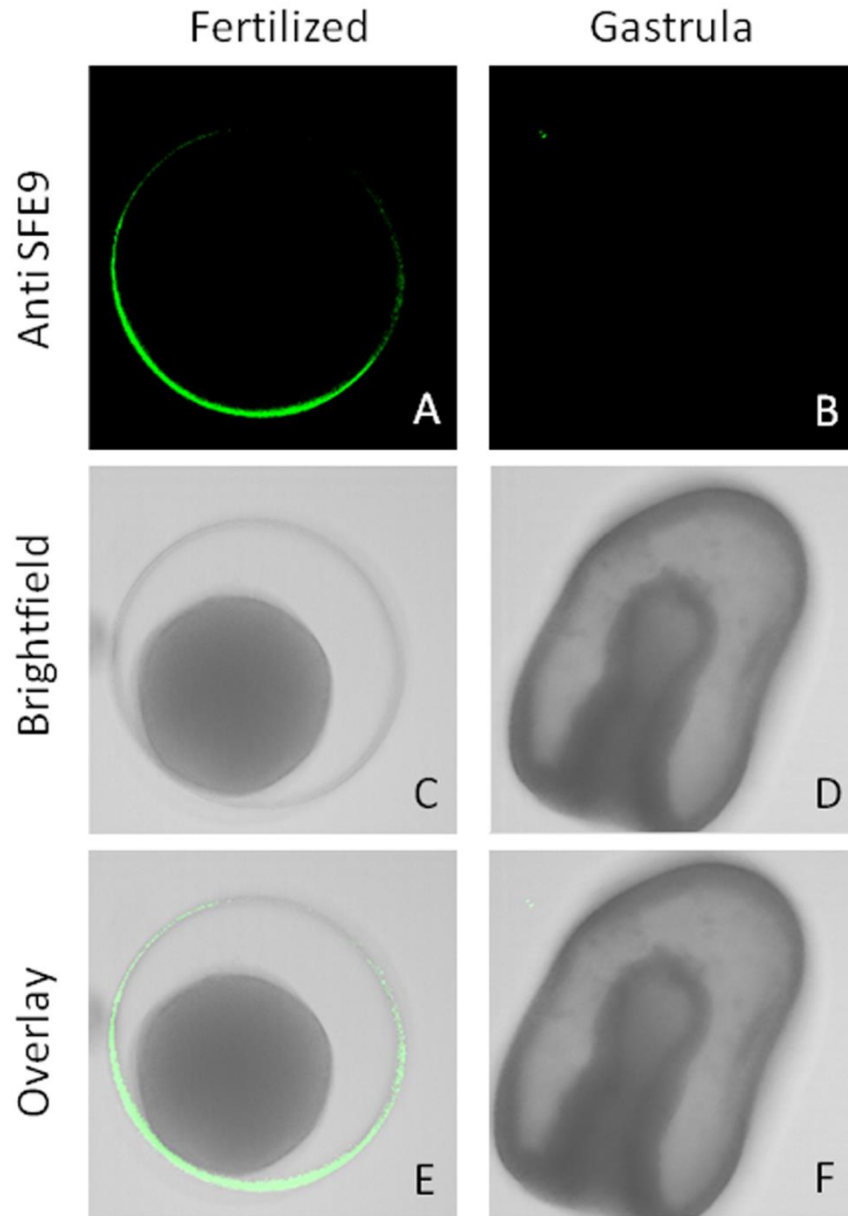
Immunofluorescence using the SFE9 preimmune serum (A,B,C,D,E,F) on young oocytes (A,B,C), full-grown immature oocytes (D), mature oocytes (E), and fertilized eggs (F). The corresponding differential interference contrast images are respectively shown in G to L. For each developmental stage, the overlay of the fluorescence and the DIC image is represented in M to R. Scale bar, 100 μ m. Pictures were taken using the same microscope settings as Figure 9 (laser intensity, pin-hole opening) at 200x magnification.

Supplemental Figure 2. After fertilization, *Pm*-SFE9 is incorporated in the fertilization envelope.



Immunofluorescence using the *Pm*-SFE9 antibody on immature (A) and mature oocytes (B), or fertilized eggs (C). For each developmental stage, the overlay of the fluorescence and the DIC image is represented. Pictures were taken using the same microscope settings (laser intensity, pin-hole opening) at 400x magnification.

Supplemental Figure 3. *Pm*-SFE9 antibody specifically labels the early developmental stage.



Immunofluorescence using the *Pm*-SFE9 antibody on fertilized eggs (A,C,E) and gastrula stage (B,D,F). For each developmental stage, the overlay of the fluorescence (E,F) and the DIC image (C,D) is represented. Pictures were taken using the same microscope settings (laser intensity, pin-hole opening) at 200x magnification.

Supplemental Table 1. Primers used to analyze the expression of the transcripts *Pm-SFE9*, *rendezvin* and *proteoliaisin*.

| <i>in situ</i> primers | Forward primer (5'→ 3') | Reverse Primer (5'→ 3') |
|-----------------------------------|--|---------------------------------------|
| <i>Pm</i> rendezvin | CCTCTGGCTGTGCGTGGCAGTTG TCTCGG | GGCGAACCCCTTTGTACGGGTCG TAAGCATTG |
| <i>Pm</i> SFE9 | CGAATGAGTTCCAATGCAACGA TAGCAG | CGCACATGAATTGACCGACAAG ACATCC |
| <i>Pm</i> proteoliaisin | TGTGATCATACTGCTAGCTTCGG TGCCTGGCGCTTC | CTGCTCAAACACTTGCCGGTATC GCAGCGGAAC |

| qPCR primers | Forward primer (5'→ 3') | Reverse Primer (5'→ 3') |
|----------------------------|-------------------------|-------------------------|
| <i>Pm</i> rendezvin | TGTGTGCAAGACCCATCAAT | CTGGTGTAGGTCCGTTTACA |
| <i>Pm</i> SFE9 | GTGTTTGTGCGAGCGAGTTCA | CATCGCAGACTTGAGTCGAA |
| <i>Pm</i> proteoliaisin | AGCAGGCTCAACAGGTCCT | TCGCCATTCTCACATTCGTA |
| <i>Pm</i> 18S | TTGGAGTGTTCAAAGCAGGC | TATCTGATCGCCTTCGAACC |

Primers used for whole mount *in situ* hybridization and qPCR.

Supplemental Table 2. Transcripts encoding for proteins involved in the formation of the fertilization envelope in sea urchins, pencil urchin, and sea stars.

| | |
|-------------------------|-------------------------|
| <i>Sp</i> SFE9 | U17377 (NCBI) |
| <i>Lv</i> SFE9 | AAT01144 (NCBI) |
| <i>Pm</i> SFE9 | Pm_1528_Trans_2 |
| <i>Af</i> SFE9 | Af_255_Trans_26 |
| <i>Et</i> SFE9 | Et_2382_Trans_4 |
| | |
| <i>Sp</i> proteoliasin | AAT01141 (NCBI) |
| <i>Lv</i> proteoliasin | AAT01142 (NCBI) |
| <i>Pm</i> proteoliasin | Pm_6062_Trans_5 |
| <i>Af</i> proteoliasin | Af_687_Trans_9 |
| <i>Et</i> proteoliasin | Et_34652_Trans_2 |
| | |
| <i>Sp</i> rendezvin | ABK35135 (NCBI) |
| <i>Lv</i> rendezvin | ABK35136 (NCBI) |
| <i>Pm</i> rendezvin | Pm0000469 and Pm0010269 |
| <i>Af</i> rendezvin | Af_487_Trans_31 |
| <i>Et</i> rendezvin | Et_1214_Trans_5 |
| | |
| <i>Sp</i> ovoperoxidase | NP_999755 (NCBI) |
| <i>Lv</i> ovoperoxidase | AAB92243 (NCBI) |
| <i>Pm</i> ovoperoxidase | Pm_8109_Trans_1 |
| <i>Af</i> ovoperoxidase | Af_1282_Trans_7 |
| <i>Et</i> ovoperoxidase | Et_826_Trans_5 |

For each transcript, this table indicates the corresponding identification in the transcriptome data. Some of the sequences used in *Sp* and *Lv* were obtained from NCBI; their accession number is given.

Supplemental Table 3. *Pm-SFE9*, *proteoliasin*, and *rendezvin* mRNA are highly expressed in young oocytes.

| | 18S | 18S | SFE9 | Proteoliasin | Rendezvin |
|------------|------|------|------|--------------|-----------|
| Young | 15.6 | 16.3 | 27.3 | 28.3 | 28.9 |
| Immature | 16.0 | 16.6 | 29.0 | 32.3 | 31.0 |
| Mature | 15.4 | 16.0 | 28.5 | 30.8 | 30.1 |
| Fertilized | 15.4 | 16.1 | 29.2 | 31.9 | 31.2 |

Ct values obtained by qPCR for the transcripts 18S, SFE9, proteoliasin and rendezvin, measured in young, immature, mature oocytes, and fertilized eggs.

Chapter V: Synthesis and future directions

Adrian Reich

INTRODUCTION

The phylogenetic relationships between the different clades of echinoderms has been contentious for over 100 years (Bather, 1900; Mac Bride, 1906) and the debate has continued through the present day, even though many more techniques and data are available (Janies *et al.*, 2011; Pisani *et al.*, 2012). Using the first well-supported echinoderm tree built from transcriptome based phylogenetic methods with representatives from all five clades, we can identify important evolutionary transitions within the phylum; while previously, the conclusions had the caveat of changing depending how the tree resolved. My work has focused on identifying a well resolved phylogenetic tree of echinoderms and using those results to elucidate evolutionary transitions within Echinodermata; with a particular emphasis on the differences between Asterozoa (sea stars) and Echinozoa (e.g. sea urchins).

RESULTS

Foremost, through the assembly and analysis of nearly two dozen *de novo* transcriptomes throughout Echinodermata, I have constructed a well-supported phylogenetic relationship with members of all five classes of extant echinoderms (Chapter II, Fig. 2). My results support the Asterozoan hypothesis of echinoderm evolution (Chapter II, Fig. 1), specifically that the sister clades Ophurozoa and Asterozoa form Asterozoa which is in turn sister to Echinozoa; comprised of the sister clades of Holothurozoa and Echinozoa. In addition to the thousands of molecular changes that I have documented in assembling the tree, morphological and embryological changes in the phylum can now be attached to specific branches with a high degree of confidence (Chapter II, Fig. 3).

Second, building upon these results, we used the transcriptome of the sea star *Patiria miniata* to identify and localize homologous sequences that are conserved in: germline determination (Chapter III, Fig. 2), germline induction (Chapter III, Fig. 3), germline association (Chapter III, Fig. 4), left/right asymmetry (Chapter III, Fig. 5), and factors important in

morphogenesis and embryogenesis (Chapter III, Fig. 6). These data taken together strongly support the conclusion that the germline in *P. miniata* is localized to the posterior enterocoel (PE; Chapter III, Fig. 1). Furthermore, the data also suggest the germline in the sea star is an inductive mechanism as opposed to inherited (Extavour and Akam, 2003).

Third, again by examining the *de novo* transcriptome of the sea star *P. miniata*, we were able to directly compare the fertilization envelopes of *P. miniata* and the sea urchin, *Strongylocentrotus purpuratus* which yielded interesting results. Of the five major components of the *S. purpuratus* fertilization envelope, we found direct evidence for the expression of three members of the fertilization envelope in sea stars (SFE9, rendezvin, and proteoliasin) and indirect evidence for a fourth (ovoperoxidase), however, no evidence was found of SFE1 (Chapter IV, Fig. 3, 4). Perhaps due to the lack of SFE1 in the sea star, the permeability of the fertilization envelope in *P. miniata* was significantly higher than that of *S. purpuratus* (Chapter IV, Fig. 1).

In total, this work facilitates many studies in echinoderms, at the very least large fractions of many transcriptomes are now known, but more significantly the analysis of the evolutionary history of a diverse group of organisms is now greatly ameliorated. Previously, evolutionary comparisons within Echinodermata were limited to only a handful of organisms concentrated in two clades (Echinoidea and Asteroidea), and the vast majority of known sequence within the phylum came from a single species (*S. purpuratus*). *S. purpuratus* has a number of derived features that may not represent echinoderms as a whole (Chapter II and Fig. 1), so having many more sequenced echinoderms available, these combined datasets allow for significant advancement in the field of evolution and development in Bilateria.

DISCUSSION

Germline determination in echinoderms

With fertilization and morphogenesis separated by a minimum of several weeks in the ancestor to extant echinoderms, there was and is ample opportunity for selective pressures to act independently on larval and adult morphologies (Smith, 1997). This is evidenced by the numerous instances of convergent evolution in echinoderm larval morphologies and developmental strategies (Chapter I). In addition to these documented cases of larval evolution, there is emerging evidence that different methods of germline determination have evolved in Echinodermata.

The method of germline determination falls broadly into two different categories, termed inductive and inherited, or epigenesis and preformation, respectively (Extavour and Akam, 2003). Broadly speaking, in an organism with inherited germline determination, whichever cells inherit a concentration of specific molecular factors (mRNA and/or protein) will develop into the germ cells. These specific factors often include *vasa*, an RNA helicase, *nanos*, a translational repressor, and *piwi*, an argonaute family member. In an inductive system, no pre-localized factors are inherited; rather a population of cells will interpret a host of signals and will be induced to differentiate into germ cells. Many classical model organisms have an inherited germ line (e.g. *D. melanogaster* or *C. elegans*); the most well-known model organism with an inductive mechanism is the mouse.

Until recently, the echinoderm germline that was studied most intensely was that of the purple sea urchin, *S. purpuratus*. Initially it was thought that the germline in this organism was inductive, because no pre-localized germ line factors were identified in the early embryo. Furthermore, the removal of the micromeres after the 4th cell cleavage (the cells that give rise to the large and small micromeres at the 5th cell cleavage), had no impact on the fertility of the adults, though there was a slight developmental delay in the larvae (Ransick *et al.*, 1996). In a following study, it was determined that after the removal of the micromeres, *vasa* expression and

accumulation was de-repressed in the remaining embryo, which was later localized to a new population of presumptive germ cells (Voronina *et al.*, 2008a). These two lines of evidence suggested that the germ line is induced later in embryogenesis or during metamorphosis. However, more recent evidence supports the hypothesis that the small micromeres are in fact presumptive primordial germ cells (Yajima and Wessel, 2011; Yajima and Wessel, 2012). The descendants of the micromere lineage behave in a cell autonomous manner in culture; isolated micromeres, upon division yield apparent large and small micromeres. Large micromeres will divide several times and migrate in the culture dish, while small micromeres remain mitotically quiescent and accumulate *nanos* and *vasa* protein; both of these phenotypes are observed in these cell populations in the embryo (Yajima and Wessel, 2012). In two different species of sea urchins, *L. variagatus* and *S. purpuratus*, removal of the micromeres results in developmental delay, but the larvae ultimately metamorphosize and become fertile adults (Ransick *et al.*, 1996; Yajima and Wessel, 2011). However, if the small micromeres are removed and the embryos raised to adulthood, the adults are sterile (Yajima and Wessel, 2011). Furthermore, the embryo wide *vasa* upregulation detected in micromere removed embryos is not observed in embryos where the small micromeres have been removed. Surprisingly, if the micromeres are removed at the 28 cell stage (instead of the 16 cell stage, as in Ransick *et al.*, 1996); just prior to the cell division that gives rise to the large and small micromere lineages, the compensatory *vasa* upregulation is absent which phenocopies the small micromere removal (Yajima and Wessel, 2011).

Evidence is accumulating rapidly for identifying the germline determination in another echinoderm, the sea star *P. miniata* (reviewed in Wessel *et al.*, 2013; Wessel *et al.*, 2014). The evidence supports the hypothesis that the germline is found in the posterior enterocoel (PE) and is induced, in contrast to the hypothesized inherited mechanism in the sea urchin *S. purpuratus* (Chapter III, Fresques *et al.*, 2013). Briefly summarized, presumptive primordial germ cells are decreased upon removal of the PE (Inoue *et al.*, 1992). Furthermore, germline factors are

enriched in the PE which is also depleted of somatic cell fate markers (Chapter III, Fresques *et al.*, 2013), and vasa protein is localized in the PE (Juliano and Wessel, 2009).

How has germline determination evolved in Echinodermata?

The ancestral mechanism for germline determination in echinoderms is either induction or inheritance of localized factors. There is strong evidence that a member of Asterozoa (*P. minata*) has an induced germline (Chapter III, Fresques *et al.*, 2013), and a member of the sister group to Asterozoa, the Echinozoid, *S. purpuratus* (Chapter II, Fig. 2), has an inherited germline (Yajima and Wessel, 2011; Yajima and Wessel, 2012). In the absence of any information outside of Echinodermata, it is equally likely that the ancestral echinoderm had an inductive germline (with a secondary gain of an inherited germline in Echinozoa), or that it had an inherited germline (with a gain of an induced germline in Asterozoa). However, the most closely related phyla to Echinodermata are Hemichordata and Xenoturbella, both of which have an induced germline (Extavour, 2007; Extavour and Akam, 2003). Therefore, the most parsimonious explanation is that the last common ancestor of Ambulacraria (Echinodermata, Hemichordata, and Xenoturbella) had an induced germline mechanism and that the sea urchin secondarily gained an inherited germline determination.

Due to the developmental plasticity conferred by feeding planktotrophic larvae, the adult and larvae are able to evolve independently (Smith, 1997). As such, sea urchins evolved a number of derived characteristics, some of which not seen in other echinoderms (Chapter II Fig. 3). Foremost among these changes is the invention of the micromere lineage. Micromeres are shared by all members of Echinozoa including the early branching cidaroids (Chapter II Fig. 2, 3; Bennett *et al.*, 2012). However, only the Euechinoids (sea urchins and sea biscuits; Chapter II Fig. 2) evolved the second asymmetric cell division at the 5th cleavage to give rise to the small micromeres. The Euechinoids have a unique combination of characteristics found nowhere else in Echinodermata (though individual characteristics are found throughout the phylum; Fig.1) including: micromeres and small micromeres (Chapter I), secretion of hyalin at fertilization

(Chapter II Fig. 3), and a broadly occluding fertilization envelope (Chapter IV, Oulhen *et al.*, 2013). I hypothesize that the evolution of an inherited germline in Euechinoids is correlated with the suite of character changes observed though it is difficult to determine causation (i.e. is an inherited germline predicated on this character suite, or does the suite make it more likely for an inherited germline to evolve).

Several of the character changes could be interpreted as protective of the newly formed small micromeres and by extension, the germ line. The feature that appears to have the most protective role is that of the fertilization envelope in *S. purpuratus* (Chapter IV, Oulhen *et al.*, 2013). The fertilization envelope is not the only method of defense used by the embryo (e.g. hyaline layer or cuticle), but it is the first line of defense. The fertilization envelope of *S. purpuratus* is able to occlude particles 50 times smaller than that of the sea star *P. miniata* (Chapter IV Fig. 1, Oulhen *et al.*, 2013). This is particularly important in the context of the marine environment, because viruses are incredibly abundant in the ocean (Suttle, 2005) and there are no intervening cells between the small micromeres and the environment. The fertilization envelope surrounds the embryo until blastula stage, just prior to gastrulation and the translocation of the small micromeres (Yajima and Wessel, 2012). Once the small micromeres cross the basal lamina during gastrulation, they are much less likely to be affected by environmental or external biological insults. It is unknown if the size disparity in occluded particles is due to a secondary loss off some kind in *P. miniata* or if the remarkably efficient fertilization envelope of *S. purpuratus* is unique to the phylum. If this dense envelope is unique to Euechinoids, it could be due in part to the presence of SFE1 in *S. purpuratus*. It will be important to identify the evolution of SFE1 in echinoderms in order to test this hypothesis.

Secreted hyalin during exocytosis of the cortical granules after fertilization can also serve a protective role. However, the role of this additional layer may be to support development as opposed to protective like the fertilization envelope. The hyaline layer provides a matrix for cell adhesion to endodermal and ectodermal cells during development but is selectively lost by

primary mesenchyme cells during gastrulation (McClay and Fink, 1982; Wessel *et al.*, 1998). Due to the very small size of small micromeres, it is likely important to provide a secondary point of attachment in addition to the large micromeres so that the small micromeres are not physically dislodged and lost during development. Before the hyaline layer transforms into the cuticle during metamorphosis, the hyaline layer could serve as a stable attachment for the small micromeres, before they translocate during gastrulation (Yajima and Wessel, 2012). From the Echinoidea clade, the early branching Cidaroids do not have hyalin, yet they do have micromeres (Bennett *et al.*, 2012). The attachment of cells to the hyaline layer is particularly stark when compared with sea star development which has a much less robust hyaline layer. During development of *P. miniata*, the blastomeres are not well anchored and with the loss of the fertilization envelope, the blastomeres form a mono layer on the substrate and development arrests (data not shown). In a closely related sea star, *P. pectinifera* the blastomeres do not aggregate together inside the fertilization envelope until much later in development (Dan-Sohkawa and Fujisawa, 1980).

The final character change that we see in Euechinoids is the formation of the small micromeres during the asymmetrical cell division of the micromeres at the 5th cell cleavage (Chapter I). This particular character change also appears to be supportive in nature. I hypothesize that the small micromeres evolved after the gain of inherited germ line simply as a matter of developmental efficiency. In a symmetrically dividing embryo, if the germ line was segregated at the 5th cell cleavage, 4 of the resulting 32 blastomeres would be germline fated. As such, these cells would not likely contribute to the growth and development of the larva. Therefore, this would constitute a loss of 12.5% of the maternal stores in the egg. In a feeding planktotrophic developmental strategy where maternal contribution is relatively low, this might constitute an unsustainable hardship if the nutrient content of the environment is too low. There would therefore be a strong selective pressure for a strategy that minimized the loss of maternal stores to the germline fated cells. The second asymmetric division of the micromeres to form the small micromeres could be one such strategy. This hypothesis is supported in part because if the small

micromeres are removed during embryogenesis at the 32 cell stage, there is no larval developmental delay (Yajima and Wessel, 2011). In addition, the small micromere transcriptome is enriched for maternally deposited transcripts, with few new transcripts (Appendix IV, Swartz *et al.*, *submitted*). This suggests that all factors necessary for larval development are depleted from the small micromeres or never transcribed.

Although causation or the particular order of evolutionary changes is difficult to attribute, some testable hypotheses are possible. Due to the presence of micromeres in cidaroids, this was likely the first change towards an inherited germline mechanism in Euechinoids, as it was unlikely to have evolved independently in both clades (Fig. 1). After Euechinoids and Cidarida diverged, I hypothesize that the first Euechinoid specific character change was the development of the dense fertilization envelope. This could be due to a duplication event leading to a paralogous SFE1 gene. The presence or absence of this particular gene outside of Euechinoids would be particularly informative. This could in turn allow for the evolution of an inherited germline cell population that arose during early cell divisions because that population would be well protected from the environment. Following the evolution of the inherited germline, I hypothesize that the last two character changes, small micromeres and the hyaline layer would co-evolve. As the small micromeres became more diminutive in comparison to the other blastomeres, a robust attachment network would be strongly selected for (Fig. 1). This hypothesis could be tested by examining the rates of evolution of brittle star hyalins compared to the sea urchin hyalins. If there was evidence of strong selective pressure on hyalin in sea urchins and relatively neutral selective pressure amongst the brittle star hyalins, this would lend support to this hypothesis.

Evidence of ancestral inductive germline determination in Euechinoids?

There is strong evidence that the ancestral mode of germline determination in echinoderms is an inductive mechanism (Extavour, 2007; Extavour and Akam, 2003), which is retained in a member of Asterozoa (Chapter III, Fresques *et al.*, 2013). In addition, recent

evidence strongly supports the hypothesis that the sea urchin uses an inherited germline specification (Yajima and Wessel, 2011; Yajima and Wessel, 2012). However, there is also evidence that sea urchins, have retained the ancestral inductive germline mechanism which can be activated under certain conditions, a limited “germline formation checkpoint”.

The first piece of evidence for this checkpoint comes from removing the micromeres, the original experiment that supported the hypothesis that sea urchins have an induced germline (Ransick *et al.*, 1996). Upon removal, there is a developmental delay and a compensatory expression of vasa protein throughout the embryo (Voronina *et al.*, 2008a). Eventually the embryo will recover and will metamorphose on time and the adults are fertile (Ransick *et al.*, 1996). However, if the descendants of the micromeres, the small micromeres (Chapter I) are removed a single cell division later, the embryo is not developmentally delayed with no ectopic expression of vasa protein and the resulting adult is infertile (Yajima and Wessel, 2011). Furthermore, if the micromeres are removed at the 28 cell stage, the embryonic phenotype is similar to the small micromere removed embryos (Yajima and Wessel, 2011) and presumably would lead to a sterile adult.

The micromeres form a powerful signaling complex, if transplanted to the animal pole, they will induce a second axis of gastrulation (Ransick and Davidson, 1993). The progenitors of these cells are critical to the development of the larvae (large micromeres) and to the fitness of the individual (small micromeres). I hypothesize that the embryo has a mechanism in place to test if the micromeres are present and furthermore that it is dependent on Nanos in somatic tissue. If the embryo does not pass the test, then a “germline formation checkpoint” is triggered and the embryo activates the ancestral inductive germline determination to recover the micromere lineage.

Nanos is selectively transcribed by the small micromere lineage and upon knocking down translation of the mRNA in the whole embryo, the embryo is unable to accumulate vasa protein, nor is it able to form the adult rudiment (Juliano *et al.*, 2010). In this scenario, the checkpoint has

been activated but is unable to progress beyond the checkpoint due to Nanos knockdown in all cells, therefore development arrests and the adult rudiment never forms. If the knockdown of Nanos is restricted to the micromere lineage, then development proceeds as normal, the checkpoint is not activated and the adult rudiment is formed (Juliano *et al.*, 2010). This is an informative experiment in light of recent results. The knockdown of Nanos allows CNOT6, a deadenylase, to accumulate in the small micromeres. The accumulation of CNOT6 leads to the loss of expression of small micromere specific transcripts, *Seawi* and *vasa* (Appendix IV, Swartz *et al.*, *submitted*). In essence, the small micromeres potentially clear the maternally inherited transcriptome and assume a more somatic transcriptomic profile. One particularly informative experiment that could be used to test this hypothesis would be to measure *vasa* protein expression in the micromere specific knockdown. If *vasa* protein is not upregulated as in the small micromere removal experiment, this would lend support for the hypothesis. If *vasa* protein is ectopically expressed in a similar manner to the micromere removal experiments, this would not support the hypothesis.

If the “germline formation checkpoint” hypothesis is correct, and the ancestral inductive germline specification pathway of echinoderms has been co-opted to test for an inherited germline in Euechinoids, it could potentially explain several curious gene expression patterns. Foremost, it could explain why *nanos* which is broadly conserved in germline determination and maintenance is critical in the somatic larval development of the sea urchin (Juliano *et al.*, 2010). Second, *vasa* mRNA is broadly expressed during early cell cleavages in *S. purpuratus* and is not enriched in the small micromeres until gastrulation, in contrast to the protein which is enriched by the 4th and 5th cell cleavages (Voronina *et al.*, 2008b). This ubiquitous *vasa* mRNA is similar to the expression pattern observed in the sea star *P. miniata* (Chapter III Fig. 2, Fresques *et al.*, 2013; Juliano and Wessel, 2009) and in the cidaroid, *E. tribuloides* (Juliano and Wessel, 2009); presumably both of these organisms use an inductive germline determination. The broad expression of these factors would be important during the activation of the checkpoint, but the

rapid turnover of these factors would also be critical in the case of a successfully passed checkpoint. Finally, there appears to be a brief window of time between the 16 cell embryo and the 28 cell embryo, where the hypothesized ancestral inductive checkpoint is permanently disabled (Yajima and Wessel, 2011). Urchins perhaps have retained the ancestral state of inductive germline specification, but that gene regulatory network is shut down just prior to the formation of small micromeres, during the slight developmental delay between the asynchronously dividing blastomeres.

Finally, if the hypothesis is correct, then it could lead to some very exciting experiments in germline determination, because in a single system, both inductive and inherited mechanisms could be found. Furthermore, Euechinoids are a diverse group of organism, which suggests that there would be a wide variety of evolutionary experiments in the clade and particular variations on the theme likely occur naturally.

CONCLUSIONS

Echinoderms are a diverse group of organisms with a rich evolutionary history. This thesis documents many transitions that have occurred within the phylum and the underlying phylogenetic relationships of extant echinoderms. As researchers branch out and explore the entire phylum instead of select model systems, a greater diversity of hypotheses can be tested and reformulated. This could lead to a rapid expansion of knowledge in the fields of evolution and development; within Bilaterians and in general.

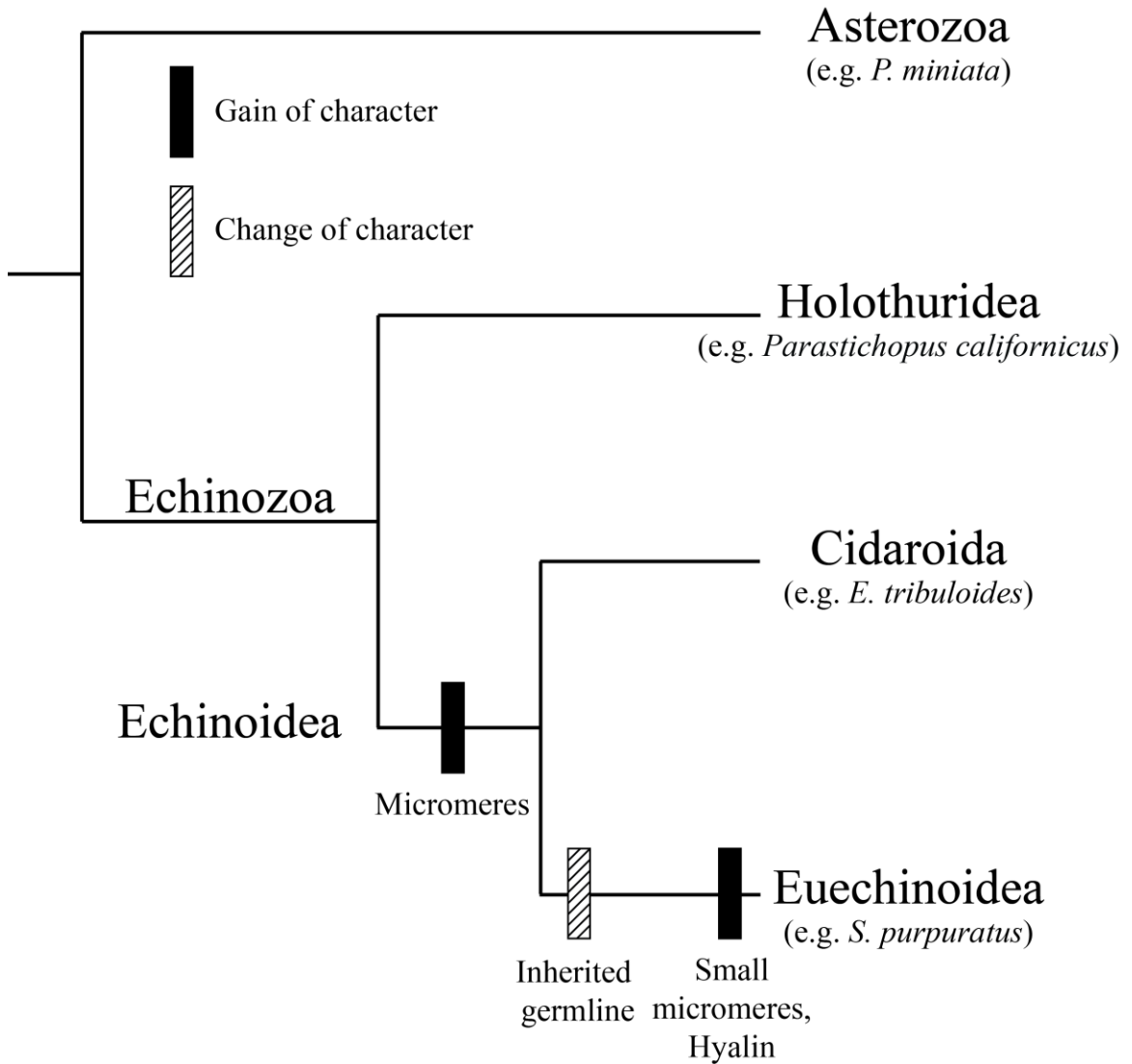
REFERENCES

- Bather, F. (1900). The Echinodermata: Treatise on Zoology, pt. 3.
- Bennett, K.C., Young, C.M., and Emler, R.B. (2012). Larval development and metamorphosis of the deep-sea cidaroid urchin *Cidaris blakei*. *Biol Bull* 222, 105-117.
- Dan-Sohkawa, M., and Fujisawa, H. (1980). Cell dynamics of the blastulation process in the starfish, *Asterina pectinifera*. *Dev Biol* 77, 328-339.
- Extavour, C.G. (2007). Evolution of the bilaterian germ line: lineage origin and modulation of specification mechanisms. *Integr Comp Biol* 47, 770-785.
- Extavour, C.G., and Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* 130, 5869-5884.
- Fresques, T., Zazueta-Novoa, V., Reich, A., and Wessel, G.M. (2013). 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*.
- Inoue, C., Kiyomoto, M., and Shirai, H. (1992). Germ cell differentiation in starfish: the posterior enterocoel as the origin of germ cells in *Asterina pectinifera*. *Dev Growth Differ* 34, 413-418.
- Janies, D.A., Voight, J.R., and Daly, M. (2011). Echinoderm phylogeny including *Xyloplax*, a progenetic asteroid. *Syst Biol* 60, 420-438.
- Juliano, C.E., and Wessel, G.M. (2009). An evolutionary transition of *Vasa* regulation in echinoderms. *Evol Dev* 11, 560-573.
- Juliano, C.E., Yajima, M., and Wessel, G.M. (2010). *Nanos* functions to maintain the fate of the small micromere lineage in the sea urchin embryo. *Dev Biol* 337, 220-232.
- Mac Bride, E.W. (1906). Echinodermata (Macmillan & Company).
- McClay, D.R., and Fink, R.D. (1982). Sea urchin hyalin: appearance and function in development. *Dev Biol* 92, 285-293.
- Oulhen, N., Reich, A., Wong, J.L., Ramos, I., and Wessel, G.M. (2013). Diversity in the fertilization envelopes of echinoderms. *Evol Dev* 15, 28-40.
- Pisani, D., Feuda, R., Peterson, K.J., and Smith, A.B. (2012). Resolving phylogenetic signal from noise when divergence is rapid: a new look at the old problem of echinoderm class relationships. *Mol Phylogenet Evol* 62, 27-34.
- Ransick, A., Cameron, R.A., and Davidson, E.H. (1996). Postembryonic segregation of the germ line in sea urchins in relation to indirect development. *Proc Natl Acad Sci U S A* 93, 6759-6763.
- Ransick, A., and Davidson, E.H. (1993). A complete second gut induced by transplanted micromeres in the sea urchin embryo. *Science* 259, 1134-1138.
- Smith, A.B. (1997). Echinoderm larvae and phylogeny. *Annual review of ecology and systematics* 28, 219-241.
- Suttle, C.A. (2005). Viruses in the sea. *Nature* 437, 356-361.
- Voronina, E., Lopez, M., Juliano, C.E., Gustafson, E., Song, J.L., Extavour, C., George, S., Oliveri, P., McClay, D., and Wessel, G. (2008a). *Vasa* protein expression is restricted to the small micromeres of the sea urchin, but is inducible in other lineages early in development. *Developmental biology* 314, 276-286.
- Voronina, E., Lopez, M., Juliano, C.E., Gustafson, E., Song, J.L., Extavour, C., George, S., Oliveri, P., McClay, D., and Wessel, G. (2008b). *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.
- Wessel, G.M., Berg, L., Adelson, D.L., Cannon, G., and McClay, D.R. (1998). A molecular analysis of hyalin--a substrate for cell adhesion in the hyaline layer of the sea urchin embryo. *Dev Biol* 193, 115-126.

- Wessel, G.M., Brayboy, L., Fresques, T., Gustafson, E.A., Oulhen, N., Ramos, I., Reich, A., Swartz, S.Z., Yajima, M., and Zazueta, V. (2013). The biology of the germ line in echinoderms. *Mol Reprod Dev*.
- Wessel, G.M., Fresques, T., Kiyomoto, M., Yajima, M., and Zazueta, V. (2014). Origin and development of the germ line in sea stars. *Genesis*.
- Yajima, M., and Wessel, G.M. (2011). Small micromeres contribute to the germline in the sea urchin. *Development* *138*, 237-243.
- Yajima, M., and Wessel, G.M. (2012). Autonomy in specification of primordial germ cells and their passive translocation in the sea urchin. *Development* *139*, 3786-3794.

FIGURES

Figure 1. Detailed trait changes in echinoderms.



Euechinoids are unique among echinoderms with having a suite of character changes. Several are shared by other members (e.g. hyalin in brittle stars, a member of Asterozoa) but no other clade has the entire suite. The hypothesized order of gains and/or changes is depicted by the order of characteristics, left to right. Not shown: changes in the density of the fertilization envelope.

**Appendix I: The transcriptome of a human polar body accurately reflects its sibling
oocyte**

Adrian Reich*, Peter Klatsky*, Sandra Carson, and Gary Wessel

* These authors contributed equally to this work

Journal of Biological Chemistry (2011) 286, no. 47: 40743-40749.

CONTRIBUTION

I conducted all experiments and analyses except for the sample biopsies and WTA amplifications.

ABSTRACT

Improved methods are needed to reliably and accurately evaluate oocyte quality prior to fertilization and transfer into the woman of human embryos created through *in vitro* fertilization (IVF). All oocytes that are retrieved and mature in culture are exposed to sperm with little in the way of evaluating the oocyte quality. Further, embryos created through IVF are currently evaluated for developmental potential by morphology, a criterion lacking in quantitation and accuracy. With the recent successes in oocyte vitrification and storage, clear metrics are needed to determine oocyte quality prior to fertilizing.

The first polar body (PB) is extruded from the oocyte before fertilization and can be biopsied without damaging the oocyte. Here we tested the hypothesis that the PB transcriptome is representative of that of the oocyte. Polar body biopsy was performed on metaphase II (MII) oocytes followed by single-cell transcriptome analysis of the oocyte and its sibling PB. Over 12,700 unique mRNAs and miRNAs from the oocyte samples were compared to the 5,431 mRNAs recovered from the sibling PBs (5,256 shared mRNAs or 97%, including miRNAs). The results show that human PBs reflect the oocyte transcript profile, and suggests that mRNA detection and quantification through high-throughput qPCR could result in the first molecular diagnostic for gene expression in MII oocytes. This could allow for both oocyte ranking and embryo preferences in IVF applications.

INTRODUCTION

The clinical importance of healthy oocyte development is evidenced by the impressive pregnancy rates seen with infertile women using assisted reproductive technology (ART) with oocytes from young, fertile donors. Oocytes from young women have lower rates of meiotic errors and aneuploidy, and although aneuploidy is the most common cause of developmental arrest, screening embryos for aneuploidy does not exclude all embryos of poor prognosis. Earlier studies have demonstrated that as the primary oocyte develops, it transcribes thousands of genes whose products are necessary for fertilization and early embryonic development. Prior to meiosis I, the germinal vesicle breaks down and transcriptional factors disengage from chromatin, rendering the cell transcriptionally silent (Sun *et al.*, 2007). This is particularly relevant since the human zygotic genome is not activated and does not transcribe its DNA for 2-3 days following fertilization (Braude *et al.*, 1988; Vassena *et al.*, 2011), the exact period during which an IVF clinician decides which embryo to transfer into the woman (Centers for Disease Control and Prevention, 2010, see especially: Fig. 37). Therefore, mRNAs needed for fertilization and early embryonic development must be present in the oocyte in sufficient quantity or ratio before the first polar body is extruded and guide the majority of embryonic processes to day 3. Thus, the transcriptome of the oocyte may predict both oocyte quality and the early developmental potential of the embryo.

The ability to measure oocyte gene expression without harming the oocyte may prove helpful to clinicians caring for patients using ART. Biopsying the polar body would allow embryologists to test for functional control of gene expression in the oocyte resulting from mRNA transcription, turnover and from epigenetic processes that depend on more complex determinants than having an appropriate number of chromosomes (Biddle *et al.*, 2009; Evsikov and Marin de Evsikova, 2009; Seli *et al.*, 2005) all without compromising the oocyte.

The transcriptome of a polar body has never been reported, and a polar body biopsy involves its careful removal through microdissection. This procedure can be performed without

damaging the sibling oocyte or developing embryo (Verlinsky *et al.*, 1990) and with advances in oocyte vitrification (Rienzi *et al.*, 2010), this could be helpful for patients with ethical objections to fertilizing multiple oocytes and creating supernumerary embryos or it can be applied to the growing practice of oocyte vitrification for donor egg banking. One can also imagine using gene expression information from a polar body to prioritize embryos for transfer in an IVF cycle. Here we are first to report the analysis of polar body transcriptomes from any organism and analyze its mRNA population with that of its sibling oocyte.

METHODS

Human Oocyte Collection and Polar Body Biopsy

Human oocytes were collected from infertility patients undergoing controlled ovarian hyperstimulation for *in vitro* fertilization (IVF) under standard clinical protocol. Germinal vesicle and MI staged oocytes that were not mature for a clinically indicated intracytoplasmic sperm injection (ICSI) procedure underwent *in vitro* maturation for 24 hours and were used in the study if they extruded a polar body. Written consent was obtained from all patients to use discarded tissue and oocytes for research, and the study was approved by the institutional review board at Women & Infants Hospital. Briefly, patients underwent controlled ovarian hyperstimulation, with either luteal downregulation using a GnRH agonist, pituitary suppression using a GnRH antagonist, or a microdose lupron “flare protocol” consisting of daily lupron injections initiated in the follicular phase with gonadotropins. Oocytes were aspirated by ultrasound guided transvaginal oocyte retrieval 36 hours after injection with recombinant HCG. Four hours after retrieval, all oocytes were mechanically stripped of cumulus cells. ICSI was performed in all oocytes with visible polar bodies. After injection of all MII oocytes, any remaining immature oocytes were cultured for 20-24 hours in IVM media. Immature oocytes were examined the next day and oocytes that extruded a polar body were used for our study. A total of 22 oocytes and sibling polar bodies were collected and individually processed in this study.

Biopsy and WTA Amplification

All biopsies were performed at 200X magnification after mechanical zona drilling with a polar body biopsy needle (Cook Medical, Bloomington, IN). Polar bodies were aspirated into a glass micropipette with an inner diameter of 20 μ m. The polar body was then processed using the lysis buffer and DNase 1 from the Ambion Cells-to-Ct Direct kit (Life Sciences, Carlsbad, CA), followed by reverse transcription and whole transcriptome amplification using the WTA2 kit (Sigma-Aldrich, St. Louis, MO). Sibling oocytes were transferred to an identical lysis solution and processed using the same protocol. The lysed specimens were stored on ice for no more than 2 hours while other oocytes were biopsied; the lysates were then processed according to the WTA2 protocol. Briefly, primers with a common ~25bp and a pseudo-random 9bp sequence, designed to favor binding to mRNA over mitochondrial and ribosomal sequences, bind to RNA and reverse transcription occurs. This reverse transcription reaction occurs at graduated temperatures over several extension phases and a final volume of 25 μ L is generated. The cDNA is then amplified using the common ~25bp sequence for 14 cycles in a final volume of 75 μ L (Fig. 1a).

In an attempt to maximize cDNA yield, 10 μ L of the amplified cDNA was added to 65 μ L of amplification mix and the contents underwent an additional 15 rounds of amplification specific for cDNA containing the WTA products. The second amplification was combined with the remainder of the first amplification step. The final concentration for all libraries was between 30 and 40 ng/ μ L as measured by QuBit (Invitrogen, Carlsbad, CA). Each individual oocyte and polar body were processed in separate reactions and for those samples that were pooled, ten oocytes or polar bodies were pooled together in a common tube after the two rounds of WTA amplification. The 22 oocyte and sibling polar body pairs (44 cells) were split into two replicates of 10 pooled cells and two replicates of a single cells, for a total of 8 samples, 4 oocyte samples and 4 polar body samples.

Illumina Library Preparation and Sequencing

The cDNA resulting from two rounds of the WTA amplification yielded fragments between 100 and 300bp in length. The cDNA was not subjected to any additional shearing and libraries were prepared using the NEBNext DNA Sample Prep Kit (NEB, Ipswich, MA) with adapters and PCR primers from IDT (Coralville, IA). The standard protocol was used with starting material of no less than 1.5µg total cDNA with one slight modification. After the ligation of the adapters to the ends of the cDNA molecules, PCR amplification was performed without an intervening gel purification. After the PCR step, gel purification of the completed library was performed and a wide band of 200-450bp was cut from the gel. The library concentration was determined by qPCR (Kapa Biosystems, Woburn, MA) and size by Bioanalyzer (Agilent Technologies, Santa Clara, CA). The samples were sequenced for 42bp on a GAIIx (Illumina, Inc., San Diego, CA) using a custom sequencing primer consisting of the 6bp most 3' of the Illumina sequencing primer fused to the WTA2 primer sequence (Fig. 1b).

Mapping and Statistical Analysis

The raw sequences were mapped against the human genome (UCSC hg18) using Illumina's software, Casava v1.7 using 32bp of the read and allowing only 2 mismatches. The raw gene counts were then loaded into edgeR (Robinson *et al.*, 2010) which normalized the counts using the TMM method (trimmed mean of M values) (Robinson and Oshlack, 2010), and these counts were used for all further analyses. The geometric means of the TMM counts of every gene for all oocyte or polar body samples was used to generate a list of the expression levels of all genes. The R package rankedListComparison (Antosh *et al.*, 2011) was used to analyze the expression level lists.

RESULTS AND DISCUSSION

Analysis of Detected Genes and Gene Expression Levels

We analyzed the transcriptomes of 22 oocytes alongside their 22 polar body siblings by high throughput DNA sequencing. The samples were grouped into two biological replicates of 10

oocytes and their sibling polar bodies, and two biological replicates of single oocytes and their single polar bodies. We developed a method for quantitative cDNA construction from both a single oocyte and its sibling polar body and we detected a total of 12,883 genes through mapping of more than 27 million reads from these oocytes and polar bodies (Table 1). From this result we estimate that between 14,000-15,000 genes are expressed in the human oocyte (Supplementary Fig. 1). The genes that were expressed in each oocyte highly correlated with those that were expressed in other oocytes. Of the 7,523 genes detected in the smallest oocyte sample, 84.5% of the genes were detected in all four oocyte samples and over 98% were detected in at least two samples (Fig. 2a). Furthermore of the four oocyte/polar body pairs in this study, greater than 90% of all the genes detected in a polar body sample were also detected in the sibling oocyte sample (Fig. 2b). This result might be expected since the polar body and oocyte shared a common ooplasm no more than 24 hours prior to polar body biopsy, but no less surprising because of the diversity of the transcripts detected in the polar body samples. Comparing the overlap of each sibling oocyte/polar body pair with every other pair reveals that 279 genes (28.0% of genes detected in the smallest overlap pool) are expressed in all four paired samples (Fig. 2c and Supplementary Table 1); 962 genes were detected in both the oocyte and sibling polar body samples in at least 3 of the four pairs (Fig. 2c).

While the sample to sample overlap of genes was very high in all examined comparisons, a critical component of the analysis was testing the abundance of those genes products. We performed a pair wise comparison of each oocyte sample and tested the levels of all gene products shared between each oocyte pair. The expression level of any gene in any oocyte sample correlated very strongly (Pearson correlation > 0.88) with the expression level in any other oocyte (Supplementary Fig. 2). One hypothesis is that the polar body is a depository of the oocyte, that transcripts or cellular contents no longer needed or undesirable in the oocyte are transported to the polar body. We tested this hypothesis by doing a differential gene enrichment analysis of all oocyte samples against all polar body samples and found no genes that were differentially

enriched between the two populations at any levels of significance (Supplementary Fig. 3). Multiple normalization methods were used in this testing and all results arrived at the same conclusion: the transcriptome of the polar body accurately reflects its sibling oocyte. The observation that transcript abundance is very similar between oocytes and polar bodies for all detected genes strongly argues against the interpretation that messages are selectively transported into or out of the polar body. A more parsimonious explanation is that as the polar body is extruded, it captures a representative portion of the ooplasm and therefore a representative transcriptome. Furthermore, no genes were sampled in all four polar body samples without being detected in a single oocyte, further evidence that specific transcripts are not localized to the site of the budding polar body and that the polar body inherits the general transcriptionally silenced state of the oocyte (Sun *et al.*, 2007). These results support the interpretation that the transcriptome of one oocyte is very similar to that of another oocyte in both the genes expressed and the expression levels and that the transcriptome of a polar body is directly representative of the transcriptome of its sibling oocyte.

Examination of Gene Expression Profiles

We took particular care to examine the data using the whole transcriptome amplification (WTA) procedure coupled with Illumina sequencing. A potential concern was that of amplification bias, but a formal test using the Ambion Cells-to-Ct kit and qPCR which has been previously tested to linearly amplify transcripts (Klatsky *et al.*, 2010b) would have been prohibitively costly. However, the relative rank abundance of half a dozen previously reported transcripts (Klatsky *et al.*, 2010b) was recapitulated in this study using the WTA method. A second concern was the introduction of contaminants or the amplification of PCR mutations especially due to the two rounds of amplification in this study, potentially leading to mapping error. To control for this variable, we examined the number of reads that mapped to genes on the Y chromosome; because this study was conducted on oocytes and polar bodies that had never been exposed to sperm, there is no biological template for Y chromosome genes. Of the more

than 27 million reads in this study, only 128 reads ($<5 \times 10^{-4}$ % of the total reads) mapped to two genes found on the Y chromosome (USP9Y and DDX3Y), both of which have paralogs on the X chromosome (USP9X and DDX3X). The percent identity for the two sets of paralogs are very high (USP9=88.3% and DDX3=88.8%), therefore Y-chromosome mapping could be due to sequencing errors or statistical mapping error. The third and final concern was the recovery of sufficient material from a single oocyte; much less from a single polar body. To address this we used two different types of samples, pooled and single specimens. We felt that the pooled specimens would decrease the oocyte to oocyte variability in gene expression and insure that the cDNA would not be limiting. The single specimens would serve as a test of the clinical feasibility of reliably detecting transcripts in single cells as well as testing the variability of oocyte to oocyte differences in transcript abundance.

We used the bioinformatics package DAVID (Huang da *et al.*, 2009) to test if the genes that were detected in all oocyte samples completed specific enzymatic pathways. Of all the detected genes in the oocyte samples, several dozen KEGG pathways were significantly complete including “Oocyte meiosis” (p-value= 4.5×10^{-7}) and “Progesterone-mediated oocyte maturation” (p-value= 3.6×10^{-4}) (Supplementary Fig. 4), providing additional evidence that the transcriptome we are detecting are representative of oocytes.

Intriguingly, a number of miRNAs in oocytes and polar bodies were detected, some of which are very abundant. Previous reports suggest that the miRNA pathway is non-functional in mouse oocytes (Ma *et al.*, 2010; Suh *et al.*, 2010), and because our sequencing technique would have missed mature miRNAs, we hypothesize that the high abundance of miRNAs is due to the presence of pre-processed miRNAs. The down-regulation of the miRNA pathway may be due to the repression of Drosha/DGCR8 which would allow pre-processed hairpins to accumulate. This hypothesis is supported in part by the observation that the transcriptome of the mouse oocyte is minimally impacted by the loss of zygotic DGCR8 during early development (Suh *et al.*, 2010). An abundance of pre-processed miRNA containing transcripts in the nucleus could explain the

slightly higher abundance (statistically non-significant) of MIR genes in the polar body compared to the oocyte. The transcriptome of the polar body may be enriched for mRNAs that associate with the meiotic spindle. Since the polar body is significantly smaller than the oocyte, it may be unduly enriched for such transcripts.

We also report that the abundance of some transcripts may vary significantly between single cells. The most notable example was the transcript level of the Anthrax receptor, ANTXR2, which varied in expression between the two single oocytes by 4 orders of magnitude. Remarkably, this difference in abundance was also reflected in the sister polar body by the same 4 orders of magnitude difference. Previous literature has shown ANTXR2 to be detected in oocytes (Grondahl *et al.*, 2010) and significantly downregulated (Kocabas *et al.*, 2006) in human oocytes compared to a reference. Such variation may in reflect atypical regulation resulting from differences in the oocyte genome, differences in nutritional status of the oocyte, of donor age, and/or environmental influences on the oocyte.

Clinical Feasibility Test and Microarray Comparison

For successful transition into a clinical application using this approach, it is important to identify a cohort of transcripts that are reliably detected, whose expression levels are predictive of the developmental competence of a given oocyte and its resulting embryo. In order to develop a list of candidate transcripts for future studies, we generated a separate rank order list of the transcript abundance for oocytes and for polar bodies by taking the geometric means of the TMM normalized gene counts in all four samples. The two separate lists were then compared with each other in discrete subsets to test the degree of overlap between the two lists within each independent subset. In the subset of the 50 most abundant genes in oocytes and polar bodies, 39 genes are shared between the two lists within that subset (p-value= $1.23e^{-74}$ when compared to a randomly ordered list). Out of the 100 most abundant genes detected in all four oocyte samples, 72 were detected in all four polar body samples and 61 of those were in the top 100 most abundant genes in polar bodies. In total the 700 most abundant genes in each list constitute the

significant overlap between the oocyte and polar body samples (combined p-value= $6.16e^{-250}$) and within the 700 genes, 460 genes are shared between the oocyte and polar body lists (Fig. 3). Nearly half of the 460 shared genes (215 genes) were detected in all oocyte and polar body samples (Figure 2c and Supplementary Table 1).

Analysis of our dataset strongly suggests that the polar body captures a representative transcriptome of the oocyte and that the transcriptomes of both single oocytes and single polar bodies can be quantitatively assessed. Although several microarray studies of human oocytes have been reported (Assou *et al.*, 2006; Bermudez *et al.*, 2004; Gasca *et al.*, 2007; Grondahl *et al.*, 2010; Jones *et al.*, 2008; Kocabas *et al.*, 2006), this is also the first study that uses deep sequencing technology on human oocytes. We compared our sequencing dataset to that of three microarray studies to validate our findings (Assou *et al.*, 2006; Grondahl *et al.*, 2010; Kocabas *et al.*, 2006). Those genes that were most highly enriched in oocytes compared to the reference in microarray studies were more likely to be the same genes that were most abundant in our dataset (Fig. 4). Additionally, the genes reported to be significantly enriched in the reference compared to the oocytes were more likely to be genes detected at very low abundance in our oocyte dataset. The total number of genes detected in polar bodies was highly variable between each sibling pair of oocytes and polar bodies, but the genes most abundant in oocytes were much more likely to be detected in one or more polar body samples. Furthermore, genes that were previously reported to be detected in oocytes and polar bodies by qPCR (Klatsky *et al.*, 2010b) were present in the same rank abundance in this study. One potential concern we had for biopsying the polar body was the fear that instead of sequencing the transcriptome of the polar body, we would contaminate the sample with a cumulus cell or other accessory cell. Comparing our data to that of a microarray study of human oocytes and cumulus cells, the polar body transcriptome is distinct from cumulus cells and instead, very closely aligns with oocyte samples (Fig. 4).

CONCLUSIONS

We demonstrated that detection and quantification of mRNA in polar bodies is possible and reflects the transcript profile of the MII oocyte. The quantification of mRNAs is of particular importance and we also report that some transcripts can have highly variable expression in different oocytes and that this variance can be reflected in the polar body. This variance may reflect many factors seen by the oocyte during its development including nutritional status and environmental insults to the oocyte. Genes with higher levels of expression in oocytes are more reliably detectable in the sibling polar body, suggesting that a failure to identify a particular mRNA in the polar body relates to transcript levels within the oocyte that fall below a critical threshold. This finding is consistent with our previous results in both human and sea star oocytes (Klatsky *et al.*, 2010a; Klatsky *et al.*, 2010b). Our results suggest that the detection and analysis of polar body mRNA may provide insight into oocyte quality, a critical metric needed by the clinical before fertilization and transfer of the resulting embryo back into the women.

DATA AVAILABILITY

Raw sequence files, raw gene counts, and TMM normalized gene counts have been deposited in NCBI Gene Expression Omnibus (Edgar *et al.*, 2002) and are available under GEO series accession number GSE32689.

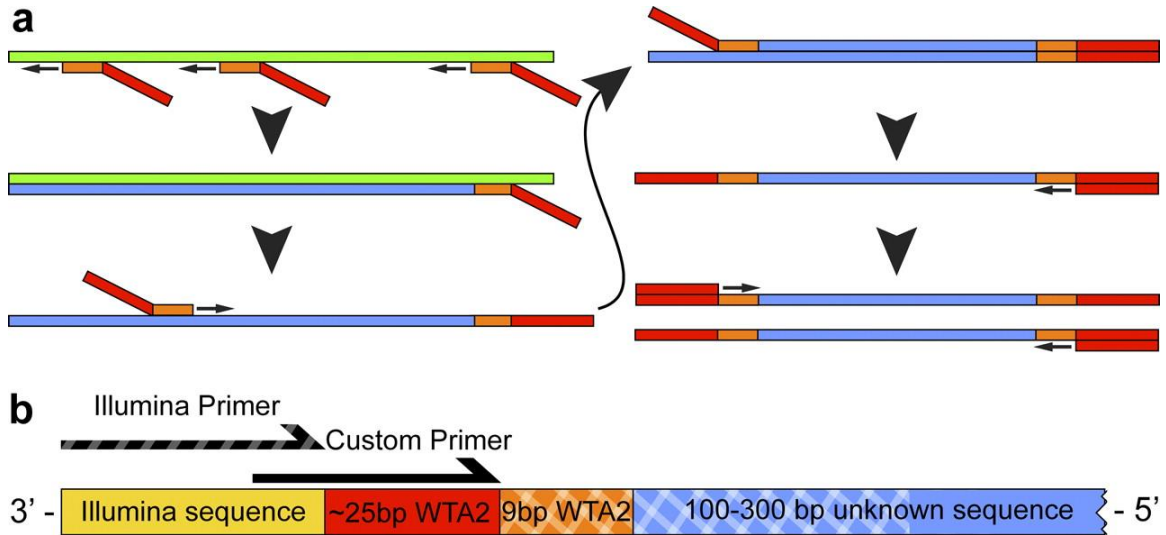
REFERENCES

- Antosh, M., Fox, D., Helfand, S.L., Cooper, L.N., and Neretti, N. (2011). New comparative genomics approach reveals a conserved health span signature across species. *Aging (Albany NY)* 3, 576-583.
- Assou, S., Anahory, T., Pantesco, V., Le Carrou, T., Pellestor, F., Klein, B., Reyftmann, L., Dechaud, H., De Vos, J., and Hamamah, S. (2006). The human cumulus--oocyte complex gene-expression profile. *Hum Reprod* 21, 1705-1719.
- Bermudez, M.G., Wells, D., Malter, H., Munne, S., Cohen, J., and Steuerwald, N.M. (2004). Expression profiles of individual human oocytes using microarray technology. *Reprod Biomed Online* 8, 325-337.
- Biddle, A., Simeoni, I., and Gurdon, J.B. (2009). *Xenopus* oocytes reactivate muscle gene transcription in transplanted somatic nuclei independently of myogenic factors. *Development* 136, 2695-2703.
- Braude, P., Bolton, V., and Moore, S. (1988). Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* 332, 459-461.
- Centers for Disease Control and Prevention, A.S.f.R.M., Society for Assisted Reproductive Technology (2010). *Assisted Reproductive Technology Success Rates: National Summary and Fertility Clinic Reports*, U.S.D.o.H.a.H. Services, ed. (Atlanta), pp. 51.
- Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30, 207-210.
- Evsikov, A.V., and Marin de Evsikova, C. (2009). Gene expression during the oocyte-to-embryo transition in mammals. *Mol Reprod Dev* 76, 805-818.
- Gasca, S., Pellestor, F., Assou, S., Loup, V., Anahory, T., Dechaud, H., De Vos, J., and Hamamah, S. (2007). Identifying new human oocyte marker genes: a microarray approach. *Reprod Biomed Online* 14, 175-183.
- Grondahl, M.L., Yding Andersen, C., Bogstad, J., Nielsen, F.C., Meinertz, H., and Borup, R. (2010). Gene expression profiles of single human mature oocytes in relation to age. *Hum Reprod* 25, 957-968.
- Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4, 44-57.
- Jones, G.M., Cram, D.S., Song, B., Magli, M.C., Gianaroli, L., Lacham-Kaplan, O., Findlay, J.K., Jenkin, G., and Trounson, A.O. (2008). Gene expression profiling of human oocytes following in vivo or in vitro maturation. *Hum Reprod* 23, 1138-1144.
- Klatsky, P.C., Carson, S.A., and Wessel, G.M. (2010a). Detection of oocyte mRNA in starfish polar bodies. *Mol Reprod Dev* 77, 386.
- Klatsky, P.C., Wessel, G.M., and Carson, S.A. (2010b). Detection and quantification of mRNA in single human polar bodies: a minimally invasive test of gene expression during oogenesis. *Mol Hum Reprod* 16, 938-943.
- Kocabas, A.M., Crosby, J., Ross, P.J., Otu, H.H., Beyhan, Z., Can, H., Tam, W.L., Rosa, G.J., Halgren, R.G., Lim, B., *et al.* (2006). The transcriptome of human oocytes. *Proc Natl Acad Sci U S A* 103, 14027-14032.
- Ma, J., Flemr, M., Stein, P., Berninger, P., Malik, R., Zavolan, M., Svoboda, P., and Schultz, R.M. (2010). MicroRNA activity is suppressed in mouse oocytes. *Curr Biol* 20, 265-270.
- Rienzi, L., Romano, S., Albricci, L., Maggiulli, R., Capalbo, A., Baroni, E., Colamaria, S., Sapienza, F., and Ubaldi, F. (2010). Embryo development of fresh 'versus' vitrified metaphase II oocytes after ICSI: a prospective randomized sibling-oocyte study. *Hum Reprod* 25, 66-73.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.
- Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11, R25.

- Seli, E., Lalioti, M.D., Flaherty, S.M., Sakkas, D., Terzi, N., and Steitz, J.A. (2005). An embryonic poly(A)-binding protein (ePAB) is expressed in mouse oocytes and early preimplantation embryos. *Proc Natl Acad Sci U S A* *102*, 367-372.
- Suh, N., Baehner, L., Moltzahn, F., Melton, C., Shenoy, A., Chen, J., and Belloch, R. (2010). MicroRNA function is globally suppressed in mouse oocytes and early embryos. *Curr Biol* *20*, 271-277.
- Sun, F., Fang, H., Li, R., Gao, T., Zheng, J., Chen, X., Ying, W., and Sheng, H.Z. (2007). Nuclear reprogramming: the zygotic transcription program is established through an "erase-and-rebuild" strategy. *Cell Res* *17*, 117-134.
- Vassena, R., Boue, S., Gonzalez-Roca, E., Aran, B., Auer, H., Veiga, A., and Izpisua Belmonte, J.C. (2011). Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development* *138*, 3699-3709.
- Verlinsky, Y., Ginsberg, N., Lifchez, A., Valle, J., Moise, J., and Strom, C.M. (1990). Analysis of the first polar body: preconception genetic diagnosis. *Hum Reprod* *5*, 826-829.

FIGURES AND TABLES

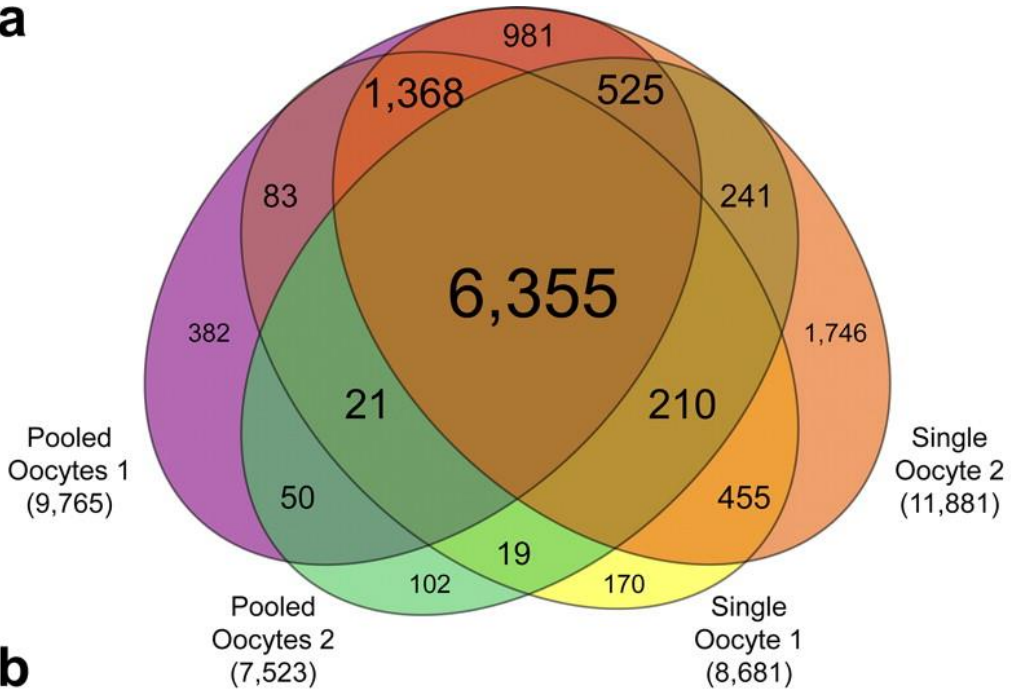
Figure 1: Amplification and sequencing strategy.



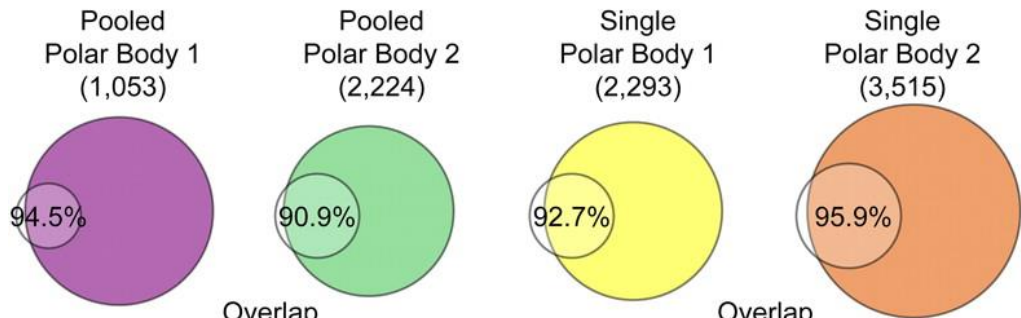
a. Reverse transcription and second strand cDNA synthesis using the WTA2 kit. Each WTA2 primer has a pseudo-random nonamer (orange) that are designed to preferentially bind to mRNA sequences (green) over rRNA. After genomic DNA digestion, the first strand of cDNA is synthesized (blue). Following an RNase H step, a second round of cDNA synthesis occurs using the same WTA2 primers with the pseudo-random nonamers. The target library was then subjected to two rounds of 15 cycles of PCR amplification using just the WTA2 primer sequence (red). The WTA2 kit produced 30ng/μL of cDNA fragments of mRNA between 100 and 300 bases long. **b.** Final library construction and sequencing primer. See Materials & Methods for library preparation procedure. The standard Illumina sequencing primer (striped primer) was not used because every sequenced cluster would have started with the same exact sequence, causing the sequencing reaction to fail. A custom sequencing primer was used which consisted of the WTA2 primer with the six most 3' bases of the Illumina sequencing primer. 42bp were sequenced (hatched) consisting of 9bp of the pseudo-random nonamer and 33bp of the unknown mRNA sequence. The first 9bp and the final base were removed, leaving 32bp sequences to map against the human genome.

Figure 2: Sample to sample overlap and comparison.

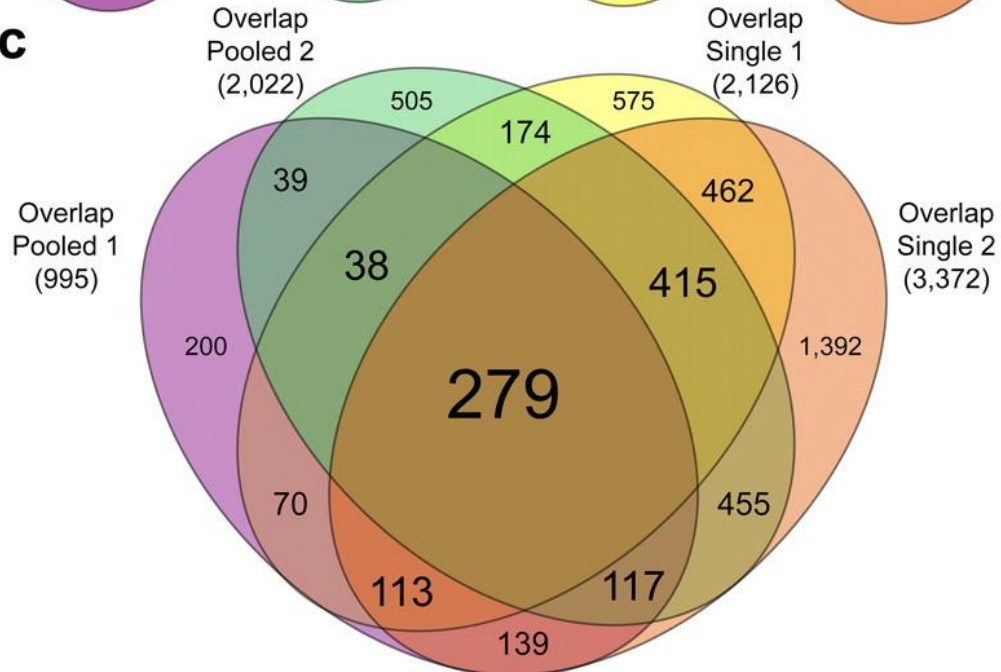
a



b

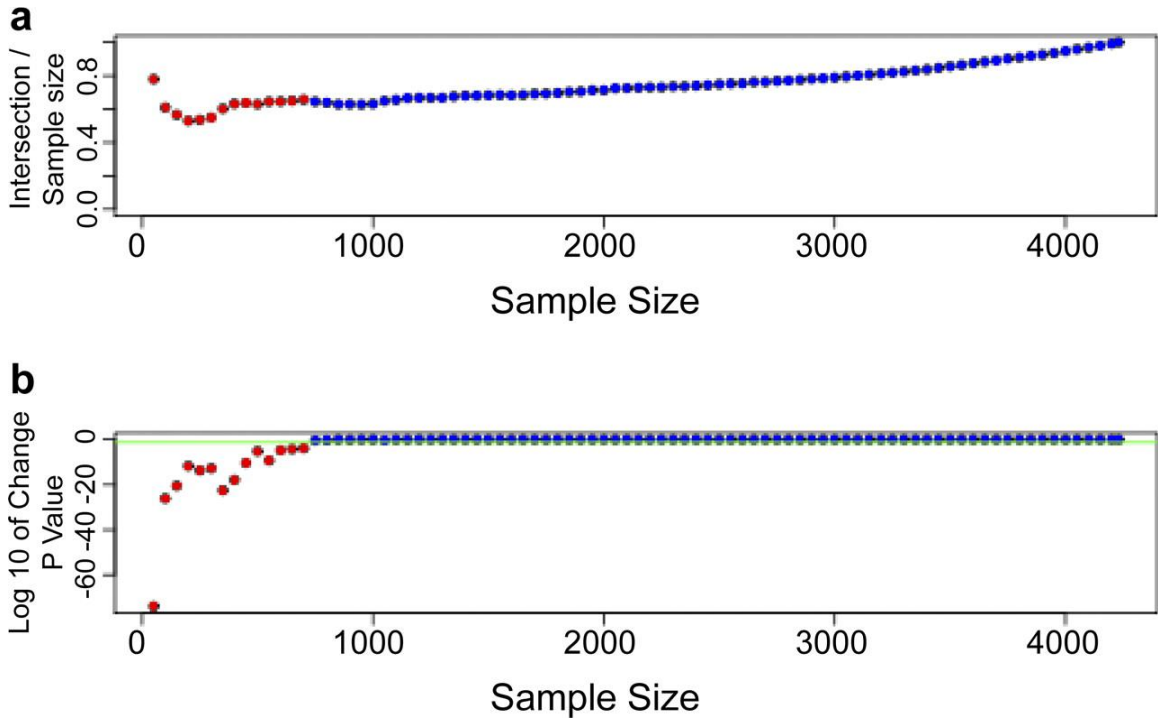


c



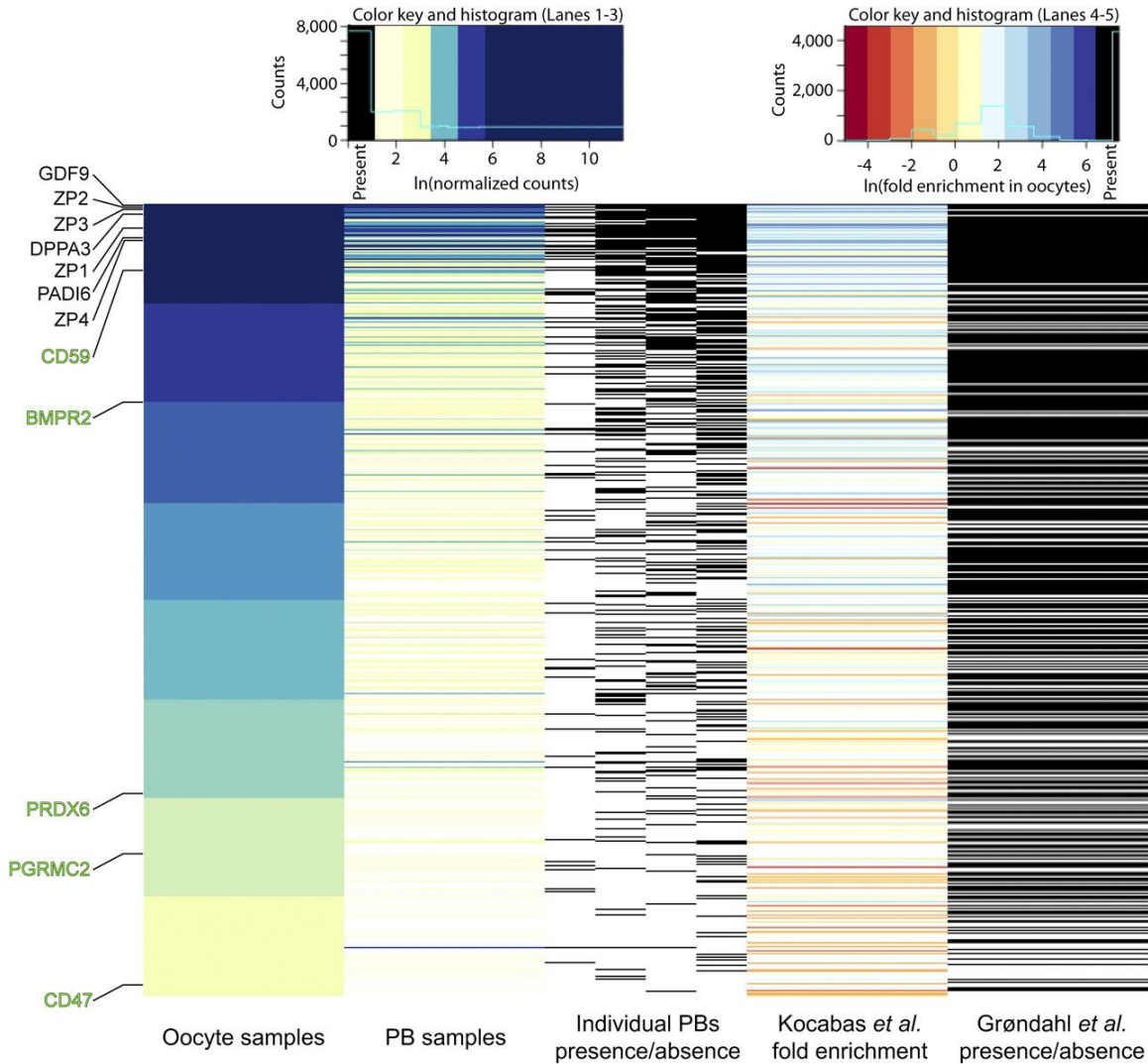
a. All four oocytes samples show a high degree of overlap with each other. The total number of genes for each sample is shown in parenthesis and the total number of all genes in all 4 samples is equal to 12,708 genes. 66.7% of all genes were detected in at least 3 of the 4 oocyte samples and 50.0% of all genes were detected in all 4 oocyte samples. **b.** The larger circle represents the total number of genes in each of the four oocyte samples and the smaller circle shows the overlap of the four sibling polar bodies. The percentage was calculated by taking the total number of genes shared between the sibling oocyte and polar body and dividing by the total number of genes in the polar body. **c.** The overlaps of genes transcribed in sibling oocyte and PB samples among the four independent comparisons are represented as in (a). In total there were 4,973 genes found between all the overlap datasets and 279 genes that were sampled in all 8 samples. Of the 4,973 overlap gene set, approximately 46% were detected in at least 2 of the overlaps.

Figure 3: The most abundant genes in oocytes are compared with the most abundant genes in polar bodies.



a. Each list is compared in increments of 50 genes testing for significant overlap in each section of 50. Iterations of 50 genes that are labeled in red have significant overlap between each list with a p-value <0.05 and the fraction of genes shared between the two lists is on the y axis. The first iteration of 50 had an overlap of nearly 80% (39 out of 50) and as each iteration adds to the total length of the shared list, the fraction of shared genes between the two lists approaches 100%. **b.** The individual p-value for each iteration of 50 is shown. The green line is the significance cutoff of $p=0.05$.

Figure 4: Comparison of our data with previously published data.



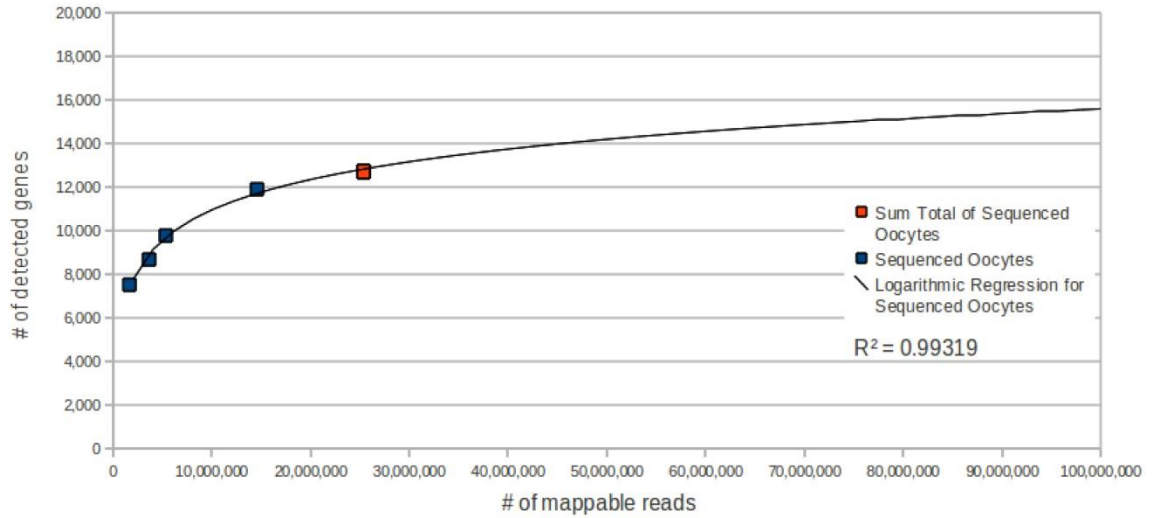
The 6,355 genes that were sampled in all 4 oocyte samples were compared to the polar body samples as well as previously published microarray studies of oocytes. Comparing the natural log of the geometric mean of the TMM normalized counts of the oocyte samples (lane 1) and the polar body samples (lane 2); genes that are most abundant in the oocytes samples (dark blue) are more likely to be sampled in the polar body samples and are also found at similar expression levels. Select genes are shown on the left including genes that have been shown to be highly expressed in oocytes (black text) (Klatsky, et al., 2010a) as well as genes significantly upregulated in cumulus cells (green text) that were found (Assou, et al., 2006). There is a high variation between the individual polar bodies (lane 3) but genes expressed at a high level in oocytes are more likely to be detected in polar bodies (black) than not (white). There is a very strong correlation between genes that are significantly upregulated (lane 4) or detected (lane 5) in oocytes with microarray studies and the genes that were most abundant in oocytes and polar bodies in this study. The inverse correlation is found with genes that were significantly upregulated in the reference compared to this dataset.

Table 1: 22 oocytes and sibling PBs were divided into eight sequencing reactions

| The 22 oocytes and sibling PBs were divided into 8 sequencing reactions | Number of mappable reads | Number of genes found |
|---|--------------------------|-----------------------|
| Pool of 10 oocytes 1 | 5,363,802 | 9,765 |
| Pool of 10 oocytes 2 | 1,731,513 | 7,523 |
| Pool of 10 polar bodies 1 | 326,176 | 1,053 |
| Pool of 10 polar bodies 2 | 207,303 | 2,224 |
| Single oocyte 1 | 3,692,168 | 8,681 |
| Single oocyte 2 | 14,573,567 | 11,881 |
| Single polar body 1 | 904,103 | 2,293 |
| Single polar body 2 | 376,441 | 3,515 |
| | | |
| Total of all oocyte samples | 25,361,050 | 12,708 |
| Total of all polar body samples | 1,814,023 | 5,431 |
| Total of all samples | 27,175,073 | 12,883 |

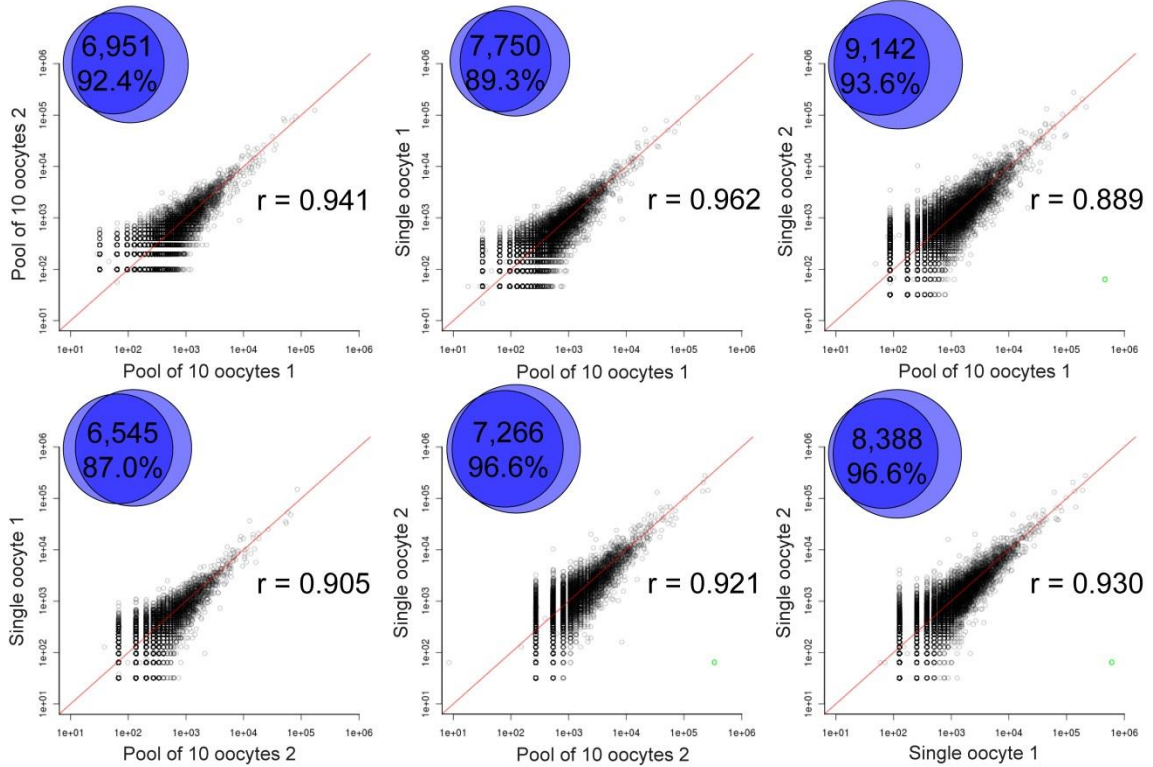
SUPPLEMENTAL INFORMATION

Supplemental Figure 1: Oocyte gene expression extrapolation.



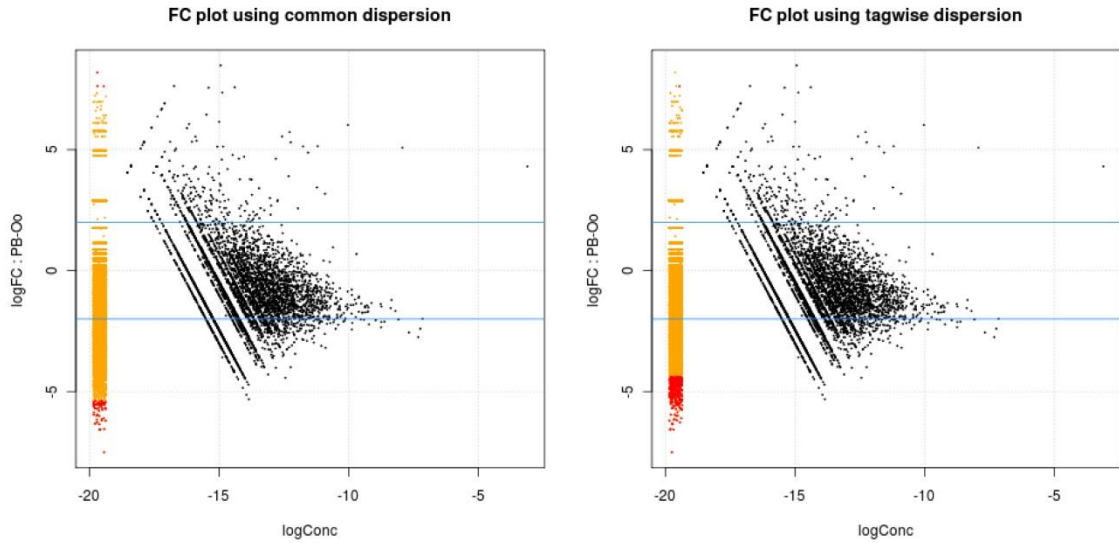
There is a strong correlation between the sequencing effort and the number of genes detected in the oocytes samples showing a diminishing return suggesting the total number of mRNA's expressed in human oocytes as less than 16,000 genes.

Supplemental Figure 2: Pair-wise comparison of all oocytes samples.



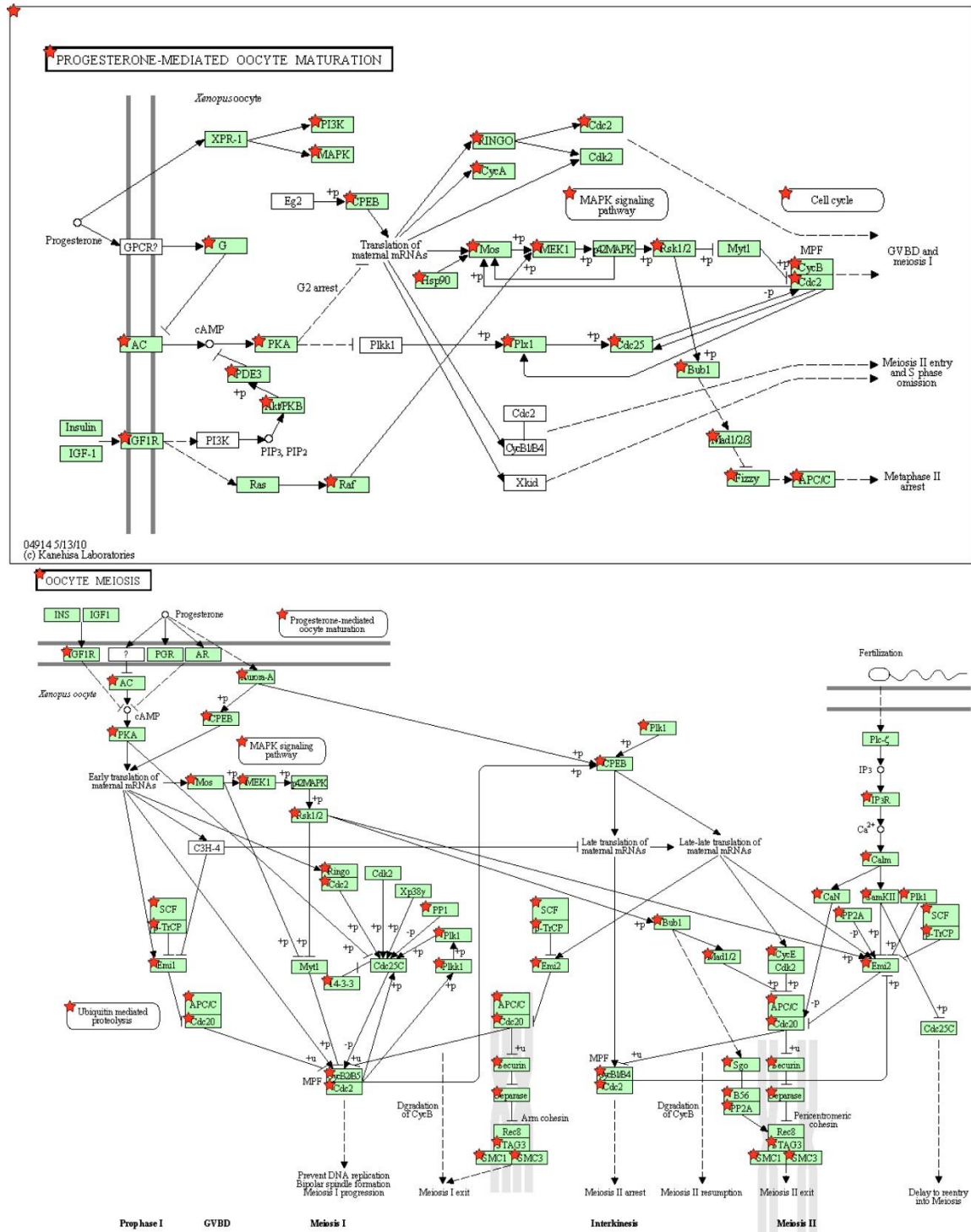
Each Venn diagram represents the gene overlap of every possible oocyte pair-wise comparison. The number in the Venn diagram is the total number of genes found in both samples and the percentage represents the fraction of genes from the smaller dataset that are also found in the larger dataset. The counts for each gene were normalized to the total size of the library and a Pearson pair-wise comparison was done on all genes found in both datasets. The red line is a Pearson correlation of 1 and the scatterplot shows how all the genes deviate from that line. The green circle in Single oocyte 2 (ANTXR2, see text) was excluded from the Pearson correlation as an outlier and the resulting r values show a very strong correlation between the expression level in one sample versus the other for all 6 pair-wise comparisons.

Supplemental Figure 3: Test for differential gene expression between oocytes and polar bodies.



No genes are differentially expressed in oocytes or polar bodies when they are also expressed in the other sample. The smear plot on the left side of each graph represents the genes that are only found in one of the samples. As expected there are very few genes found only in the polar bodies (greater than zero) and a maximum of three genes are deemed significant (red dots) depending on the dispersion method used. Many more genes are only found in the oocyte samples and a number of these are found to be significantly enriched.

Supplemental Figure 4: Select KEGG pathway maps.



KEGG maps of select pathways that were selectively enriched using all the genes detected in all oocytes using DAVID (Huang Da, et al., 2009). Genes with red stars are those that were detected in oocytes.

Supplemental Table 1: The 279 genes that are expressed in all four paired samples

| Gene Name | Chromosome | Oocyte Rank | Polar Body Rank | Oocyte Geometric Mean | Polar Body Geometric Mean |
|-----------|------------|-------------|-----------------|-----------------------|---------------------------|
| C13orf23 | 13 | 1 | 8 | 14144.3 | 1776.5 |
| WEE2 | 7 | 2 | 4 | 13580.5 | 2894.6 |
| DNMT1 | 19 | 3 | 3 | 12402 | 3826.4 |
| NLRP4 | 19 | 4 | 5 | 10607.1 | 2149 |
| KPNA7 | 7 | 5 | 6 | 6682 | 1860.7 |
| OTX2 | 14 | 6 | 10 | 4993.6 | 1446.7 |
| FAM46C | 1 | 7 | 7 | 4614.7 | 1796.7 |
| PTTG1 | 5 | 8 | 26 | 4530.5 | 825.2 |
| DLGAP5 | 14 | 9 | 19 | 4195 | 955.1 |
| TCL1A | 14 | 10 | 12 | 4117 | 1327.5 |
| HIST1H1A | 6 | 11 | 16 | 4061.6 | 1017 |
| UBB | 17 | 12 | 22 | 3558.9 | 893.5 |
| ZFAND2A | 7 | 13 | 11 | 3369.3 | 1419.4 |
| GEMIN5 | 5 | 14 | 9 | 3331.5 | 1492.6 |
| ANTXR2 | 4 | 15 | 1 | 3262.8 | 87920.6 |
| HIST1H4H | 6 | 16 | 36 | 3216.6 | 661.4 |
| HSPA8 | 11 | 18 | 35 | 2601.4 | 665.5 |
| BOD1 | 5 | 19 | 20 | 2592.1 | 925.2 |
| FN1 | 2 | 20 | 13 | 2548.4 | 1073.1 |
| AURKA | 20 | 21 | 37 | 2431.4 | 660.8 |
| GDF9 | 5 | 22 | 32 | 2368.6 | 682.9 |
| ACTL8 | 1 | 23 | 14 | 2364.6 | 1057.5 |
| MTMR3 | 22 | 24 | 15 | 2332 | 1029.3 |
| SIN3A | 15 | 25 | 39 | 2269.6 | 647.3 |
| RGS2 | 1 | 26 | 28 | 2241.7 | 811 |
| ZP2 | 16 | 27 | 18 | 2237.3 | 966.7 |
| CSDE1 | 1 | 28 | 24 | 2232.4 | 892.6 |
| TUBBP5 | 9 | 29 | 29 | 2220.9 | 804.8 |
| NLRP11 | 19 | 30 | 68 | 2184.3 | 399.2 |
| H3F3B | 17 | 32 | 49 | 2033.3 | 490.3 |
| TET3 | 2 | 33 | 31 | 1966.4 | 724.7 |
| NLRP9 | 19 | 34 | 23 | 1951.9 | 893.2 |
| TNRC6B | 22 | 35 | 21 | 1689.9 | 918.9 |
| TIAM1 | 21 | 36 | 30 | 1669.4 | 729.7 |
| FTL | 19 | 37 | 25 | 1623.7 | 842.6 |
| GIT2 | 12 | 38 | 34 | 1604.3 | 673.6 |
| ZP3 | 7 | 40 | 38 | 1552.1 | 647.7 |
| CDK7 | 5 | 42 | 74 | 1504.5 | 383.2 |
| TUBA1C | 12 | 44 | 42 | 1453 | 560.9 |
| CNNM2 | 10 | 45 | 43 | 1440.2 | 559.3 |
| MED13 | 17 | 47 | 118 | 1422.4 | 264.1 |
| NLRP13 | 19 | 48 | 203 | 1422.1 | 175.3 |
| UCHL1 | 4 | 49 | 47 | 1387.7 | 492.2 |
| NPC1 | 18 | 51 | 44 | 1371.6 | 546.2 |
| BMP6 | 6 | 53 | 67 | 1356.2 | 400.3 |
| LIMA1 | 12 | 54 | 72 | 1351 | 385 |
| ESRP1 | 8 | 55 | 33 | 1304.3 | 679.3 |
| GAB1 | 4 | 56 | 91 | 1263.1 | 323.1 |

| | | | | | |
|----------|----|-----|-----|--------|-------|
| CKAP5 | 11 | 57 | 82 | 1255.6 | 341.4 |
| NLRP2 | 19 | 59 | 73 | 1208.3 | 384.1 |
| TUBA4A | 2 | 60 | 41 | 1182.5 | 604.1 |
| ALAS1 | 3 | 62 | 59 | 1152.7 | 410.3 |
| FAM13A | 4 | 65 | 110 | 1145.1 | 279.6 |
| MXD1 | 2 | 68 | 56 | 1132.1 | 433.8 |
| DPPA3 | 12 | 69 | 61 | 1117.6 | 406.2 |
| CDH3 | 16 | 71 | 46 | 1039 | 492.3 |
| ADD3 | 10 | 72 | 53 | 1027.2 | 441.8 |
| B3GNT2 | 2 | 73 | 105 | 1021.1 | 287.1 |
| TMEM2 | 9 | 74 | 114 | 1019.1 | 271.4 |
| SYNE2 | 14 | 76 | 66 | 1014 | 402 |
| TUBB8 | 10 | 77 | 106 | 1003.2 | 286.9 |
| PLAC1L | 11 | 78 | 156 | 998.1 | 212.2 |
| KIAA0922 | 4 | 79 | 222 | 990.9 | 163.4 |
| NEO1 | 15 | 80 | 98 | 973.1 | 307.1 |
| PPRC1 | 10 | 82 | 108 | 960.6 | 284 |
| UNC13C | 15 | 84 | 62 | 918.1 | 406.2 |
| ADAR | 1 | 86 | 55 | 902.1 | 434.2 |
| TAX1BP1 | 7 | 87 | 48 | 898.8 | 491.1 |
| ARIH1 | 15 | 90 | 130 | 870.5 | 245.3 |
| SH2D4A | 8 | 92 | 128 | 865.8 | 245.7 |
| PDE8B | 5 | 96 | 60 | 836.2 | 407.7 |
| ZMYM2 | 13 | 100 | 75 | 801.8 | 373.5 |
| CNN3 | 1 | 101 | 157 | 793.1 | 208.8 |
| PIK3C2A | 11 | 104 | 194 | 786.4 | 181.3 |
| C10orf18 | 10 | 105 | 158 | 782.7 | 208.5 |
| EIF4G2 | 11 | 106 | 103 | 777.8 | 289.8 |
| NUP153 | 6 | 107 | 69 | 777.5 | 389.5 |
| SLC29A1 | 6 | 109 | 40 | 772.4 | 605.4 |
| NLRP5 | 19 | 110 | 76 | 766.9 | 368.5 |
| RMRP | 9 | 111 | 17 | 762.4 | 998 |
| BCL2L10 | 15 | 112 | 84 | 762.4 | 337.7 |
| NUMB | 14 | 114 | 136 | 748.5 | 238.8 |
| IPO8 | 12 | 115 | 51 | 748.3 | 449.1 |
| DTL | 1 | 125 | 228 | 729.4 | 152.6 |
| TNRC6A | 16 | 128 | 77 | 725.3 | 361 |
| ODC1 | 2 | 129 | 90 | 724.5 | 326 |
| ERBB4 | 2 | 130 | 134 | 722.2 | 243.3 |
| CCNB3 | X | 132 | 247 | 719.6 | 137.5 |
| TPT1 | 13 | 135 | 109 | 715.7 | 283.9 |
| ARID1A | 1 | 138 | 141 | 704.1 | 228.2 |
| AMOT | X | 139 | 123 | 699.9 | 251.3 |
| CPEB4 | 5 | 147 | 54 | 679.7 | 438.7 |
| PREPL | 2 | 150 | 83 | 671.3 | 338.1 |
| GYG1 | 3 | 152 | 58 | 668.1 | 425.2 |
| C11orf40 | 11 | 156 | 196 | 657.5 | 181.3 |
| PCGF1 | 2 | 162 | 112 | 645.3 | 279.4 |
| FMN1 | 15 | 164 | 85 | 643.4 | 335.7 |
| SLC6A5 | 11 | 165 | 111 | 640.4 | 279.6 |
| CNBP | 3 | 174 | 81 | 623.5 | 341.4 |
| BAZ2A | 12 | 176 | 80 | 614.9 | 341.6 |

| | | | | | |
|-----------|----|-----|-----|-------|--------|
| CRKL | 22 | 180 | 148 | 610 | 215.8 |
| DAAM1 | 14 | 181 | 64 | 608.8 | 403.6 |
| LAMB1 | 7 | 184 | 216 | 601.1 | 168.6 |
| CDK12 | 17 | 185 | 97 | 596.8 | 307.5 |
| UHRF1 | 19 | 187 | 79 | 589 | 342.5 |
| WHSC1 | 4 | 188 | 175 | 588.4 | 196 |
| ACLY | 17 | 189 | 185 | 585.6 | 188 |
| RNF122 | 8 | 193 | 115 | 580.3 | 266.9 |
| MSL2 | 3 | 194 | 236 | 579.8 | 145.3 |
| GPR37 | 7 | 201 | 87 | 568.3 | 326.9 |
| DDX3X | X | 202 | 170 | 566.4 | 200.3 |
| TRIO | 5 | 203 | 252 | 564.8 | 137.4 |
| AKAP11 | 13 | 207 | 274 | 555.9 | 115.6 |
| TACC3 | 4 | 208 | 135 | 555.2 | 241.4 |
| ALKBH5 | 17 | 210 | 57 | 554.2 | 430.3 |
| ZP1 | 11 | 211 | 167 | 553.7 | 200.6 |
| EIF4ENIF1 | 22 | 212 | 71 | 552.6 | 387.2 |
| RPA1 | 17 | 214 | 241 | 550.6 | 141.5 |
| MED13L | 12 | 215 | 131 | 548.3 | 245.3 |
| HSP90AA1 | 14 | 216 | 63 | 548.2 | 404.2 |
| PITRM1 | 10 | 219 | 88 | 543.7 | 326.8 |
| KPNA2 | 17 | 223 | 171 | 539.7 | 200.2 |
| CLDN12 | 7 | 225 | 160 | 536.3 | 206.3 |
| SLC43A3 | 11 | 228 | 147 | 523.5 | 215.9 |
| ZHX3 | 20 | 230 | 174 | 521.8 | 196.4 |
| HABP2 | 10 | 233 | 89 | 516.2 | 326.7 |
| MTUS1 | 8 | 243 | 176 | 509.4 | 194.8 |
| ZNF215 | 11 | 244 | 104 | 506.7 | 289.7 |
| LARP1 | 5 | 245 | 121 | 506.7 | 256.8 |
| USP2 | 11 | 247 | 45 | 506.2 | 505.5 |
| DNAJC13 | 3 | 248 | 209 | 506 | 173.4 |
| CKS1B | 1 | 252 | 165 | 502 | 200.7 |
| SLC39A10 | 2 | 258 | 161 | 494.4 | 200.8 |
| CUL5 | 11 | 262 | 220 | 487.7 | 163.5 |
| CAPZA1 | 1 | 266 | 100 | 484.3 | 300.4 |
| MAEL | 1 | 267 | 169 | 484.3 | 200.4 |
| PADI6 | 1 | 270 | 107 | 479.7 | 284 |
| PRPF8 | 17 | 281 | 219 | 470 | 163.5 |
| TSC22D2 | 3 | 282 | 260 | 469.8 | 127.8 |
| MEPCE | 7 | 295 | 163 | 453.4 | 200.8 |
| MED16 | 19 | 296 | 50 | 452.8 | 462.5 |
| DMXL1 | 5 | 297 | 240 | 451.2 | 141.5 |
| CMTM4 | 16 | 298 | 186 | 450.7 | 188 |
| ZFH3 | 16 | 301 | 280 | 446.2 | 107.5 |
| PAXIP1 | 7 | 302 | 95 | 445.7 | 308.6 |
| LAD1 | 1 | 305 | 122 | 442.2 | 252.4 |
| RND1 | 12 | 307 | 96 | 441 | 307.8 |
| RAPH1 | 2 | 308 | 2 | 440.8 | 6117.7 |
| ARG2 | 14 | 310 | 191 | 439.9 | 184.9 |
| PDPN | 1 | 316 | 133 | 435.7 | 243.6 |
| AMBRA1 | 11 | 320 | 117 | 431.3 | 264.1 |
| SRCAP | 16 | 322 | 159 | 429.1 | 208.2 |

| | | | | | |
|-----------|----|-----|-----|-------|-------|
| SF3A3 | 1 | 334 | 232 | 419.9 | 152.3 |
| KALRN | 3 | 335 | 78 | 419.9 | 355 |
| SKP1 | 5 | 339 | 152 | 418.6 | 215.5 |
| FAM193A | 4 | 344 | 182 | 411.8 | 191.6 |
| CTTNBP2NL | 1 | 345 | 198 | 411.7 | 177.4 |
| PRPF18 | 10 | 347 | 205 | 410.6 | 175.3 |
| SLC7A8 | 14 | 349 | 238 | 409.2 | 141.7 |
| QKI | 6 | 355 | 129 | 406 | 245.3 |
| MLL2 | 12 | 359 | 153 | 401.7 | 215.4 |
| HIPK2 | 7 | 360 | 137 | 401.4 | 238.7 |
| ITPR1 | 3 | 365 | 119 | 396.7 | 262.6 |
| ASPM | 1 | 367 | 27 | 394.9 | 817.1 |
| BHLHE40 | 3 | 372 | 149 | 392.8 | 215.7 |
| RUNX1T1 | 8 | 376 | 249 | 391.4 | 137.4 |
| SUN1 | 7 | 380 | 189 | 388.7 | 186.4 |
| STARD7 | 2 | 385 | 93 | 381.6 | 314.2 |
| CSF1R | 5 | 387 | 162 | 380.8 | 200.8 |
| RNF10 | 12 | 391 | 183 | 378.7 | 188.7 |
| GBF1 | 10 | 396 | 65 | 376.2 | 402 |
| AMD1 | 6 | 407 | 180 | 373 | 191.7 |
| EPAS1 | 2 | 413 | 146 | 366.4 | 215.9 |
| TTL4 | 2 | 432 | 120 | 355.6 | 258.6 |
| WIPF2 | 17 | 438 | 246 | 354.1 | 137.5 |
| GPR137B | 1 | 444 | 52 | 349.8 | 444.9 |
| ESCO2 | 8 | 450 | 351 | 346.3 | 81.4 |
| FIP1L1 | 4 | 456 | 145 | 340.6 | 222.1 |
| CLSTN1 | 1 | 460 | 124 | 340.1 | 251 |
| EBF1 | 5 | 465 | 116 | 336.8 | 264.4 |
| CHD7 | 8 | 478 | 132 | 329.1 | 245.1 |
| DCDC2 | 6 | 479 | 99 | 329 | 300.5 |
| YWHAE | 17 | 484 | 255 | 326.1 | 127.9 |
| NAA11 | 4 | 486 | 221 | 325.2 | 163.4 |
| COG5 | 7 | 487 | 276 | 324.9 | 115.5 |
| BAIAP2 | 17 | 489 | 94 | 324.6 | 312.4 |
| INTS9 | 8 | 508 | 187 | 315.8 | 186.7 |
| C16orf72 | 16 | 510 | 177 | 315.4 | 194.8 |
| DPF2 | 11 | 520 | 193 | 311.1 | 181.4 |
| FZD3 | 8 | 525 | 261 | 307.4 | 127.8 |
| ELAVL1 | 19 | 526 | 300 | 306.8 | 97.1 |
| RBM33 | 7 | 528 | 143 | 305.5 | 227.8 |
| IWS1 | 2 | 541 | 140 | 300.3 | 230.8 |
| CETN3 | 5 | 545 | 166 | 298.5 | 200.6 |
| PAIP1 | 5 | 547 | 262 | 298 | 127.8 |
| THAP1 | 8 | 550 | 309 | 297.4 | 97 |
| METAP2 | 12 | 562 | 275 | 291.5 | 115.6 |
| HHLA2 | 3 | 573 | 213 | 286.3 | 168.8 |
| DCAF5 | 14 | 576 | 188 | 284.6 | 186.5 |
| ANKS1A | 6 | 582 | 168 | 283.3 | 200.4 |
| MON2 | 12 | 590 | 206 | 279.9 | 175.3 |
| TAF5L | 1 | 594 | 144 | 279.2 | 227.8 |
| RMND5A | 2 | 616 | 250 | 274.7 | 137.4 |
| FBLN7 | 2 | 618 | 253 | 273.3 | 128.1 |

| | | | | | |
|----------|----|------|-----|-------|-------|
| C22orf30 | 22 | 634 | 173 | 269.1 | 197.7 |
| TDRD1 | 10 | 635 | 164 | 269 | 200.7 |
| CHEK1 | 11 | 643 | 199 | 265.7 | 177.3 |
| SAP18 | 13 | 646 | 155 | 264.6 | 215.2 |
| SVOPL | 7 | 655 | 190 | 262.7 | 184.9 |
| RC3H1 | 1 | 657 | 251 | 262.1 | 137.4 |
| ZBTB10 | 8 | 658 | 230 | 262 | 152.4 |
| CASC5 | 15 | 661 | 301 | 261.4 | 97.1 |
| EIF1 | 17 | 673 | 281 | 257.7 | 107.5 |
| COPS7B | 2 | 698 | 202 | 252 | 175.3 |
| CALM2 | 2 | 699 | 306 | 251.6 | 97 |
| LRRC8A | 9 | 701 | 231 | 251.4 | 152.3 |
| EIF1B | 3 | 704 | 113 | 250.9 | 274.5 |
| G3BP1 | 5 | 711 | 235 | 248.9 | 145.4 |
| CCDC21 | 1 | 737 | 179 | 244.2 | 191.9 |
| WFDC2 | 20 | 738 | 86 | 243.7 | 334.5 |
| NR3C2 | 4 | 770 | 248 | 236 | 137.4 |
| ARF3 | 12 | 786 | 92 | 233.9 | 314.5 |
| KDM6B | 17 | 797 | 154 | 230.3 | 215.3 |
| COG8 | 16 | 840 | 181 | 220.8 | 191.6 |
| MAGOH | 1 | 846 | 139 | 220.2 | 230.9 |
| GNG12 | 1 | 869 | 215 | 214.8 | 168.6 |
| PSME3 | 17 | 872 | 310 | 214.2 | 97 |
| MYLIP | 6 | 894 | 259 | 210.9 | 127.8 |
| PODXL | 7 | 895 | 208 | 210.8 | 173.4 |
| UNKL | 16 | 909 | 350 | 208.2 | 81.4 |
| DNMT3A | 2 | 915 | 211 | 206.4 | 173.1 |
| RNF20 | 9 | 923 | 101 | 205.5 | 300.3 |
| MCM3AP | 21 | 933 | 267 | 203.7 | 122.2 |
| KTN1 | 14 | 941 | 271 | 203.1 | 115.7 |
| ZMAT2 | 5 | 980 | 234 | 198.5 | 145.5 |
| TBC1D1 | 4 | 987 | 151 | 197.1 | 215.6 |
| BAT2 | 6 | 1031 | 200 | 192.3 | 176.9 |
| FIGNL1 | 7 | 1037 | 126 | 191.6 | 248.1 |
| ACSL3 | 2 | 1042 | 125 | 190.8 | 248.4 |
| PPP2R5C | 14 | 1082 | 308 | 185.6 | 97 |
| HIP1R | 12 | 1088 | 195 | 184.6 | 181.3 |
| YES1 | 18 | 1096 | 204 | 183.5 | 175.3 |
| LASS2 | 1 | 1122 | 207 | 179.7 | 174.8 |
| JAM3 | 11 | 1148 | 257 | 176.4 | 127.9 |
| PRDM4 | 12 | 1221 | 210 | 166.2 | 173.3 |
| PDLIM1 | 10 | 1223 | 315 | 166.1 | 96.8 |
| TARDBP | 1 | 1237 | 305 | 164.4 | 97 |
| AGK | 7 | 1293 | 258 | 159.4 | 127.8 |
| CGREF1 | 2 | 1357 | 302 | 153.1 | 97.1 |
| TBPL1 | 6 | 1366 | 229 | 152.6 | 152.5 |
| SMARCA5 | 4 | 1377 | 303 | 151.4 | 97 |
| ETV5 | 3 | 1422 | 102 | 146.3 | 290.8 |
| TOX2 | 20 | 1460 | 150 | 143.1 | 215.7 |
| FAM54A | 6 | 1469 | 172 | 141.6 | 197.7 |
| DSTN | 20 | 1483 | 237 | 140.4 | 145.2 |
| HMGCR | 5 | 1489 | 313 | 139.9 | 96.9 |

| | | | | | |
|-----------|----|------|-----|-------|-------|
| EBF2 | 8 | 1546 | 225 | 134.8 | 161.1 |
| FANCI | 15 | 1562 | 224 | 133.3 | 161.2 |
| CFL1 | 11 | 1576 | 142 | 132.5 | 228 |
| ZNF761 | 19 | 1580 | 312 | 132.3 | 96.9 |
| KIAA1598 | 10 | 1679 | 244 | 124.8 | 137.6 |
| CIR1 | 2 | 1702 | 192 | 123.3 | 181.4 |
| PCID2 | 13 | 1739 | 256 | 120.8 | 127.9 |
| ERMP1 | 9 | 1846 | 277 | 115.3 | 115.5 |
| EIF5 | 14 | 1895 | 278 | 111.8 | 115.4 |
| SLC25A46 | 5 | 1974 | 304 | 107 | 97 |
| SECISBP2L | 15 | 2007 | 307 | 104.7 | 97 |
| TCL1B | 14 | 2134 | 227 | 99.5 | 158.1 |
| UBLCP1 | 5 | 2235 | 214 | 95 | 168.6 |
| WWC2 | 4 | 2288 | 218 | 93 | 163.8 |
| CHN2 | 7 | 2291 | 283 | 92.9 | 107.2 |
| KLHDC3 | 6 | 2312 | 273 | 92 | 115.6 |
| CNTD2 | 19 | 2392 | 184 | 87.8 | 188.5 |
| CTDSPL | 3 | 2509 | 279 | 83.2 | 115.4 |
| UIMC1 | 5 | 2625 | 226 | 78.4 | 160.7 |
| MCM7 | 7 | 2684 | 245 | 76.5 | 137.6 |
| ZC3HC1 | 7 | 3008 | 282 | 66.4 | 107.3 |
| TMEM132D | 12 | 3407 | 223 | 55.9 | 163.4 |
| ADAT1 | 16 | 3442 | 272 | 55.3 | 115.6 |

**Appendix II: Transcriptome variance in single oocytes within, and between,
genotypes**

Adrian Reich, Nicola Neretti, Richard N. Freiman, and Gary M. Wessel

Mol. Reprod. Dev (2012),79, 502-503.

CONTRIBUTION

I conducted all experiments and analyses.

ABSTRACT

The zygote of sexually reproducing organisms contains a combination of parental genomes, and all subsequent cells of the embryo are derived from this original genotype. Although clonal, it is not known how much genetic variation exists in progeny of this original cell, or between cells of the same lineage resulting from this zygote. Oocytes in mammals, especially humans, have prolonged developmental histories and each may be quite different in terms of gene expression. It is clear that oocyte quality can differ significantly within a cohort, and the variation in early developmental success from each oocyte can be dramatic. Oocyte quality is ultimately best measured by the success of the embryo, but other features, such as normalcy of the mRNA population, may be important criteria to identify such potential. Here, we test the variation in steady-state levels of mRNAs in mouse oocytes to establish a baseline of “normal” variation, and compare it mRNA levels of individual oocytes of poor quality. We sequenced to saturation the mRNA from five wild-type oocyte samples (three individual oocytes, and two pools of five oocytes each from two wild-type mice) and 16 Taf4b-deficient oocyte samples (12 individual oocytes and four pools of 5 or 10 oocytes each from two Taf4b-deficient mice). The Taf4b-deficient mice are known to have oocytes that appear morphologically normal (Fig.1a,b), but are of poor quality with regard to successful embryogenesis. This genotype was selected as a model for human premature ovarian insufficiency (POI; Lovasco *et al.*, 2010). Taf4b-null animals are viable as adults, but the oocytes they make die prematurely in adults, leading to a POI phenotype, and any oocytes that mature and are fertilized do not develop past the two- to four-cell stage (Falender *et al.*, 2005; Lovasco *et al.*, 2010).

RESULTS AND DISCUSSION

The hypothesis tested here is that the transcriptome of the Taf4b-deficient oocyte differs significantly from that of the wild-type oocyte. To properly assess this, we also needed to determine the variance between individual oocytes to ascribe significance to the comparison. This dataset was generated by high-throughput DNA sequencing following transcriptome amplification (Reich *et al.*, 2011) and compared within and between genotypes to determine the variance. To test the fidelity of the amplification process for this protocol, prior to and independent of high-throughput DNA sequencing, oocytes from a wild-type mouse were isolated and pooled before lysing. Following DNase treatment, one oocyte-equivalent was isolated and the cDNA library was synthesized. The resulting library was diluted 100 times, the approximate volume of a single polar body, which is important if a polar body were to be used to determine the oocyte quality without harming the oocyte (Reich *et al.*, 2011). Three samples from this pool were independently amplified, and each technical replicate was tested by quantitative RT-PCR (qPCR) as a measure of the fidelity of the amplification procedure (Reich *et al.*, 2011). Overall, low technical variation was detected, providing confidence in the protocol (Fig. 1c). We do not know what kinds of bias the amplification procedure may have, but based on these results, the amplification appears to be consistent. The starting material for a polar body is so limiting, however, that even with this cDNA amplification, qPCR is only able to consistently amplify some transcripts—most rare transcripts have high C_t values, thus the sensitivity of sequencing is therefore preferred.

In order to test the inter- and intra-genotype variation, we collected oocytes from Taf4b-null and wild-type oviducts after ovulation, mechanically and enzymatically stripped of all granulosa cells, and processed the cells for cDNA synthesis and amplification for sequencing as described (Reich *et al.*, 2011). The libraries were sequenced on a HiSeq 2000, and the reads were mapped to the mouse genome (mm9) using TopHat (Trapnell *et al.*, 2009), yielding an average of 219,207 (SD 138,190) mappable reads per sample. These reads were tested for differential

expression using edgeR (Robinson and Smyth, 2007). A total of 11,373 genes were detected across all 21 samples that were also above a filter threshold of greater than 20 raw counts across all 21 libraries, and a total of 3,242 genes were differentially expressed with a false discovery rate (FDR) of <0.05 (Supplemental Table 1). A large number of genes are upregulated in the Taf4b mutant samples, including 3,465 genes undetected in the wild-type samples; 1,037 of these genes achieve significance (Supplemental Table 1 and Fig. 1d). The gene-by-gene average of the RPKM (reads per kilobase of transcript per million mapped reads) from one genetic background is very similar to the average RPKM from another background (Fig. 1d). The log-transformed standard deviations of the RPKMs of wild-type and knockout samples (Supplemental Fig. 2) closely mirrors the graph of the means of the RPKMs (Fig. 1d), suggesting: (a) as genes become more abundant, the variation increases, (b) different genomic backgrounds have similar rates of variation, and (c) assuming the qPCR results from (Fig. 1c) represent the technical variability of all genes, then any bias introduced by the amplification process appears significantly less than the biological variability within a population.

Although the gene-by-gene standard deviation of the RPKM scales linearly with the abundance of the gene, suggesting that samples within a background are similar, we compared how the entire gene set of a sample compared with another sample within the same genetic background and also across backgrounds. The five samples isolated from the two wild-type mice (WT1 and WT2) and 16 samples from the two Taf4b mutant mice (KO1 and KO2) clearly segregate by genotype into two main groups; within a group, the samples segregate by mouse to a great degree (Fig. 2 and Supplemental Fig. 1). Only one of the wild-type samples clustered together with the knockout samples, although the distance between this wild-type sample and all knockout samples (cophenetic distance) is larger than any of the other samples within this group; this indicates that its transcriptional profile is intermediate between the two genotypes.

CONCLUSIONS

We conclude that the biological variability of transcriptomes can be quantified between single cells within a genotype, and the comparison between genotypes can reveal genes that are differentially expressed in a robust manner. This approach may help reveal oocyte quality by use of the polar body metric without harm to the oocyte (Reich *et al.*, 2011).

DATA AVAILABILITY

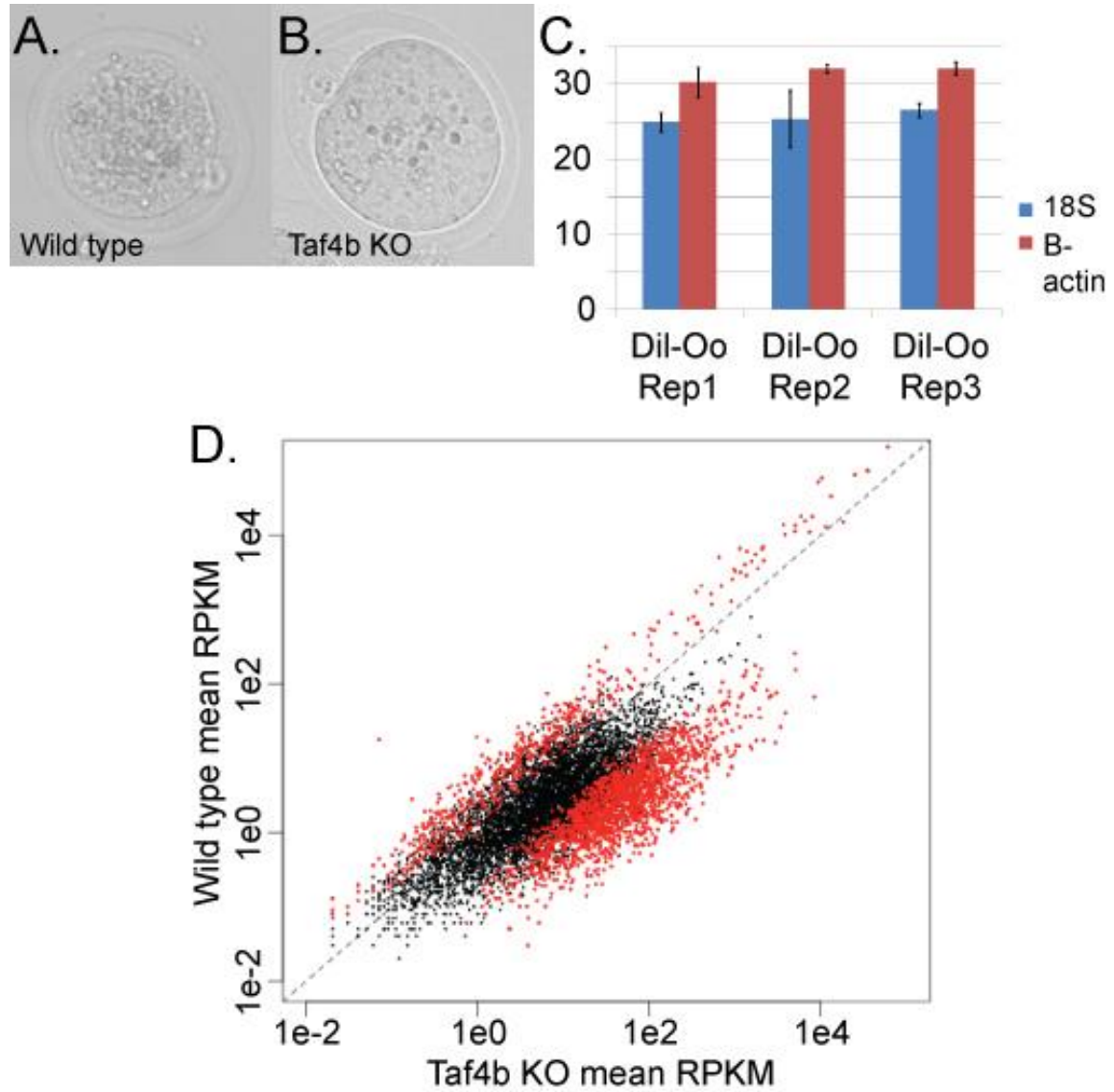
The sequence reported in this paper has been deposited in the GenBank database (NCBI BioProject no. PRJNA236019).

REFERENCES

- Falender, A.E., Shimada, M., Lo, Y.K., and Richards, J.S. (2005). TAF4b, a TBP associated factor, is required for oocyte development and function. *Developmental biology* 288, 405-419.
- Lovasco, L.A., Seymour, K.A., Zafra, K., O'Brien, C.W., Schorl, C., and Freiman, R.N. (2010). Accelerated ovarian aging in the absence of the transcription regulator TAF4B in mice. *Biology of reproduction* 82, 23-34.
- Reich, A., Klatsky, P., Carson, S., and Wessel, G. (2011). The transcriptome of a human polar body accurately reflects its sibling oocyte. *Journal of Biological Chemistry* 286, 40743-40749.
- Reich, A., Neretti, N., Freiman, R.N., and Wessel, G.M. (2012). August 2012 cover image. *Molecular Reproduction & Development* 79, C1.
- Robinson, M.D., and Smyth, G.K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* 23, 2881-2887.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105-1111.

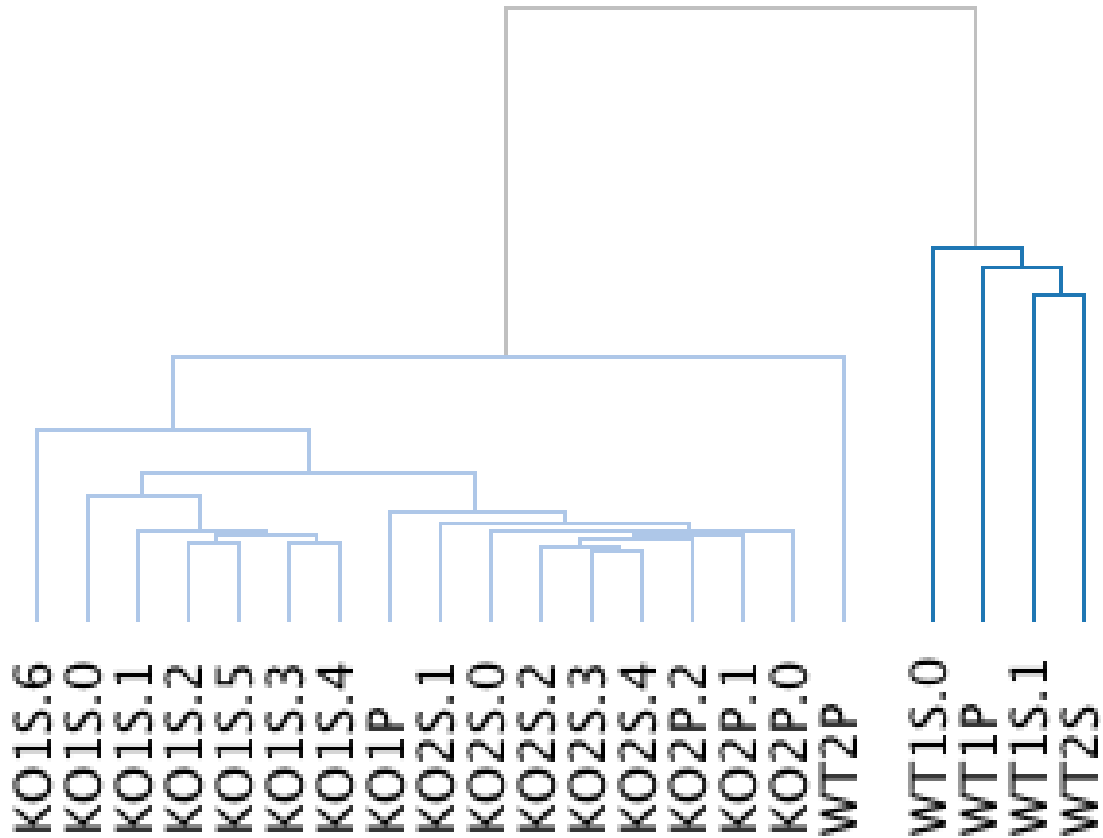
FIGURES

Figure 1: KO and WT morphological and molecular comparisons.



A,B: Photos of sequenced wild-type and Taf4b-knockout oocytes. **C:** qPCR was performed in triplicate on the amplified library for beta-actin and 18S rRNA, and the C_t values are reported here. Dil-Oo refers to oocyte samples diluted 100-fold (see text). **D:** Knockout versus wild-type means: all genes above the filter are plotted with respect to the log of the average RPKM across all wild-type or knockout samples. Genes with a significant p -value and an FDR < 0.05 are shown in red.

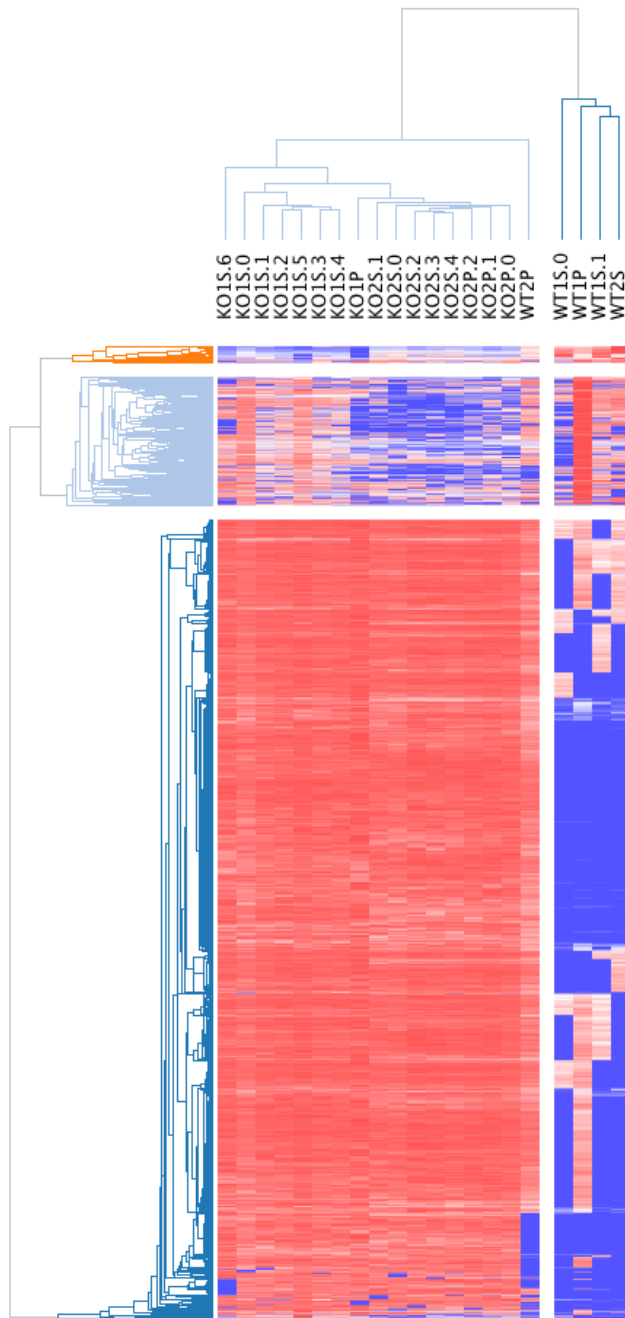
Figure 2: Dendrogram of 5 wild-type samples and 16 Taf4b-knockout samples.



This shows the high sample-to-sample relatedness within genotypes and even within individual mice (see Supplemental Fig. 1 for full details). Each sample name also describes the type of sample and the biological replicate of that type from the same mouse. S = single oocyte and P = pool of oocytes; KO2S.5 = knockout mouse #2, single oocyte biological replicate #5.

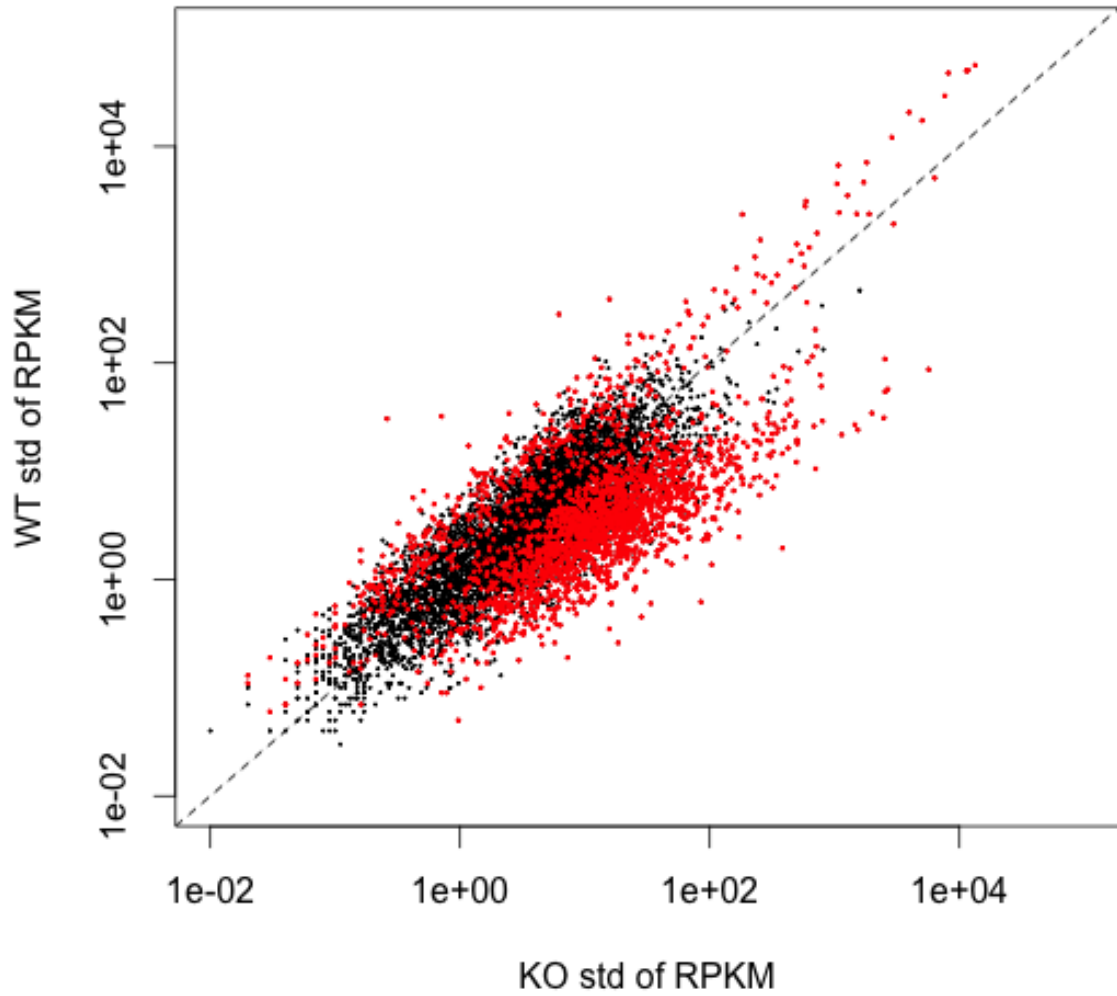
SUPPLEMENTAL INFORMATION

Supplemental Figure 1: Heatmap comparing significant genes between KO and WT.



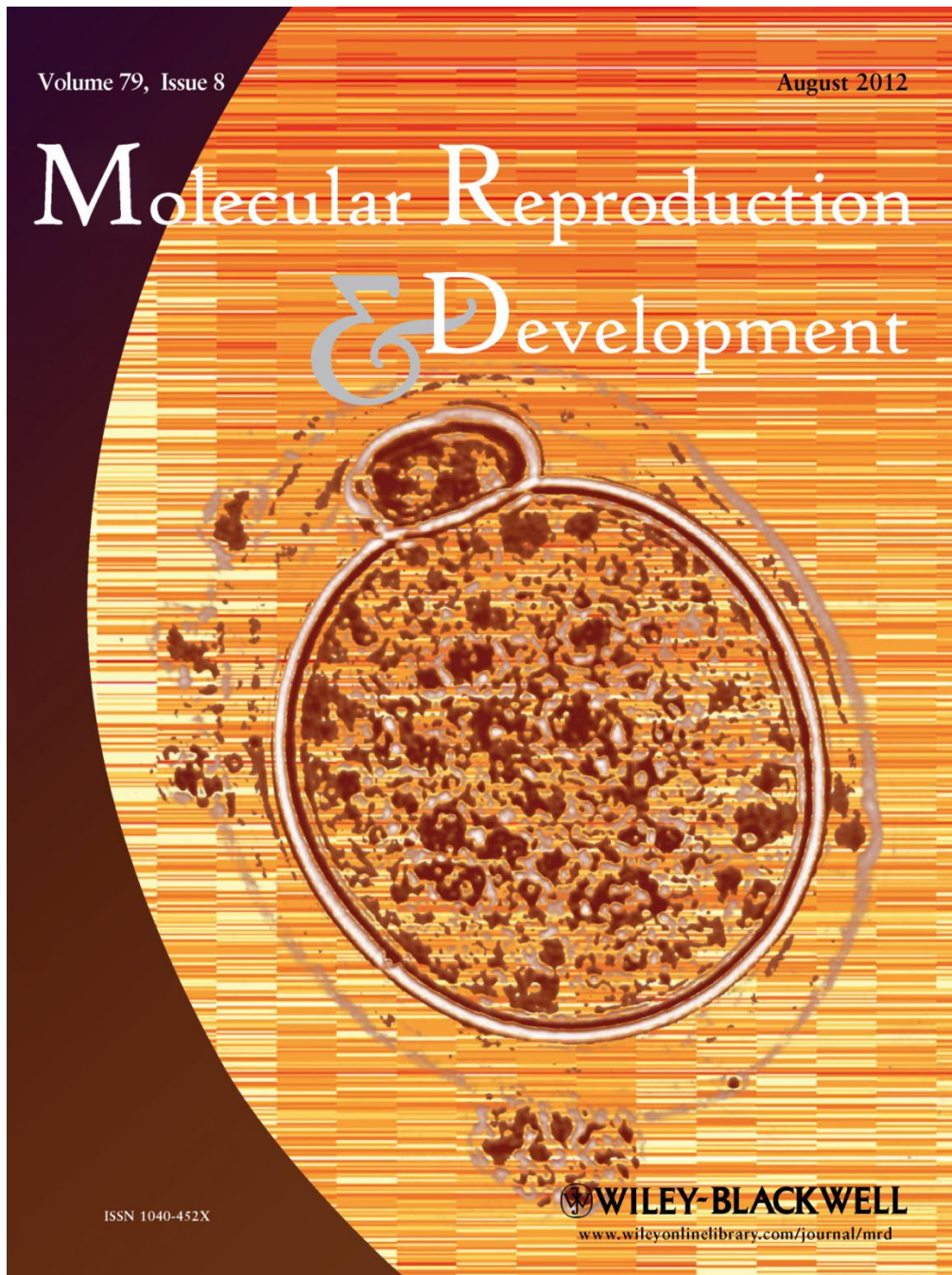
The 3,242 genes were selected as significantly different in the two genotypes ($FDR \leq 0.05$). Clustering was done on both rows and columns by using $(1 - \text{Pearson correlation})$ as metrics and the average as linkage. Blue corresponds to values less than the mean-value of a given gene across all samples, red corresponds to values greater than the mean value of a given gene across all samples.

Supplemental Figure 2: Knockout versus wildtype standard deviations.



Same data as in Supplementary Table 1 and Figure 1D, here expressed as comparison of log of standard deviations.

Supplemental Figure 3: A single mouse oocyte superimposed on a heatmap of differentially expressed genes.



A single mouse oocyte whose transcriptome was amplified, sequenced, and compared to individual transcriptomes of twenty other oocytes from two different genotypes. The oocyte is shown within a heat map from the gene-by-gene expression analysis of all detected transcripts from these oocytes. Reich et al. (2012) and Appendix II test the transcript variance among individual oocytes and determine that they cluster by genotype.

Supplemental Table 1: Database of KO vs WT

Please refer to the online copy of Supplemental Table 1 here: DOI: 10.1002/mrd.22061 The table is too large to fit within this document.

All genes with greater than 20 reads across all 21 libraries (11,373) were tested for differential expression with edgeR using TMM normalization, resulting in 3,243 genes with a P-value and FDR less than 0.05. The lengths of all isoforms of all genes were downloaded from ensembl.org and averaged to generate a typical transcript length. This length and the raw counts from the RNA-seq were used to generate the RPKM measurement for all genes. Even though two different normalization metrics were used (TMM and RPKM) for different parts of the data analysis, the two methods agree on which background is enriched for a particular gene, but not always on the scale of the enrichment.

**Appendix III: PIWI proteins and PIWI-interacting RNAs function in Hydra
somatic stem cells**

Celina E Juliano, **Adrian Reich**, Na Liu, Jessica Götzfried, Mei Zhong, Selen Uman,
Robert A Reenan, Gary M Wessel, Robert E Steele, Haifan Lin

Proceedings of the National Academy of Sciences (2014),*111*(1), 337-342.

CONTRIBUTION

I processed the RNA and constructed the Illumina library; sequenced and assembled the *de novo* transcriptome and conducted all transcriptome-based bioinformatic analyses. This data is found in Figure 3C-H, Supplemental Figure 6 and Supplemental Tables 1-4, which I prepared.

ABSTRACT

PIWI proteins and their bound piRNAs are found in animal germlines and are essential for fertility, but their functions outside of the gonad are not well understood. The cnidarian *Hydra* is a simple metazoan with well-characterized stem/progenitor cells that provides an important new model for analysis of PIWI function. Here we report that *Hydra* has two PIWI proteins, Hywi and Hyli, both of which are expressed in all *Hydra* stem/progenitor cells, but not in terminally differentiated cells. We identified ~15 million piRNAs associated with Hywi and/or Hyli and found that they exhibit the ping-pong signature of piRNA biogenesis. *Hydra* PIWI proteins are strictly cytoplasmic and thus likely act as post-transcriptional regulators. To explore this function, we generated a *Hydra* transcriptome for piRNA mapping. piRNAs map to transposons with a 25- to 35-fold enrichment compared to the abundance of transposon transcripts. By sequencing the small RNAs specific to the interstitial, ectodermal, and endodermal lineages we found that the targeting of transposons appears to be largely restricted to the interstitial lineage. We also identified putative non-transposon targets of the pathway unique to each lineage. Finally we demonstrate that *hywi* function is essential in the somatic epithelial lineages. This comprehensive analysis of the PIWI-piRNA pathway in the somatic stem/progenitor cells of a non-bilaterian animal suggests that this pathway originated with broader stem cell functionality.

INTRODUCTION

PIWI proteins and their bound small PIWI-interacting RNAs (piRNAs) are central players in a regulatory pathway that is essential for germline establishment and maintenance. Loss of PIWI proteins in *Drosophila*, mice, and zebrafish leads to a loss of fertility, due to a disruption in germline stem cell (GSC) formation or maintenance, arrest in meiosis, and other gametogenic defects (Thomson and Lin, 2009). *Piwi* is also expressed outside the germline, largely in various kinds of stem and progenitor cells. For example, *piwi* genes are expressed in the pluripotent stem cells of planarians, sponges, and tunicates, and are required for regeneration (Juliano *et al.*, 2011). *Piwi* expression is also found in hematopoietic stem cells in humans, mesenchymal stem cells in mice, and somatic stem cells in cnidarians and ctenophores (Alie *et al.*, 2011; Seipel *et al.*, 2004; Sharma *et al.*, 2001; Wu *et al.*, 2010). However, detailed analyses have been largely confined to the function of the PIWI-piRNA pathway in the germline and the gonadal somatic cells in a few model bilaterians, with a focus on transposon silencing (Siomi *et al.*, 2011). The potential significance of the pathway in stem cells outside the gonad and on non-transposon sequences is largely unexplored.

Hydra is a morphologically simple multicellular organism belonging to the phylum Cnidaria, which is the sister group to bilaterians (Supplemental Fig. 1a). The adult *Hydra* polyp is composed of three distinct cell lineages: the two epithelial lineages (ectoderm and endoderm) and the interstitial lineage (Fig. 1a). The multipotent interstitial stem cells that support the interstitial lineage give rise to three somatic cell types (nerves, nematocytes, and gland cells) and to germ cells (Fig. 1a) (David, 2012). The epithelial lineages do not have a true stem cell population, but they are mitotic along the entire length of the body column and these progenitor/stem cells are responsible for maintaining the lineage (Holstein *et al.*, 1991). These cells indefinitely self-renew and retain the capability of differentiating into the non-mitotic cells that function in the tentacles and foot. (Fig. 1a). In this study we provide a comprehensive analysis of both PIWI protein

expression and piRNA sequences in *Hydra*, which demonstrates that the PIWI-piRNA pathway has ancient and broadly conserved stem cell functions, including somatic functions.

RESULTS

***Hydra* PIWI proteins, Hywi and Hyli, are expressed in multipotent stem cells.**

Computational searches of the *Hydra* genome (Chapman *et al.*, 2010) revealed four Argonaute proteins: two AGO family proteins (*Hy-ago1* and *Hy-ago2*) and two PIWI family proteins (Supplemental Fig. 1b). The *Hydra* PIWI family proteins were named Hywi and Hyli for their PIWI and PIWI-like orthologs (Supplemental Fig. 1b). We generated polyclonal antibodies against the N-terminal and MID-domains of both Hywi and Hyli and demonstrated their specificity with immunoprecipitation experiments (Supplemental Fig. 1c-g). The antibodies stained numerous cells throughout the body column, but not in the extremities (Fig. 1b,c and Supplemental Fig 2a-c). The restriction of Hywi and Hyli expression to the body column, where the stem/progenitor cells reside, was also seen by immunoblot analysis of body columns and heads (Fig. 1d). Co-labeling with C41 antibody, an interstitial stem cell marker (David *et al.*, 1991), demonstrated that Hywi and Hyli are expressed in interstitial stem cells (Fig. 1e-g and Supplemental Fig. 1d-f). In addition, both Hywi and Hyli are expressed in nematoblast nests, which are nematocyte progenitor cells of the interstitial lineage (Supplemental Fig. 3) (Bosch and David, 1987; David and Murphy, 1977). Hywi and Hyli proteins are diffusely distributed in the cytoplasm of interstitial stem cells and are enriched in punctate foci around the nucleus (Fig. 1h-j). Immuno-electron microscopy demonstrated that both Hywi and Hyli are associated with electron-dense perinuclear structures similar to what is seen in the germlines of several animals, including *Drosophila*, mice, and zebrafish (Fig. 1k and l) (Brennecke *et al.*, 2007; Carmell *et al.*, 2007; Houwing *et al.*, 2008; Houwing *et al.*, 2007; Kuramochi-Miyagawa *et al.*, 2004; Unhavaithaya *et al.*, 2009; Wang *et al.*, 2009).

Hywi and Hyli accumulate in perinuclear granules of epithelial stem/progenitor cells

Hywi and Hyli staining is prominent in the interstitial stem cells and nematoblasts (see Fig. 1e-g; Supplemental Fig. 2d-f; Supplemental Fig. 3), but immunoblotting of *Hydra* that are depleted of the interstitial lineage revealed Hywi and Hyli accumulation outside this lineage (Fig. 2a). For more detailed cell type analysis, we used transgenic animals with lineage specific GFP or DsRed2 expression and dissociated whole animals into single cells for both immunoblotting and immunostaining (Fig. 2b) (Dana *et al.*, 2012; Glauber *et al.*, 2013). Both Hywi and Hyli were detected by immunoblotting in both ectodermal and endodermal cell populations isolated by FACS (Fig. 2c and Supplemental Fig. 4). Furthermore, both Hyli and Hywi proteins accumulate in puncta around the nuclei of ectodermal and endodermal epithelial cells, but do not accumulate significantly elsewhere in the cytoplasm (Fig. 2e-j and Supplemental Fig. 2g-j). Immunostaining experiments revealed that both Hywi and Hyli are absent from the nucleus (e.g. Fig. 1h-j and Fig. 2e-j). This is in contrast to PIWI proteins in *Drosophila* and the mouse, some of which are nuclear and likely act as epigenetic regulators (Aravin *et al.*, 2008; Huang *et al.*, 2013; Sienski *et al.*, 2012). To test if the cytoplasmic localization of Hywi and Hyli in situ is due to antigen masking or low abundance in the nucleus, we analyzed nuclear and cytoplasmic fractions by immunoblotting and found that Hywi and Hyli are apparently exclusively cytoplasmic (Fig. 2d).

Isolation and characterization of *Hydra* piRNAs reveals conserved mechanisms of piRNA biogenesis

To investigate the function of the PIWI-piRNA pathway in *Hydra*, piRNAs bound to Hywi and Hyli were isolated by immunoprecipitation and sequenced (Supplemental Fig. 5a). Analysis of the size distribution revealed that Hywi and Hyli bind piRNAs of different sizes, which is consistent with PIWI proteins in *Drosophila*, mice, and zebrafish (Fig. 3a) (Aravin *et al.*, 2008; Brennecke *et al.*, 2007; Houwing *et al.*, 2008). Over 90% of piRNAs bound to Hywi have a uridine at their 5' end (Supplemental Fig. 5d) and over 80% of piRNAs bound to Hyli have an adenine at their 10th position (Supplemental Fig. 5d). Furthermore, we found a complementary 10-base pair overlap between the 5' ends of Hywi-bound and Hyli-bound piRNAs (Fig. 3b).

These features are identical to the ping-pong signature of biogenesis that was first described in *Drosophila* (Brennecke *et al.*, 2007; Gunawardane *et al.*, 2007) and also observed in mice and zebrafish (Aravin *et al.*, 2008; Aravin *et al.*, 2007; Houwing *et al.*, 2008). Previous sequencing of total *Nematostella vectensis* and *Hydra* RNAs identified putative piRNAs (Grimson *et al.*, 2008; Krishna *et al.*, 2013). Here we have identified bona fide cnidarian piRNAs bound to specific PIWI proteins, thus allowing for comparisons between piRNAs bound to different PIWI proteins. Finally, we show that *Hydra* piRNAs are 2'-O-methylated at their 3' ends similar to bilaterian piRNAs (Supplemental Fig 5e) (Ohara *et al.*, 2007; Saito *et al.*, 2007). Our data definitively demonstrate that Hywi and Hyli participate in ping-pong biogenesis and prove that this mechanism has a deep evolutionary origin in metazoans.

The *Hydra* PIWI-piRNA targets transposon transcripts

The prevailing model posits that ping-pong piRNA biogenesis results in decreased transposon expression due to post-transcriptional processing of transposon RNAs into piRNAs (Brennecke *et al.*, 2007). To test if Hywi and Hyli function in post-transcriptional transposon repression, we first mapped the piRNAs to the *Hydra* genome (Chapman *et al.*, 2010). Approximately 50% of the sequenced piRNAs were mapped to unique sites in the *Hydra* genome. 55-65% of *Hydra* piRNAs map to repeat sequences that were previously identified by RepeatMasker (Supplemental Fig 5b,c). Since the total repeat content in the *Hydra* genome is 57%, this mapped population of piRNAs is not significantly enriched for repeat sequences (Chapman *et al.*, 2010).

To better characterize the piRNA targets in *Hydra* we focused our attention on transcripts that are expressed in the adult. To this end, we sequenced and assembled a *Hydra* transcriptome containing ~27,000 sequences, which we curated to obtain a set of 9,986 transcripts with a significant BLAST (1×10^{-5}) match to the Swiss-Prot database. This allowed for definition of open reading frames and transcript orientation. Of the curated transcriptome data set, 622 transcripts were identified as arising from transposons by BLAST (1×10^{-5}) analysis against the *Hydra*

transposons in Repbase. Of our sequenced piRNAs, 1.7 million mapped to the transcriptome when allowing up to a three-base pair mismatch. Among these, 72% of Hywi-bound piRNAs and 58% of Hyli-bound piRNAs map to transposon transcripts, which is a significant enrichment over the abundance of transposons in the transcriptome (Fig. 3c). Furthermore, significantly more piRNAs map per transposon transcript than per non-transposon transcript (Fig. 3d). The majority of Hywi-bound piRNAs map to transposons in the antisense orientation whereas the Hyli-bound piRNAs map largely in the sense orientation (Fig. 3e and f and Supplemental Table 1); this sense/antisense bias is consistent with the ping-pong model for piRNA biogenesis and post-transcriptional repression of transposons (Brennecke *et al.*, 2007). Although the majority of transposons are lowly expressed, they have a high number of piRNAs mapping to their transcripts (Fig. 3d,f). This is also consistent with the ping pong model, which posits that transposon mRNAs are repressed by processing them into piRNAs (Brennecke *et al.*, 2007). Taken together, these data strongly suggest that one role of the *Hydra* Piwi-piRNA pathway is to regulate transposon expression.

Identification of candidate non-transposon PIWI-piRNA pathway targets

The processing of mRNAs into piRNAs is also a possible mechanism of post-transcriptional repression for non-transposon genes. We found that both Hywi- and Hyli-bound piRNAs predominantly map to the non-transposon genes of the transcriptome in the sense orientation, which suggests that piRNAs are being made from these transcripts, similar to observations in *Drosophila* and mice (Fig. 3e and Supplemental Table 1) (Robine *et al.*, 2009). A group of non-transposon transcripts with more than 10 piRNAs mapping per kilobase were selected as putative targets and subjected to gene ontology analysis (Fig. 3d, Supplemental Table 2 and Supplemental Table 3). We find significant differences in the enriched GO categories between transcripts with high numbers of Hywi piRNAs mapping to them as compared to those with high numbers of Hyli piRNAs mapping to them. This suggests selectivity in the mRNAs that are processed into piRNAs. However, we also found a correlation between the expression level of

non-transposon transcripts and the number of piRNAs mapped to them (Fig. 3g). Therefore, some piRNA production may occur from highly expressed transcripts simply due to their high abundance.

To test if the PIWI-piRNA pathway in *Hydra* has targets that are specific to each developmental lineage we isolated each lineage by FACS for small RNA sequencing. Transgenic *Hydra* were used that express GFP in the endoderm and DsRed2 in the ectoderm (Fig. 2b; Supplemental Fig 4a). The interstitial lineage was collected as the population of cells without fluorescence (Supplemental Fig 4a). We found that the most abundant small RNAs in the interstitial lineage are between 26 and 32 nucleotides in length, with a peak at 28. By contrast, in both the ectodermal and endodermal lineages the most abundant small RNAs are between 26 and 34 nucleotides, with a peak at 32 (Supplemental Fig. 6a). For all three lineages, there is a bias for uridine at the 5' end of small RNAs between 26 and 34 nucleotides long (Supplemental Fig. 6b). To test for potential lineage-specific targets of the PIWI-piRNA pathway we mapped small RNAs greater than 23 nucleotides from each lineage to the transcriptome. Transcripts that had at least 10 times more mapped piRNAs from one lineage as compared to the other two lineages were considered putative lineage-specific targets. Approximately 50% of the targets specific to the interstitial lineage are transposons, whereas only one putative transposon target was enriched in epithelial cells (Fig. 3h). Generally, more piRNAs from the interstitial lineage map to transposons in the transcriptome as compared to piRNAs from the epithelial lineages and this trend was not observed for non-transposon transcripts (Supplemental Fig. 6d,e). These data suggest that transposon regulation is largely specific to the interstitial lineage, which is further supported by the observation that the ping-pong biogenesis signature is significantly stronger in the interstitial lineage (Supplemental Fig. 6c). In addition, we identified putative non-transposon targets and subjected these to gene ontology analysis; the results strongly suggest that the pathway has specific functions in each lineage (Supplemental Table 4).

Hywi has an essential function in *Hydra* epithelial cells

To gain insight into the function of the PIWI-piRNA pathway in *Hydra* somatic cells, we sought to knockdown *hywi* expression in the epithelial lineages. We modified our previously described operon vector by placing an RNA hairpin in the upstream position and the DsRed2 gene in the downstream position to mark transgenic cells (Fig. 4a) (Dana *et al.*, 2012). Expression of the two genes is driven by an actin promoter that is not active in the interstitial stem cells, but is active in the differentiated cells of the interstitial lineage and throughout the ectodermal and endodermal lineages (Supplemental Fig. 7a-c). Therefore, the RNAi transgene is predicted to affect *hywi* expression in the epithelial cell lineages, but not the interstitial lineage. Injection of plasmid DNA into early *Hydra* embryos results in random integration and the generation of mosaic patches of stably transgenic tissue (Wittlieb *et al.*, 2006). We tested two different constructs targeting *hywi* and one control construct with a hairpin from the GFP gene. Hatchlings from these injections were scored for transgene (DsRed2) expression in the epithelial cells (Fig. 4b). Fifty-eight percent of control hatchlings showed DsRed2 expression in the epithelial cells, whereas significantly fewer hatchlings from the *hywi* RNAi injections showed DsRed2 expression in the epithelial cells (15.5% and 25.8%; Fig. 4b). By contrast, the *hywi* RNAi and control transgenes were integrated into the interstitial lineage at the same rate (Fig. 4b). Fully transgenic ectodermal or endodermal lines are established by asexual propagation and continual selection of buds with the most transgenic tissue (Wittlieb *et al.*, 2006). From the initial hatchlings expressing the GFP control transgene we established lines that are fully transgenic in the ectoderm or endoderm, thus the control transgene does not negatively affect either tissue. However, we were unable to establish lines with *hywi* knocked down in the ectoderm or endoderm. These data suggest that *hywi* is essential in the epithelial lineages.

We established three lines expressing either the *hywi* RNAi-1 or the *hywi* RNAi-2 construct in the interstitial lineage (as observed by fluorescence in the differentiated cells) (Fig. 4b; Supplemental Fig. 7d). In one of these lines the *hywi* RNAi-1 transgene is regularly transmitted through the germline which results in F1 hatchlings that are uniformly transgenic in

both the endodermal and ectodermal epithelial layers (Fig. 4c and Supplemental Fig. 7e,f). Both qRT-PCR and western blot analysis of transgenic F1 hatchlings demonstrated significant down regulation of *hywi* as compared to nontransgenic F1 siblings (Fig. 4d,e and Supplemental Fig. 7m). By contrast, the RNA and protein levels of *hyli* are not significantly affected (Fig. 4d,e). *Hywi* is not detected in the epithelial cells by immunostaining, but is still detected in interstitial stem cells as expected (Supplemental Fig. 7g-l). *Hywi* knockdown F1 hatchlings initially appear normal, and eat shortly after hatching similar to nontransgenic F1 control siblings (Fig. 4g,j). However, the *hywi* knockdown F1 hatchlings begin to lose epithelial integrity as early as 5 days, and die between 8 and 12 days after their first meal (Fig. 4f,h,k). A small number of both control and knockdown hatchlings never eat; all of these animals die of starvation rather than loss of epithelial integrity. These observations provide further evidence that *hywi* is an essential gene in the somatic epithelium of this organism.

DISCUSSION

The PIWI-piRNA pathway is best known for repressing transposon expression in the germline to maintain genomic integrity (Siomi *et al.*, 2011). In this study we report that *Hydra* PIWI proteins accumulate in the cytoplasm of all stem/progenitor cells of the adult and are essential for the animal. These data reveal crucial functions of the PIWI proteins beyond transposon silencing and strongly suggest that the primary function of the PIWI-piRNA pathway in *Hydra* stem cells is in post-transcriptional regulation. These data also imply that cytoplasmic function of the pathway is primitive and that nuclear function is derived, although sampling from more non-bilaterian taxa is required before definitive conclusions can be made. Beyond a handful of well-studied bilaterian models, very little is known about the localization of PIWI proteins (reviewed in Mani and Juliano, 2013). Interestingly, *Drosophila* Piwi protein and the zebrafish PIWI protein Zili can be nuclear or cytoplasmic depending on the developmental stage (Houwing *et al.*, 2008; Megosh *et al.*, 2006). Thus, it is possible that during *Hydra* embryonic development

either Hywi or Hyli has a nuclear function. Nonetheless, our data point to a conserved broader functional importance for this pathway in the cytoplasm of adult stem cells.

We found that the subpopulation of piRNAs that map to the transcriptome are highly enriched for transposon transcripts, which is in contrast to no enrichment for transposon/repeat sequences when we map the total population of piRNAs to the genome. Aside from *Drosophila*, genomic mapping of putative piRNAs in several other organisms also revealed very little enrichment for repeat sequences over total genomic content (reviewed in Mani and Juliano, 2013). Therefore, when considering piRNAs associated with cytoplasmic PIWI proteins, our transcriptome-mapping approach may be preferable to genomic mapping for drawing conclusions about the function of the pathway. Our results support the conclusion that the PIWI-piRNA pathway targets transposon mRNAs via a cytoplasmic pathway in *Hydra*. This function appears to be largely specific to the interstitial lineage, which is of particular interest because this lineage is capable of giving rise to the germline. Thus, the control of transposon expression is likely an ancient function of the PIWI-piRNA pathway in germ cells.

In addition to transposon expression, putative non-transposon targets were identified in the interstitial lineage including several involved in cell cycle regulation (Supplemental Table 4). This is consistent with studies in *Drosophila* GSCs and in mouse mesenchymal stem cells that demonstrate a potential role for the pathway in controlling cell division (Cox *et al.*, 1998; Cox *et al.*, 2000; Wu *et al.*, 2010). Our data also suggest that the PIWI-piRNA pathway has an essential function in the two strictly somatic epithelial lineages, which is likely due to a function in regulating non-transposon genes. The putative targets in the ectoderm are enriched for genes encoding cell adhesion proteins and extracellular matrix components (ECM) (Supplemental Table 4). In the endoderm there is an enrichment of both ECM and proteolysis genes among putative targets. The misregulation of the genes in these categories may lead to the defects observed in the *hywi* knockdown hatchlings, perhaps due to loss of epithelial integrity.

The presence of genes with shared expression in the germline and in stem cells has led to speculation that these cells have a common evolutionary origin, with germ cells arising as a lineage-restricted stem cell population (Agata *et al.*, 2006; Extavour, 2007). In addition, several lines of evidence suggest that germline genes are also more broadly expressed in metazoan stem cells. For example, *piwi*, *vasa*, and *nanos* are expressed and often required in many multipotent and pluripotent stem cells, both with and without germline potential (Juliano *et al.*, 2011; Juliano and Wessel, 2010). A handful of expression studies in ctenophores and cnidarians reveal *piwi* expression in somatic stem and/or progenitor cells, which suggests an ancient role for PIWI in stem cell regulation (Alie *et al.*, 2011; Juliano and Wessel, 2010; Seipel *et al.*, 2004). Our study provides a comprehensive analysis of PIWI proteins and piRNAs in a cnidarian. These data strongly suggest that the PIWI-piRNA pathway has ancient and conserved stem cell functions beyond the germline and sets the stage for a mechanistic understanding of the pathway in adult somatic stem cells.

MATERIALS AND METHODS

Animals and Culturing Conditions

Hydra magnipapillata strain 105 and *Hydra vulgaris* strain AEP were cultured by standard procedures (Lenhoff and Brown, 1970). See Supplemental Materials and Methods for details.

Hywi and Hyli Antibody Generation

His-tagged recombinant proteins were made used to raise antisera in rabbits (Hywi) or guinea pigs (Hyli). See SI Materials and Methods for details on protein purification, antibody purification, immunoblotting procedures, immunofluorescence procedures, and immunoelectron microscopy.

Nuclear-Cytoplasmic Fractionation

Fractionation was done using the ProteoExtract Subcellular Proteome Extraction Kit #539790. See SI Materials and Methods for details.

Fluorescence Activated Cell Sorting (FACS)

For small RNA sequencing, animals were prepared as previously described (Hemrich *et al.*, 2012). For immunoblot analysis, transgenic *Hydra* were dissociated with 0.25% Trypsin-EDTA solution. See SI Materials and Methods for details.

Immunoprecipitation and piRNA Sequencing

Trizol-LS was added directly to the Protein A bead/antibody complexes to isolate total RNA. Small RNA libraries were prepared using Illumina Small RNA Preparation Kit v1.5 following the manufacturer's protocol. Libraries were gel-purified and sequenced using the Genome Analyzer II. See SI Materials and Methods for further details about procedures, bioinformatics analysis and genomic mapping of piRNAs.

Sequencing of Lineage-Specific Small RNAs

Each lineage was collected by FACS, RNA was isolated using Trizol and used to generate small RNA libraries using the TruSeq Small RNA Sample Prep Kit according to the manufacturer's protocol. The libraries were sequenced using the HiSeq™ 2000. See SI Materials and Methods for details.

Assembly of the Hydra Transcriptome and Small RNA Mapping

The transcriptome was assembled using a previously described pipeline (Howison *et al.*, 2012). piRNA and lineage-specific small RNA mapping was done using Bowtie 0.1.0 (Langmead *et al.*, 2009). Gene ontology analysis of putative PIWI-piRNA pathway targets was done using DAVID (Dennis *et al.*, 2003). See SI Materials and Methods for details.

Generation of Transgenic Hydra

The generation of transgenic *Hydra* was performed as previously described (Dana *et al.*, 2012; Wittlieb *et al.*, 2006). See Supplemental Materials and Methods for details on plasmid construction and injection methods.

DATA AVAILABILITY

The sequence reported in this paper has been deposited in the GenBank database (NCBI BioProject no. PRJNA213706).(David *et al.*, 1991; Marcum *et al.*, 1980; Terada *et al.*, 1988)

REFERENCES

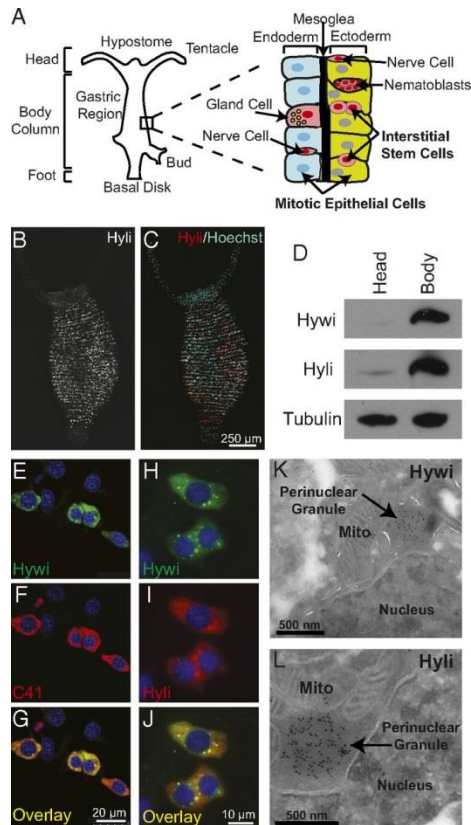
- Agata, K., Nakajima, E., Funayama, N., Shibata, N., Saito, Y., and Umesono, Y. (2006). Two different evolutionary origins of stem cell systems and their molecular basis. *Semin Cell Dev Biol* 17, 503-509.
- Alie, A., Leclere, L., Jager, M., Dayraud, C., Chang, P., Le Guyader, H., Queinnec, E., and Manuel, M. (2011). Somatic stem cells express Piwi and Vasa genes in an adult ctenophore: Ancient association of "germline genes" with stemness. *Dev Biol* 350, 183-197.
- Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* 31, 785-799.
- Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* 316, 744-747.
- Bosch, T., and David, C. (1987). Stem Cells of *Hydra magnipapillata* can differentiate into somatic cells and germ line cells. *Dev Biol* 121, 182-191.
- Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128, 1089-1103.
- Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* 12, 503-514.
- Chapman, J.A., Kirkness, E.F., Simakov, O., Hampson, S.E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, P.G., Borman, J., Busam, D., *et al.* (2010). The dynamic genome of *Hydra*. *Nature* 464, 592-596.
- Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev* 12, 3715-3727.
- Cox, D.N., Chao, A., and Lin, H. (2000). piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* 127, 503-514.
- Dana, C.E., Glauber, K.M., Chan, T.A., Bridge, D.M., and Steele, R.E. (2012). Incorporation of a horizontally transferred gene into an operon during cnidarian evolution. *PLoS One* 7, e31643.
- David, C.N. (2012). Interstitial stem cells in *Hydra*: multipotency and decision-making. *Int J Dev Biol* 56, 489-497.
- David, C.N., Fujisawa, T., and Bosch, T.C. (1991). Interstitial stem cell proliferation in *hydra*: evidence for strain-specific regulatory signals. *Dev Biol* 148, 501-507.
- David, C.N., and Murphy, S. (1977). Characterization of interstitial stem cells in *hydra* by cloning. *Dev Biol* 58, 372-383.
- Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4, P3.
- Extavour, C. (2007). Evolution of the bilaterian germ line: lineage origin and modulation of specification mechanisms. *Integrative and Comparative Biology* 47, 770-785.
- Glauber, K.M., Dana, C.E., Park, S.S., Coby, D.A., Noro, Y., Fujisawa, T., Chamberlin, A.R., and Steele, R.E. (2013). A small molecule screen identifies a novel compound that induces a homeotic transformation in *Hydra*. in press at *Development*.
- Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* 455, 1193-1197.

- Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* *315*, 1587-1590.
- Hemmrich, G., Khalturin, K., Boehm, A.M., Puchert, M., Anton-Erxleben, F., Wittlieb, J., Klostermeier, U.C., Rosenstiel, P., Oberg, H.H., Domazet-Loso, T., *et al.* (2012). Molecular signatures of the three stem cell lineages in Hydra and the emergence of stem cell function at the base of multicellularity. *Mol Biol Evol.*
- Holstein, T.W., Hobmayer, E., and David, C.N. (1991). Pattern of epithelial cell cycling in hydra. *Dev Biol* *148*, 602-611.
- Houwing, S., Berezikov, E., and Ketting, R.F. (2008). Zili is required for germ cell differentiation and meiosis in zebrafish. *EMBO J* *27*, 2702-2711.
- Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B., *et al.* (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* *129*, 69-82.
- Howison, M., Sinnott-Armstrong, N., and Dunn, C.W. (2012). BioLite, a lightweight bioinformatics framework with automated tracking of diagnostics and provenance. Proceedings of the 4th USENIX Workshop on the Theory and Practice of Provenance (TaPP '12).
- Huang, X.A., Yin, H., Sweeney, S., Raha, D., Snyder, M., and Lin, H. (2013). A Major Epigenetic Programming Mechanism Guided by piRNAs. *Dev Cell* *24*, 502-516.
- Juliano, C., Wang, J., and Lin, H. (2011). Uniting germline and stem cells: the function of Piwi proteins and the piRNA pathway in diverse organisms. *Annu Rev Genet* *45*, 447-469.
- Juliano, C., and Wessel, G. (2010). Developmental biology. Versatile germline genes. *Science* *329*, 640-641.
- Krishna, S., Nair, A., Cheedipudi, S., Poduval, D., Dhawan, J., Palakodeti, D., and Ghanekar, Y. (2013). Deep sequencing reveals unique small RNA repertoire that is regulated during head regeneration in Hydra magnipapillata. *Nucleic Acids Res* *41*, 599-616.
- Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., *et al.* (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* *131*, 839-849.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* *10*, R25.
- Lenhoff, H.M., and Brown, R.D. (1970). Mass culture of hydra: an improved method and its application to other aquatic invertebrates. *Lab Anim* *4*, 139-154.
- Mani, S.R., and Juliano, C.E. (2013). Untangling the web: The diverse functions of the PIWI/piRNA pathway. *Mol Reprod Dev.*
- Marcum, B.A., Fuijsawa, T., and Sugiyama, T. (1980). A mutant strain (sf-1) containing temperature-sensitive interstitial cells. In *Developmental and Cellular Biology of Coelenterates*, P. Tardent, and R. Tardent, eds. (Amsterdam: Elsevier), pp. 429-434.
- Megosh, H.B., Cox, D.N., Campbell, C., and Lin, H. (2006). The role of PIWI and the miRNA machinery in *Drosophila* germline determination. *Curr Biol* *16*, 1884-1894.
- Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., and Miyauchi, K. (2007). The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. *Nat Struct Mol Biol* *14*, 349-350.
- Robine, N., Lau, N.C., Balla, S., Jin, Z., Okamura, K., Kuramochi-Miyagawa, S., Blower, M.D., and Lai, E.C. (2009). A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol* *19*, 2066-2076.
- Saito, K., Sakaguchi, Y., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev* *21*, 1603-1608.
- Seipel, K., Yanze, N., and Schmid, V. (2004). The germ line and somatic stem cell gene Cniwi in the jellyfish *Podocoryne carnea*. *Int J Dev Biol* *48*, 1-7.

- Sharma, A.K., Nelson, M.C., Brandt, J.E., Wessman, M., Mahmud, N., Weller, K.P., and Hoffman, R. (2001). Human CD34(+) stem cells express the hiwi gene, a human homologue of the *Drosophila* gene piwi. *Blood* 97, 426-434.
- Sienski, G., Donertas, D., and Brennecke, J. (2012). Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression. *Cell* 151, 964-980.
- Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* 12, 246-258.
- Terada, H., Sugiyama, T., and Shigenaka, Y. (1988). Genetic analysis of developmental mechanisms in hydra. XVIII. Mechanism for elimination of the interstitial cell lineage in the mutant strain Sf-1. *Dev Biol* 126, 263-269.
- Thomson, T., and Lin, H. (2009). The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. *Annu Rev Cell Dev Biol* 25, 355-376.
- Unhavaithaya, Y., Hao, Y., Beyret, E., Yin, H., Kuramochi-Miyagawa, S., Nakano, T., and Lin, H. (2009). MILI, a PIWI-interacting RNA-binding Protein, Is Required for Germ Line Stem Cell Self-renewal and Appears to Positively Regulate Translation. *J Biol Chem* 284, 6507-6519.
- Wang, J., Saxe, J., Tanaka, T., Chuma, S., and Lin, H. (2009). Mili interacts with tudor domain-containing protein 1 in regulating spermatogenesis. *Curr Biol* 19, 640-644.
- Wittlieb, J., Khalturin, K., Lohmann, J.U., Anton-Erxleben, F., and Bosch, T.C. (2006). Transgenic Hydra allow in vivo tracking of individual stem cells during morphogenesis. *Proc Natl Acad Sci U S A* 103, 6208-6211.
- Wu, Q., Ma, Q., Shehadeh, L.A., Wilson, A., Xia, L., Yu, H., and Webster, K.A. (2010). Expression of the Argonaute protein PiwiL2 and piRNAs in adult mouse mesenchymal stem cells. *Biochem Biophys Res Commun* 396, 915-920.

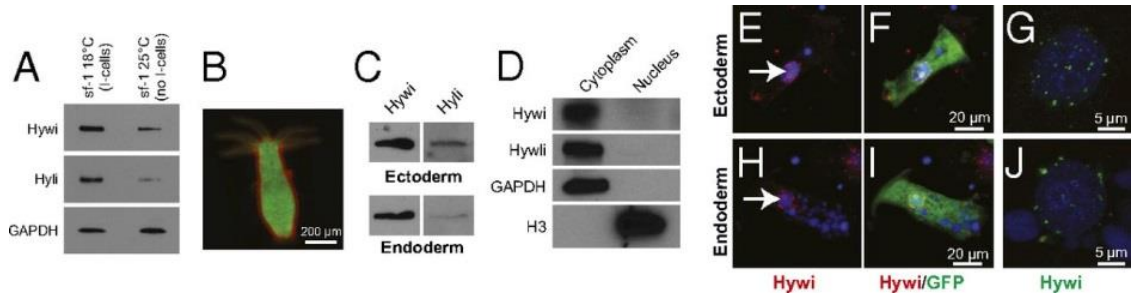
FIGURES

Figure 1: PIWI proteins are expressed in the interstitial stem cells and are enriched in perinuclear granules.



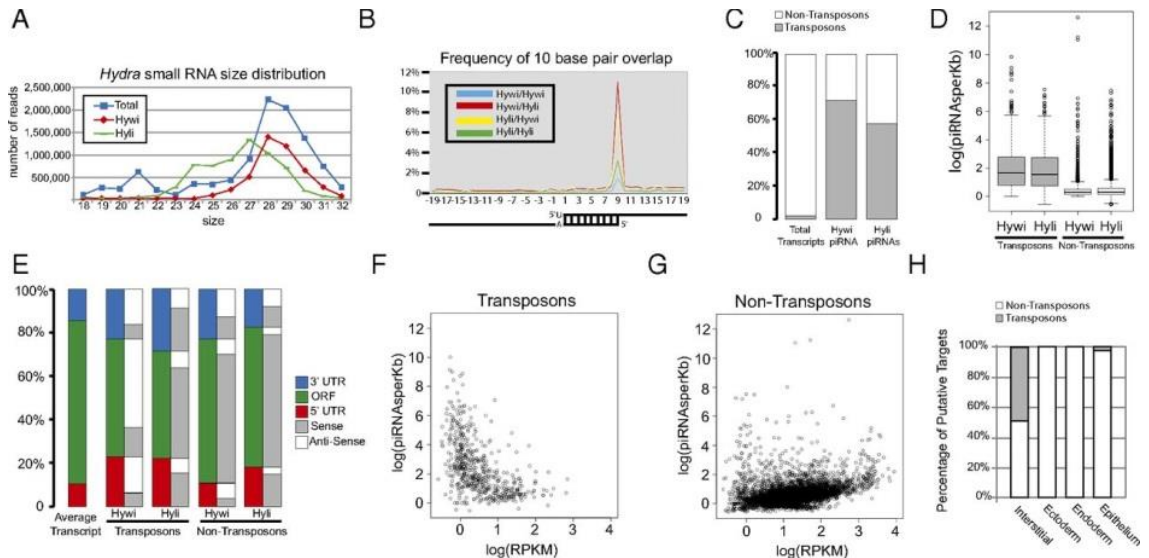
(A) A schematic of *Hydra* showing that it is composed of three cell lineages. The ectodermal (green) and endodermal (blue) epithelial cell lineages form the inside and outside of the body column. All of the epithelial cells in the body column are mitotic and maintain the lineage (Holstein, *et al.*, 1991). As they divide they are displaced towards the extremities where they become post-mitotic, differentiate, and eventually are sloughed from the tips of the tentacles and the foot. The interstitial cell lineage (pink) consists of the interstitial stem cells, which give rise to the differentiated nerve cells, gland cells, nematocytes (from precursor nematoblast nests), and germ cells (David, 2012). The expression of *Hydra* PIWI protein, Hyli and Hywi, is restricted to the body column as shown by Hyli whole-mount immunofluorescence (B,C) and Hywi and Hyli immunoblot analysis (D). (E-G) Hywi (green) and Hyli (Supplemental Fig. 2d-f) are expressed in the interstitial stem cells as shown by co-labeling with the C41 antibody, which labels interstitial stem cells (red) (David, *et al.*, 1991). Hywi (green) and Hyli (red) proteins are distributed diffusely in the cytoplasm of interstitial stem cells, and are enriched in perinuclear granules, as demonstrated by immunofluorescence (H-J) and immunoelectron microscopy (K,L). DNA is labeled with Hoechst 33342.

Figure 2: PIWI proteins are cytoplasmic and expressed in the mitotically active somatic epithelial cells



(A) Expression of Hywi and Hyli protein is detected by immunoblot in epithelialized *sf-1 Hydra*, which lose the interstitial lineage when cultured at 25°C (Marcum, *et al.*, 1980, Terada, *et al.*, 1988). (B) To test for epithelial expression of Hywi and Hyli, ectodermal cells expressing DsRed2 and endodermal cells expressing GFP were isolated by FACS (Supplemental Fig 4) and subjected to immunoblot analysis with Hywi and Hyli antibodies. (C) Hywi and Hyli were detected in both the ectodermal and endodermal epithelial cells. (D) To test the subcellular localization of Hywi and Hyli, nuclear (histone H3) and cytoplasmic (GAPDH) fractions were probed with the Hywi and Hyli antibodies; both proteins were detected selectively in the cytoplasmic fractions. To determine the subcellular localization of Hywi in epithelial cells, staining was performed on transgenic *Hydra* that express GFP in either the ectodermal (E-G) or endodermal (H-J) epithelial cells. (E-J) Hywi accumulates in perinuclear granules (arrows in E and H) in ectodermal (E-G) and endodermal (H-J) epithelial cells. (G, J) Hywi-positive granules are detected around the nucleus of epithelial cells in confocal Z-stack projections. (E-J) DNA is labeled with Hoechst 33342; vacuoles in endodermal cells (H, I) are also Hoechst-positive.

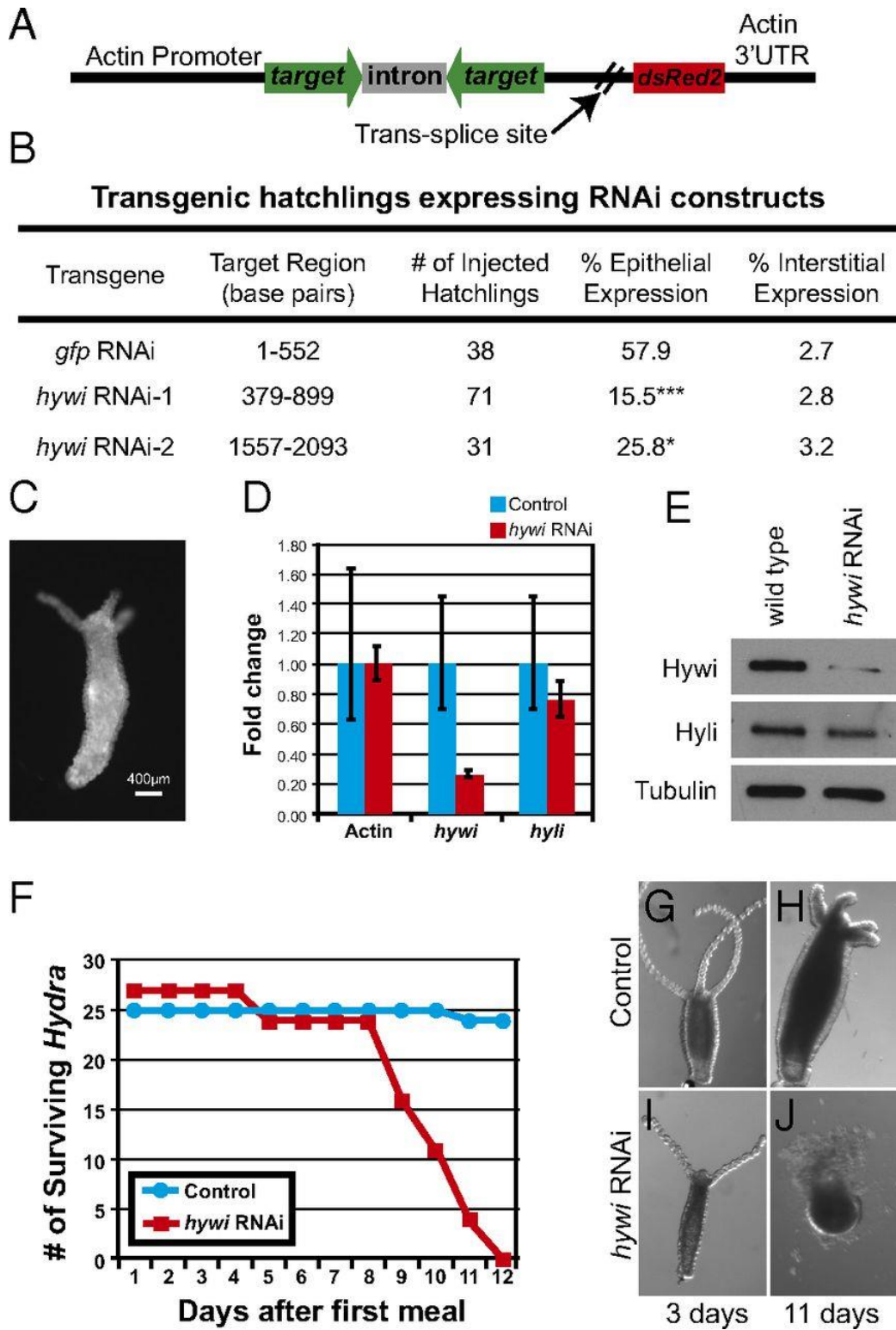
Figure 3: Sequencing and mapping of Hywi- and Hyli-bound piRNAs reveals conserved mechanisms of piRNA biogenesis and candidate post-transcriptional targets



Small RNAs isolated by size and piRNAs bound to Hywi or Hyli were sequenced. (A) Analysis of the size distribution of total small RNAs (blue squares) in *Hydra* reveals a peak at 21 nucleotides in length and a peak at 28 nucleotides in length. piRNAs bound to Hywi (red diamonds) have a peak at 28 nucleotides and piRNAs bound to Hyli (green rectangles) have a peak at 27 nucleotides. (B) Hywi and Hyli bound piRNAs have a high frequency of complementary overlap 10 bases from their 5' end. (C) A *Hydra* transcriptome was assembled and piRNAs were mapped to it. Transposon sequences represent 2.2% of the sequences in the transcriptome. By contrast, 72% of mapping Hywi-bound piRNAs and 58% of mapping Hyli-bound piRNAs map to transposons. (D) A box and whisker plot analyzing the number of piRNAs mapping per kilobase of sequence demonstrated that on average, 46 Hywi-bound piRNAs and/or 36 Hyli-bound piRNAs mapped to each kilobase of transposon transcript. By contrast, non-transposon transcripts have on average only two piRNAs mapping per kilobase, but 371 and 536 transcripts have more than 10 Hywi- or Hyli-bound piRNAs mapped per kilobase respectively (with 188 transcripts common to both populations). (E) piRNAs mapping to the UTRs are slightly overrepresented as compared to the coding region per kilobase of transcript; the architecture of the average transcript in the assembly is represented by the first bar. For transposon transcripts, the majority of Hywi-bound piRNAs map in the anti-sense orientation (white) and the majority of Hyli-bound piRNAs map in the sense orientation (grey). The majority of both Hywi- and Hyli-bound piRNAs that map to non-transposon transcripts map in the sense orientation (see Table S1 for percentages). (F,G) Number of piRNAs bound per transcript (normalized by transcript size) as a function of transcript abundance (RPKM value). (F) The majority of transposon transcripts are in low abundance and have a high number of piRNAs mapping to them. (G) By contrast, the majority of non-transposon transcripts show a correlation between transcript abundance and the number of piRNAs mapping to them, suggesting that some transcripts may be processed due to their abundance. (H) To identify lineage-specific targets of the PIWI-piRNA pathway, small RNAs were isolated from FACS separated interstitial, ectodermal, and endodermal lineages. Small RNAs greater than 23 nucleotides long were mapped to the transcriptome. Approximately 50% of the putative targets in the interstitial lineage are transposons, whereas no transposon targets were identified as specific to the ectoderm or endoderm. One putative transposon target

was identified in the epithelium (combination of ectoderm and endoderm) that is enriched over the interstitial lineage.

Figure 4: *Hywi* has an essential function in *Hydra* epithelial cells



(A) Schematic of the RNAi construct used to knockdown *hywi* in the epithelial cells (ectoderm and endoderm). Expression of the transgene is driven by an actin promoter, which is active in all

cells except the interstitial stem cells (see Supplemental Fig. 7a-c). The RNA hairpin (inverted repeats separated by an actin intron spacer) and the DsRed2 transcript are arranged in an operon configuration which are spliced apart after transcription. (B) *Hydra* embryos were injected with control (*gfp* RNAi) and *hywi* RNAi knockdown plasmids and the percentage of injected hatchlings that have transgene expression were quantified. Significantly fewer hatchlings expressed the *hywi* knockdown transgene in the epithelium as compared to the control; the p-values for epithelial expression of the *hywi* RNAi constructs are 0.0001 (*hywi* RNAi-1) and 0.04 (*hywi* RNAi-2). (C) The *hywi* RNAi-1 transgene was stably incorporated into the interstitial lineage and underwent germline transmission (see Supplemental Fig. 7d-f). This resulted in F1 *Hydra* expressing the transgene in both the ectodermal and endodermal epithelial layers, but not the interstitial stem cells (see Supplemental Fig 7g-m). (D) By qRT-PCR, the *hywi* mRNA levels (normalized to GAPDH) are reduced by ~80% in transgenic F1 hatchlings (samples taken 7 days after hatchling) and (E) the protein levels are reduced. (D,E) By contrast, *hyli* RNA and protein levels are similar between control and *hywi* knockdown F1 animals. (F) *hywi* knockdown F1 animals all die between 8 and 12 days after eating their first meal, whereas control hatchlings are normal. (G,H) Control F1 animals (non-transgenic siblings) look normal after 11 days. (J,K) *hywi* knockdown F1 animals are initially normal, but lose epithelial integrity between 8 and 12 days after eating.

SUPPLEMENTAL INFORMATION

***Hydra* strains and culturing conditions**

Hydra magnipapillata strain 105 and *Hydra vulgaris* strain AEP were cultured at 18°C by standard procedures (Lenhoff and Brown, 1970). All experiments described were performed using the 105 strain unless otherwise noted. Transgenic AEP *Hydra* expressing GFP in either the ectoderm or endoderm were used for immunofluorescent labeling and transgenic animals expressing DsRed2 in the ectoderm and GFP in the endoderm were used for FACS. These transgenic animals were made as previously described (Dana *et al.*, 2012; Wittlieb *et al.*, 2006). *sf-1* strain *Hydra* were cultured at 25°C for 4 days to remove the interstitial cell lineage and then processed for immunoblot and immunofluorescence analysis as described below (Marcum *et al.*, 1980).

Hywi and Hyli identification and antibody generation

Hywi and *Hyli* were identified by BLAST analysis of the *Hydra magnipapillata* 105 genome (<http://hydrazome.metazome.net/cgi-bin/gbrowse/hydra/>) (Altschul *et al.*, 1990). Full-length cDNA *hywi* and *hyli* sequences, including UTRs, were obtained using the First Choice RLM-RACE Kit (Life Technologies; Carlsbad, CA) and deposited into GenBank (*Hywi*, KF411461; *Hyli*, KF411462). Using PAUP (Phylogenetic Analysis Using Parsimony), an unrooted neighbor-joining phylogram was made from full-length *piwi* coding sequences; bootstrap replicate values are from 1000 iterations (Swofford, 2002). The *Hywi* N-terminus (amino acids 1-227), the *Hywi* mid-domain (amino acids 444-583), the *Hyli* N-terminus (amino acids 1-270), and the *Hyli* mid-domain (amino acids 479-622) were cloned into the Gateway expression vector pDEST17, which has an N-terminal 6xHis tag (Life Technologies; Carlsbad, CA). Recombinant protein was expressed in BL21-AI bacterial cells (Life Technologies; Carlsbad, CA), purified on Ni-NTA resin (Qiagen; Valencia, CA), and further purified by SDS-PAGE separation and electro-elution. Purified proteins were used to raise antisera in rabbits (*Hywi*) or guinea pigs (*Hyli*) (Cocalico Biologicals Inc.; Reamstown, PA). The *Hyli* N-terminus

antibody was affinity purified for immune-EM and immunoprecipitation. Hyli N-terminus recombinant protein was immobilized on an Affi-gel 10 column per the manufacturer's instructions (Bio-Rad; Hercules, CA). Heat-inactivated antiserum was passed over the antigen-immobilized column and bound antibodies were eluted with 1 ml 100 mM glycine (pH 2.5) into 50 ul 1 M Tris (pH 9.5) and used directly for immuno-EM and IP.

Immunoblot and immunofluorescence analysis

For immunoblot analysis, protein extracts were made by removing the culture medium and adding 1X SDS-PAGE loading buffer with 5 mM DTT to Hydra. Approximately 10 µl of buffer was used per Hydra polyp. To obtain equal loading, the Pierce™ BCA Protein Assay Kit (Thermo Scientific; Rockford, IL) was used for quantitation of total protein. Samples were vortexed, heated at 100°C for 10 minutes, and spun at 15K RPM for 1 minute. Samples were loaded onto 4-15% Mini-PROTEAN TGX Precast Gels (Bio-Rad; Hercules, CA). After transfer to nitrocellulose (Bio-Rad; Hercules, CA) proteins were exposed to primary antibodies overnight at 4°C in TBST + 3% dry milk. Primary antibody dilutions were as follows: Hywi N-terminal serum 1:20,000; Hywi Mid-domain serum 1:2000; Hyli N-terminal serum 1:2000; Hyli Mid-domain serum 1:2000; α -Alpha-Tubulin 1:10,000 (12G10 from the Hybridoma bank), GAPDH 1:1,000 (Sigma 9545; St. Louis, MO); Histone H3 1:1,000 (Abcam 1791; Cambridge, MA). HRP-conjugated secondary antibodies (Jackson ImmunoResearch; West Grove, PA) were diluted 1:10,000 and incubated in blocking buffer for 1 hour at room temperature and visualized by standard ECL detection (ThermoScientific; Rockford, IL).

Whole mount immunofluorescence was performed as previously described (Munder *et al.*, 2010). Briefly, *Hydra* were relaxed in 2% urethane in *Hydra* medium, fixed in 4% PFA in *Hydra* medium, washed with PBS, and permeabilized with 0.5% Triton X-100 in PBS. Samples were incubated with blocking solution (1% BSA; 10% normal goat serum; 0.1% Triton X-100 in PBS) for one hour. Primary antibodies were diluted in blocking solution and incubations were done overnight at 4°C. Antibody dilutions were as follows: Hywi N-terminal serum 1:1,000; Hyli

N-terminal serum 1:1,000; C41 monoclonal 1:2 (David *et al.*, 1991); GFP 1:500 (Roche Cat #11814460001; Indianapolis, IN); dsRed2 1:50 (Santa Cruz # sc-81595; Santa Cruz, CA). Alexa Fluor-conjugated secondary antibodies were diluted 1:500 in blocking buffer and incubations were done for 1 hour at room temperature (Invitrogen; Carlsbad, CA). For labeling of cells from dissociated *Hydra* polyps, macerations were done as previously described (David, 1973). Slides were dried for at least 3 hours and labeling was then performed using the same steps described for whole-mount labeling. Whole-mount *Hydra* preparations were imaged on a Leica TCS SP5 confocal microscope (Leica Microsystems, Bannockburn, IL) and single cells were imaged either on the Leica TCS SP5 or on a Zeiss AxioImager Z1 microscope (Carl Zeiss, Inc.; Thornwood, NY). DNA was labeled with 1:000 Hoechst 33342 diluted in PBS (Life Technologies; Carlsbad CA).

Immuno-electron microscopy

Samples were fixed in 4% paraformaldehyde/0.1% glutaraldehyde in PBS for 15 minutes, followed by 4% PFA in PBS for 1 hour. Samples were cryoprotected in 2.3 M sucrose overnight at 4°C. The samples were rapidly frozen onto aluminum pins in liquid nitrogen. The frozen block was trimmed on a Leica Cryo-EMUC6UltraCut (Leica Microsystems, Bannockburn, IL) and 60 nm sections were collected as previously described (Tokuyasu, 1973). The frozen sections were thawed and placed on a nickel formvar /carbon-coated grid floated in a dish of PBS ready for immunolabeling. For immunolabeling, samples on grids were placed section side down on drops of 0.1M ammonium chloride to quench unreacted aldehyde groups, then blocked for nonspecific binding with 1% fish skin gelatin in PBS. The grids were incubated with primary antibodies for 30 minutes, Hywi serum 1:150 or purified Hyli antibody 1:50. For Hyli, a rabbit anti-guinea pig bridging serum (Jackson ImmunoResearch; West Grove, PA) was used. Rinsed grids were placed on Protein A-gold 10 nm (UtrechtUMC) for 30 minutes. All grids were rinsed in PBS, fixed using 1% glutaraldehyde, then rinsed and transferred to a UA/methylcellulose drop for 10 minutes. Samples were viewed using a FEI Tencai Biotwin Transmission Electron Microscope (FEI;

Hillsboro, Oregon) at 80Kv. Images were captured using Morada CCD and iTEM (Olympus) software.

Nuclear-Cytoplasmic Fractionation

Approximately 100 AEP Hydra polyps were dissociated into single cells as previously described and filtered through a 100 μ m cell strainer (Gierer *et al.*, 1972). 3×10^5 cells were collected and fractionation was performed following the cell culture protocol of the ProteoExtract Subcellular Proteome Extraction Kit #539790 (EMD Millipore; Darmstadt, Germany). Nuclear and cytoplasmic fractions were analyzed by immunoblot analysis as described above.

Fluorescence Activated Cell Sorting (FACS)

Transgenic AEP Hydra polyps expressing DsRed2 in the ectoderm and GFP in the endoderm were used for all FACS experiments (Glauber *et al.*, 2013). For small RNA sequencing the animals were dissociated using Pronase E (Sigma; St. Louis, MO) as previously described (Hemrich *et al.*, 2012). However, this method did not work for immunoblotting because proteins were severely degraded. For immunoblotting Hydra were dissociated with 0.25% Trypsin-EDTA solution (Life Technologies; Carlsbad, CA). Approximately 120 Hydra were divided evenly into 3 wells of a 24-well plate, the Hydra medium was removed and 1 ml of trypsin solution was added. Hydra were incubated twice at 37°C for 5 minutes and pipetted up and down after each incubation to dissociate cells. Cells were moved to a 15 ml conical tube and volume was increased to 5 ml with dissociation medium (Gierer *et al.*, 1972). Trypsin was neutralized with fetal bovine serum. For both Pronase E and trypsin dissociation procedures, cells were filtered through a 100 μ m filter and washed twice with dissociation medium. Cells were collected after each wash by spinning at 200xg. Cells were sorted on a FACS Aria Cell Sorter with 100 μ m nozzle (see Supplemental Fig. 4) (BD Biosciences; San Jose, CA).

Immunoprecipitation and piRNA sequencing

For immunoprecipitations (IPs), approximately 100 *Hydra magnipapillata* strain105 polyps were homogenized in 1 ml MCB buffer [50 mM HEPES, pH7.5; 150 mM KOAc; 2 mM

Mg(OAc)₂; 10% glycerol; 0.1% TritonX-100; 0.1% NP-40; 1 mM DTT] and complete protease inhibitor cocktail (Roche; Indianapolis, IN). The protein concentration of the resulting protein extract was ~1 mg/ml. The total protein extract was pre-cleared by incubating with 50 mg of protein A sepharose CL-4B (GE Healthcare; Piscataway, NJ) for 1 hour at 4°C on a rotator. Protein extracts were incubated with antibody overnight at 4°C on a rotator. For Hywi IP, 5 µl of N-terminal antibody serum or 5 µl of pre-bleed serum was added to ~250 mg of protein extract. For Hyli IP, 4 µg of affinity-purified N-terminal antibody, 5 µl of N-terminal serum, or 5 µl of pre-bleed serum was added to ~250 mg of protein extract. Protein A beads (60 µl of a 2X slurry) were added to each IP and incubated for 1 hour at 4°C on a rotator. Beads were washed 5 times with MCB buffer. For immunoblotting, 40 µl of SDS Sample buffer plus 5 mM DTT was added to beads after removal of the last wash. Samples were vortexed and incubated at 100°C for 5 minutes. Beads were removed by centrifugation and the resulting supernatant was used for immunoblotting.

For isolation of Hywi- and Hyli-bound piRNAs, RNase OUT was added to the lysate (Life Technologies; Carlsbad, CA). After IP, 300 µl of Trizol-LS was added directly to the beads (Life Technologies; Carlsbad, CA) and RNA was isolated according to manufacturer's protocol. RNA pellet was re-suspended in 10 µL of nuclease-free water: 4 µL was used for 5'-end labeling with [γ -³²P] ATP by polynucleotide kinase and 6 µL was used for piRNA sequencing. For total small RNA sequencing 10 µg of total RNA was used as starting material. Small RNA libraries were prepared using Illumina small RNA Preparation Kit v1.5 (Illumina Inc.; San Diego, CA) following the manufacturer's protocol. In brief, RNA was electrophoresed in a 15% TBE urea gel and small RNAs were eluted from the gel. Adaptors were ligated to the 3' and 5' ends followed by reverse transcription and low cycle PCR. Libraries were gel-purified and sequenced using the Genome Analyzer II (Illumina Inc.; San Diego, CA).

β-elimination and small RNA northern blot

β -elimination reactions and small RNA northern blots were performed as previously described (Vagin *et al.*, 2006; Watanabe *et al.*, 2007). 20 μ g of total RNA from Hydra AEP in 13.5 μ l of water was combined with 4 μ l of 5x borate buffer (148 mM borax, 148 mM boric acid, pH 8.6) (Polysciences; Warrington, PA) and 2.5 μ l of freshly prepared 200 mM NaIO₄. The reaction was incubated for 15 minutes in the dark at room temperature and then 2 μ l of 100% glycerol was added to quench unreacted NaIO₄. Reactions were then incubated for an additional 15 minutes at room temperature. The reactions were dried in a SpeedVac evaporator for 75 minutes at room temperature. Pellets were resuspended in 50 μ l 1x borate buffer with 50mM NaOH (pH 9.5) and incubated at 45°C for 90 minutes. 2 μ l of glycogen was added and RNA was EtOH precipitated. Pellets were collected by centrifugation, washed with 70% EtOH and 95% EtOH respectively, and air-dried pellets were resuspended in 15 μ l of water. β -elimination reactions and untreated RNA samples were electrophoresed for 5 hours at 200V in 15% polyacrylamide midi-gels with 8M Urea and 0.5X TBE as buffer. RNA was transferred from the gels onto Hybond N+ membrane (GE Healthcare Life Sciences; Pittsburgh, PA) in 0.5X TBE at 100 mA for 2.5 hours. The membranes were then washed two times for 10 minutes with 2X SSC and then UV cross-linked using the UV Stratalinker 1800 (Stratagene; Santa Clara, CA) on the “Auto crosslink” setting. The membranes were incubated for 1 hour at 42°C in hybridization buffer (0.2M NaHPO₄ pH 7.2, 1mM EDTA, 1% BSA, 7%SDS). Probes for hybridization were an oligo complementary to miR2030 (probe sequence: CAAATTTATTTTTGCGCTCTCA) (Krishna *et al.*, 2013) and an oligo complementary to an abundant transposon-derived piRNA (probe sequence: AATCCAAACGCCAGGAATTCGATCACC). 10 pmol of each oligo was 5'-end labeled with [γ -³²P] ATP by polynucleotide kinase for 1 hour at 37°C. Labeled oligos were purified using oligo quick spin columns (Roche; Indianapolis, IN). The entire labeled oligo sample was added to the hybridization buffer for incubation overnight at 42°C with rotation. Membranes were washed four times for 10 minutes with 2xSSC/0.1%SDS at 50°C with rotation. Finally, the membranes were wrapped in Saran Wrap, exposed to phosphor plates for 6 hours, and

imaged on the Typhoon Trio Variable Mode Imager (GE Healthcare Life Sciences; Pittsburgh, PA).

Sequencing of lineage-specific small RNAs

After FACS, RNA was isolated from each lineage by Trizol extraction (Life Technologies; Carlsbad, CA). Approximately 2 µg of RNA was collected from each of the epithelial lineages, ~7 µg of RNA was collected from the interstitial lineage. This RNA was used to generate small RNA libraries using the TruSeq Small RNA Sample Prep Kit (Illumina Inc.; San Diego, CA) according to the manufacturer's protocol. Briefly, the libraries were generated by ligation of specific 5' and 3' adapters to the RNA and ligated products were reverse transcribed and amplified by PCR. PCR products were separated by polyacrylamide gel electrophoresis (PAGE), and products corresponding to adaptor ligated 18 – 35 nt long RNAs (~150 bp) were used for pooled gel purification. The libraries were sequenced using the HiSeq™ 2000 (Illumina Inc.; San Diego, CA). The quality of the RNA and the corresponding cDNA was analyzed by the Agilent 2100 Bioanalyzer (Agilent Technologies; Santa Clara, CA). For the interstitial lineage there were 15,302,191 trimmed reads less than 23 nucleotides and 23,713,174 trimmed reads 23 nucleotides or greater. For the ectodermal lineage there were 14,873,803 trimmed reads less than 23 nucleotides and 12,715,400 trimmed reads 23 nucleotides or greater. For the endodermal lineage there were 26,964,740 trimmed reads less than 23 nucleotides and 21,073,941 trimmed reads 23 nucleotides or greater.

Bioinformatic analysis and genomic mapping of small RNAs

For analysis of PIWI-bound piRNAs, three small RNA libraries were sequenced: 1. Total small RNAs (total), 2. Hywi-bound piRNAs (Hywi), and 3. Hyli-bound piRNAs (Hyli). Linker sequences were successfully trimmed from more than 90% of the sequences. 18-32nt RNAs were selected and mapped to the *Hydra* genome for further analysis. From the total library, ~10 million sequences were mapped to the genome, from the Hywi library ~4.7 million were mapped, and from the Hwyl library ~6.3 million were mapped. Approximately 50% of the sequenced small

RNAs were mapped to the genome without ambiguity. First, the small RNAs that mapped to tRNAs were annotated. Because *Hydra* tRNA annotation is not yet complete, up to 2 mismatches were allowed including insertions/deletions. For miRNA annotation, we extended 30 nucleotides on both ends of the small RNA sequences obtained by the 454 sequencing carried out as part of the *Hydra* genome project (Chapman *et al.*, 2010). For transposon/repeat annotation, we referred to the annotation information obtained via RepeatMasker as part of the *Hydra* genome project (Chapman *et al.*, 2010). For the gene annotation, we adopted the Berkeley group's annotation (Chapman *et al.*, 2010). To test whether the ping-pong mechanism of piRNA biogenesis functions in *Hydra*, we searched for the ping-pong signature among these libraries by determining whether the piRNAs have sequence partners with a 10 nucleotide off-set. The partners were defined as piRNAs whose 5' 10 nucleotides are reverse compliments, as previously described (Brennecke *et al.*, 2007). The same methods were used to test for the presence of the ping-pong signature in the lineage-specific small RNA sequencing data (26 to 34 nucleotides): 1. Interstitial, 2. Ectoderm, and 3. Endoderm.

Assembly of the *Hydra* transcriptome and mapping of piRNAs and lineage-specific small RNAs

Total RNA was isolated from 2-day starved *Hydra magnipapillata* strain 105. A cDNA library was constructed using the Illumina TruSeq RNA Sample Preparation Kit using a slightly modified procedure. The total RNA was only sheared for 1.5 minutes and prior to the final PCR enrichment, the library was run on a LabChip XT DNA 750 (PerkinElmer; Waltham, Massachusetts) to select for 500bp fragments (+/-5%). The final library was sequenced on the Illumina HiSeq 2000 (Illumina Inc.; San Diego, CA) yielding 43.6 million 100 base pair paired-end reads. The transcriptome was assembled using a previously described pipeline (Howison *et al.*, 2012). Approximately 27,000 sequences were initially assembled and then curated to 9,986 sequences with BLAST hits to the Swiss-Prot database (E value of 1×10^{-5} or less), thus allowing for identification of open reading frames and transcript orientation. To identify transposon

sequences, all known *Hydra* transposons were first downloaded from RepBase (current version August, 5th 2013). One sequence was removed (hAT-64_HM|hAT|Hydra_magnipapillata) due to ambiguous characters, yielding 565 transposons. We compared the *de novo* transcriptome against the *Hydra* transposon database by BLAST analysis (tblastx -evalue .00001) and any transcripts with a BLAST hit were flagged as transposons. Small RNAs were trimmed of adapter sequence using CutAdapt (version 1.0 “-a ATCTCGTATGCCGTCTTCTGCTTG -m 10 --too-short-output --untrimmed-output”, though subsequent analysis ignored small RNAs shorter than 23nt) and were mapped to the transcriptome allowing up to a 3 base-pair mismatch using Bowtie 1 (version 0.12.9) with the following settings: “-a --best --strata -v 3”. 1.4 million small RNAs at least 23 base pairs long from the total small RNA library were mapped to the transcriptome, 63% of these mapped to transposon sequences. Hywi-bound piRNAs and Hyli-bound piRNAs at least 23 base pairs long were also mapped to the transcriptome: 725,000 Hywi-bound piRNAs mapped and 963,000 Hyli-bound piRNAs mapped. For the lineage-specific small RNA sequences, adapter sequences were trimmed, using CutAdapt with the following settings: “-a TGGAAATTCTCGGGTGCCAAGGC -m 15”. The trimmed reads were mapped against the transcriptome with Bowtie 1 (version 0.12.9) with the same settings as above and yielded: 1,247,033 interstitial, 488,903 ectodermal, and 848,964 endodermal small RNAs 23 nucleotides or greater. Transcripts with at least 10 times more mapped piRNAs from a specific lineage as compared to other lineages (normalized to the size of the libraries) were considered putative lineage-specific targets for that particular lineage. Gene ontology analysis of putative PIWI-piRNA pathway targets was done using DAVID (Dennis *et al.*, 2003), employing the Swiss-Prot identification from the closest BLAST hit. All small RNA short reads, RNA-seq short reads and the assembled transcriptome were deposited under NCBI BioProject PRJNA213706.

Real-time quantitative PCR to test *hywi* knockdown levels

RNA was isolated from F1 *Hydra* hatchlings expressing the *hywi* RNAi-1 transgene and from wild type F1 siblings by Trizol extraction (Life Technologies; Carlsbad, CA). Further

purification of the RNA and on-column DNase-treatment was performed using the RNeasy Mini Kit (Qiagen; Valencia, CA) according to the manufacturer's protocol. Reverse transcription was performed using the High Capacity cDNA Reverse Transcription Kit (Life Technologies; Carlsbad, CA) according to the manufacturer's protocol. Real-time quantitative PCR was carried out using iQ SYBR Green 2x Supermix on a CFX96™ thermal cycler (Bio-Rad; Hercules, CA).

RNAi plasmid description and construction

The RNAi plasmids were designed in an operon configuration such that the upstream hairpin and the downstream DsRed2 gene are transcribed together from an actin gene promoter, with the bicistronic primary transcript then being separated into a hairpin RNA and a DsRed2 mRNA by trans-splicing (Dana *et al.*, 2012) (see Supplemental Fig. 4a). The RNA hairpin consists of a ~500 base pair fragment from the target gene cloned in an inverted orientation around an actin intron sequence (483 base pairs). Downstream of the hairpin cassette is the DsRed2 gene followed by the actin 3'UTR (500 base pairs). In between the hairpin cassette and the DsRed2 gene is the RFC140/*flp* intergenic sequence, which contains an acceptor for trans-spliced leader addition (Dana *et al.*, 2012). Thus, the RNA hairpin and DsRed2 transcript are arranged in an operon configuration and are spliced apart after transcription.

The operon plasmid pHyVec11 (2) was modified by the insertion of a GFP hairpin (nucleotides 1-552) separated by the intron from a *Hydra* actin gene (Fisher and Bode, 1989) in the upstream position of the operon. This plasmid, named pHyVec12 (*gfp* RNAi), was then used to construct the *hywi* RNAi-1 and *hywi* RNAi-2 plasmids. pHyVec12 was cut with NheI and BamHI to remove the GFP hairpin and the actin intron. *Hywi* forward and reverse sequences (379-899 or 1557-2093), surrounding the actin intron sequence, were then inserted into the plasmid using the Cold Fusion Cloning Kit (System Biosciences; Mountain View, CA).

Generation of transgenic *Hydra*.

The generation of transgenic *Hydra* was performed as previously described (Dana *et al.*, 2012; Wittlieb *et al.*, 2006). *Hywi* RNAi and *gfp* RNAi plasmids were prepared by Maxiprep

(Qiagen, Valencia, CA) and eluted in RNase-free water. Plasmid DNA was injected at a final concentration of 1 mg/mL using an IM-9B Narishige microinjector (Narishige; East Meadow, NY) under a Zeiss dissecting scope (Carl Zeiss, Inc.; Thornwood, NY).

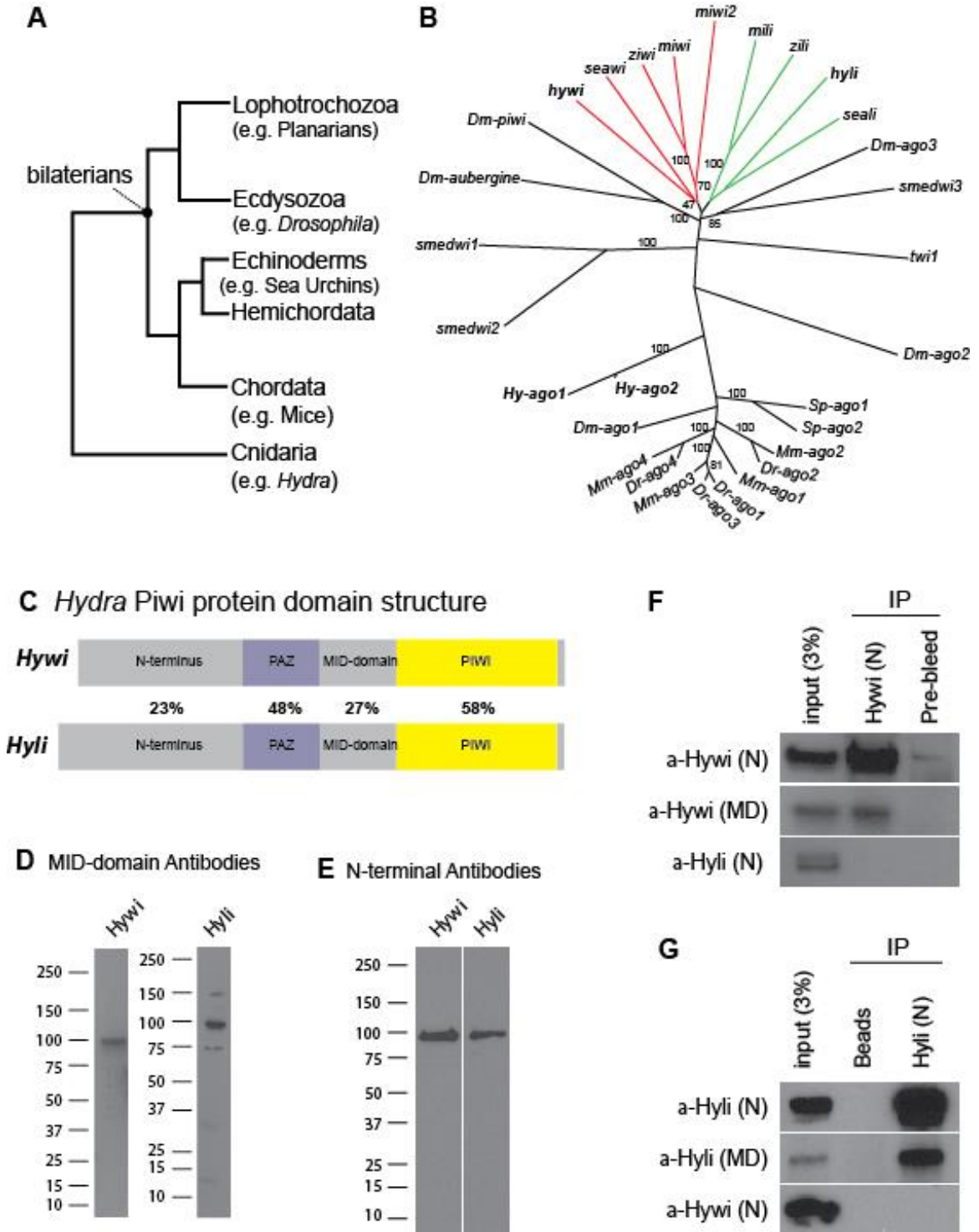
SUPPLEMENTAL REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-410.
- Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128, 1089-1103.
- Chapman, J.A., Kirkness, E.F., Simakov, O., Hampson, S.E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, P.G., Borman, J., Busam, D., *et al.* (2010). The dynamic genome of *Hydra*. *Nature* 464, 592-596.
- Dana, C.E., Glauber, K.M., Chan, T.A., Bridge, D.M., and Steele, R.E. (2012). Incorporation of a horizontally transferred gene into an operon during cnidarian evolution. *PLoS One* 7, e31643.
- David, C.N. (1973). A Quantitative Method for Maceration of *Hydra* Tissue. *Wilhelm Roux' Archiv* 171, 259-268.
- David, C.N., Fujisawa, T., and Bosch, T.C. (1991). Interstitial stem cell proliferation in *hydra*: evidence for strain-specific regulatory signals. *Dev Biol* 148, 501-507.
- David, C.N., and Gierer, A. (1974). Cell cycle kinetics and development of *Hydra attenuata*. III. Nerve and nematocyte differentiation. *J Cell Sci* 16, 359-375.
- Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4, P3.
- Fisher, D.A., and Bode, H.R. (1989). Nucleotide sequence of an actin-encoding gene from *Hydra attenuata*: structural characteristics and evolutionary implications. *Gene* 84, 55-64.
- Gierer, A., Berking, S., Bode, H., David, C.N., Flick, K., Hansmann, G., Schaller, H., and Trenkner, E. (1972). Regeneration of *hydra* from reaggregated cells. *Nat New Biol* 239, 98-101.
- Glauber, K.M., Dana, C.E., Park, S.S., Coby, D.A., Noro, Y., Fujisawa, T., Chamberlin, A.R., and Steele, R.E. (2013). A small molecule screen identifies a novel compound that induces a homeotic transformation in *Hydra*. in press at *Development*.
- Hemmrich, G., Khalturin, K., Boehm, A.M., Puchert, M., Anton-Erxleben, F., Wittlieb, J., Klostermeier, U.C., Rosenstiel, P., Oberg, H.H., Domazet-Loso, T., *et al.* (2012). Molecular signatures of the three stem cell lineages in *Hydra* and the emergence of stem cell function at the base of multicellularity. *Mol Biol Evol*.
- Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol* 17, 1265-1272.
- Howison, M., Sinnott-Armstrong, N., and Dunn, C.W. (2012). BioLite, a lightweight bioinformatics framework with automated tracking of diagnostics and provenance. *Proceedings of the 4th USENIX Workshop on the Theory and Practice of Provenance (TaPP '12)*.
- Kirino, Y., and Mourelatos, Z. (2007). Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nat Struct Mol Biol* 14, 347-348.
- Krishna, S., Nair, A., Cheedipudi, S., Poduval, D., Dhawan, J., Palakodeti, D., and Ghanekar, Y. (2013). Deep sequencing reveals unique small RNA repertoire that is regulated during head regeneration in *Hydra magnipapillata*. *Nucleic Acids Res* 41, 599-616.
- Lenhoff, H.M., and Brown, R.D. (1970). Mass culture of *hydra*: an improved method and its application to other aquatic invertebrates. *Lab Anim* 4, 139-154.
- Marcum, B.A., Fujisawa, T., and Sugiyama, T. (1980). A mutant strain (sf-1) containing temperature-sensitive interstitial cells. In *Developmental and Cellular Biology of Coelenterates*, P. Tardent, and R. Tardent, eds. (Amsterdam: Elsevier), pp. 429-434.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5, 621-628.

- Munder, S., Kasbauer, T., Prexl, A., Aufschnaiter, R., Zhang, X., Towb, P., and Bottger, A. (2010). Notch signalling defines critical boundary during budding in Hydra. *Dev Biol* 344, 331-345.
- Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., and Miyauchi, K. (2007). The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. *Nat Struct Mol Biol* 14, 349-350.
- Saito, K., Sakaguchi, Y., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes Dev* 21, 1603-1608.
- Slautterback, D.B., and Fawcett, D.W. (1959). The development of the cnidoblasts of Hydra; an electron microscope study of cell differentiation. *J Biophys Biochem Cytol* 5, 441-452.
- Swofford, D.L. (2002). PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. (Sunderland, MA: Sinauer Associates).
- Tokuyasu, K.T. (1973). A technique for ultracryotomy of cell suspensions and tissues. *J Cell Biol* 57, 551-565.
- Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313, 320-324.
- Watanabe, T., Totoki, Y., Sasaki, H., Minami, N., and Imai, H. (2007). Analysis of small RNA profiles during development. *Methods Enzymol* 427, 155-169.
- Wittlieb, J., Khalturin, K., Lohmann, J.U., Anton-Erxleben, F., and Bosch, T.C. (2006). Transgenic Hydra allow in vivo tracking of individual stem cells during morphogenesis. *Proc Natl Acad Sci U S A* 103, 6208-6211.

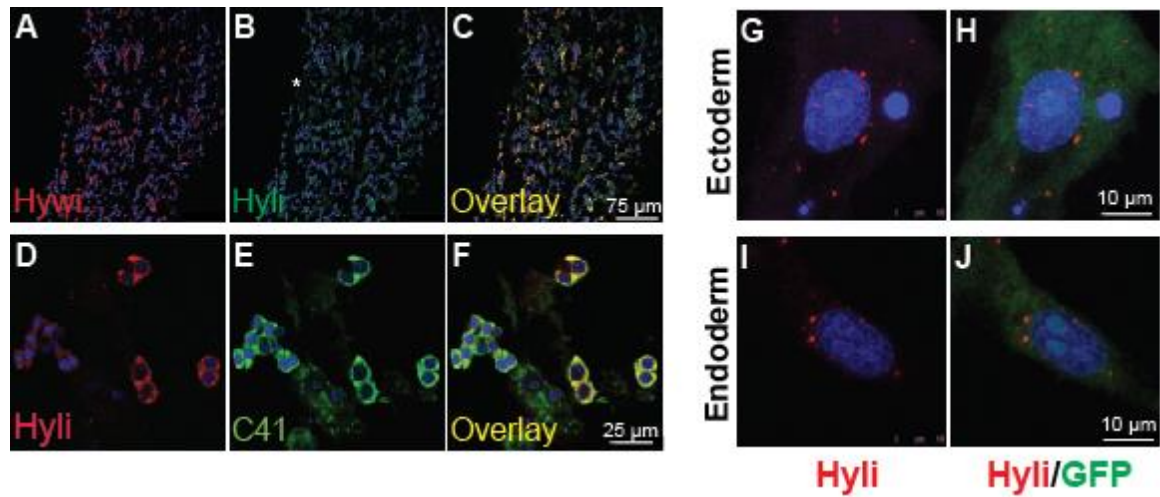
SUPPLEMENTAL FIGURES AND TABLES

Supplemental Figure 1: Generation and characterization of antibodies against Hywi and Hyli.



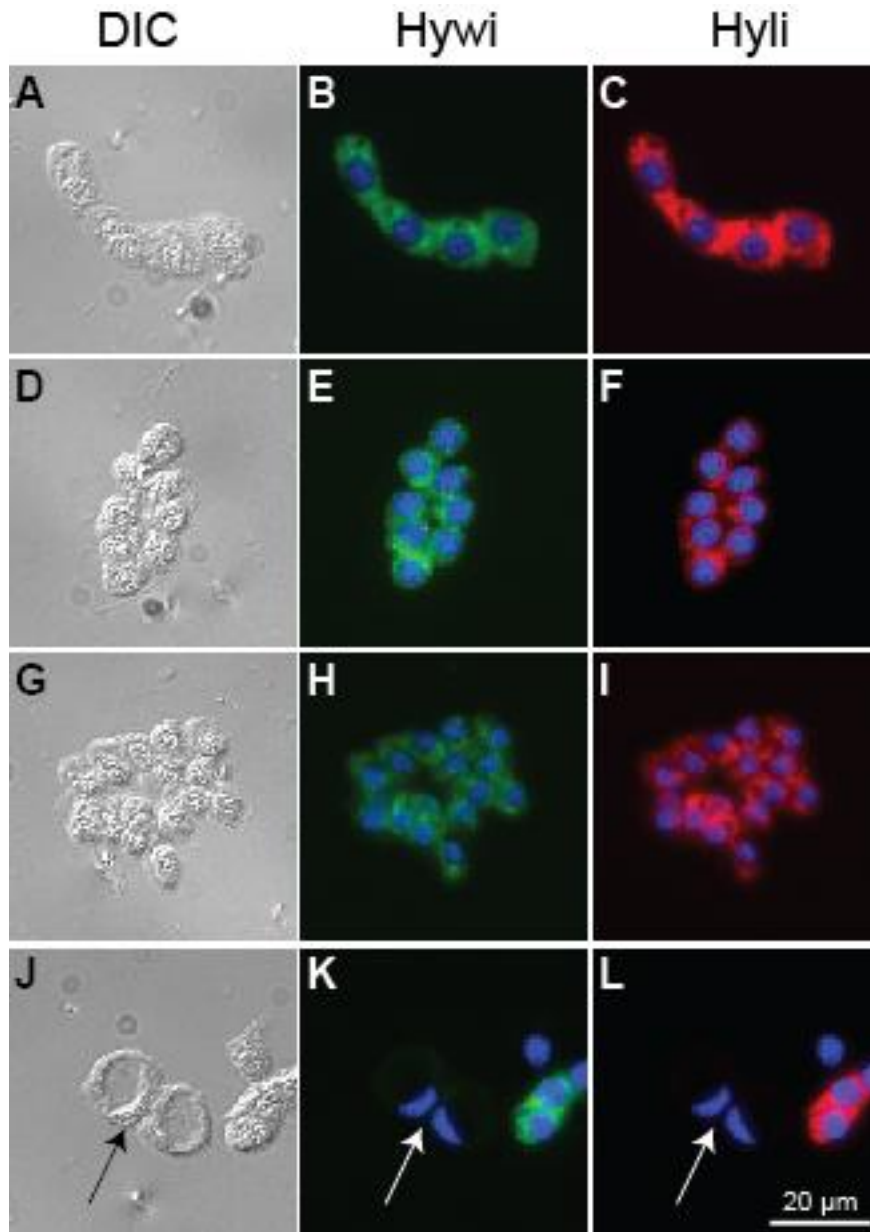
(A) *Hydra* is a member of the phylum Cnidaria, which is the sister group to the bilaterians. (B) An unrooted neighbor-joining phylogram demonstrates that *hywi* and *hyli* cluster with *piwi* family genes and that *Hy-ago1* and *Hy-ago2* cluster with *ago* family genes. *Hywi* clusters with *miwi*, *ziwi*, and *seawi* (red lines) and *hyli* clusters with *mili*, *zili*, and *seali* (green lines). Numbers indicate bootstrap replicate values from 1000 iterations. Abbreviations are as follows: Mm (*Mus musculus*), Dr (*Danio rerio*, i.e. zebrafish), Sp (*Strongylocentrotus purpuratus*, i.e. sea urchin), Sm (*Schmidtea mediterranea*, i.e. planarian), Dm (*Drosophila Melanogaster*). (C) *Hywi* and *hyli* have conserved ARGONAUTE family domain structures. Sequence comparisons between the protein domains of *Hywi* and *Hyli* show that the N-terminal and Mid-domains have lower sequence identity (23% and 27%) as compared to the PAZ and PIWI domains (48% and 58%). (D,E) Polyclonal antibodies were raised against the N-terminal and Mid-domains of *Hywi* (in rabbit) and *Hyli* (in guinea pig). (F) The 100kDa protein immunoprecipitated with the *Hywi* N-terminus (N) antibody is recognized by the antibody raised against the Mid-domain (MD) of *Hywi*, but not by the N-terminal *Hyli* antibody. (G) The 100 kDa protein immunoprecipitated with the *Hyli* N-terminus antibody is recognized by the antibody raised against the Mid-domain of *Hyli*, but not by the N-terminal *Hywi* antibody.

Supplemental Figure 2: Hyli protein is expressed in interstitial stem cells and mitotically active epithelial stem/progenitor cells.



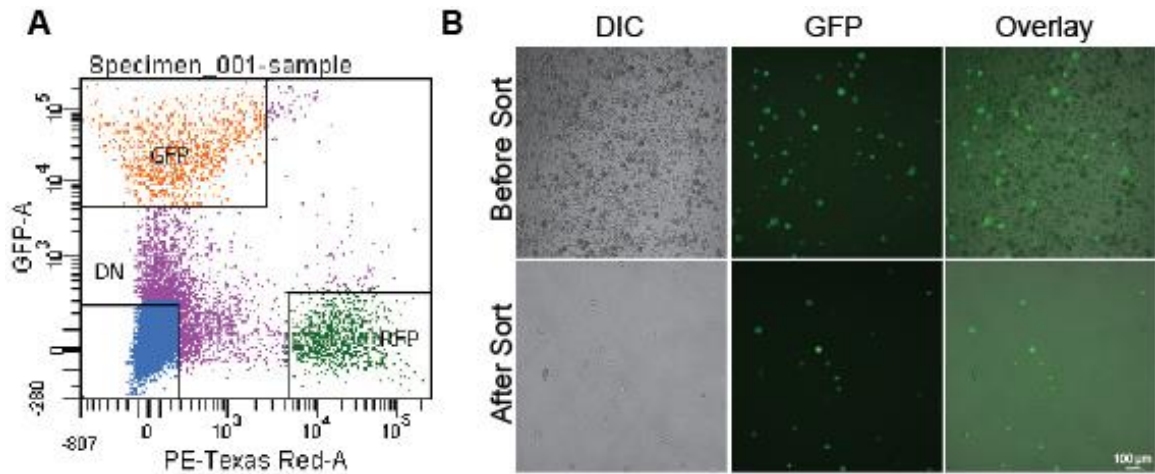
(A-C) Hywi (red) and Hyli (green) antibodies stain the same population of cells throughout the body column. Green nematocyte (e.g. asterisk) labeling in panel B is non-specific labeling from the secondary antibody. (D-F) Hyli (red) is expressed in the C41-positive (green) I-cells. (G-J) Confocal images of Hyli accumulation in perinuclear granules in ectodermal (G,H) and endodermal (I,J) epithelial cells. Staining was done on transgenic *Hydra* that express GFP in either the ectodermal (H) or endodermal (J) lineages. DNA is labeled with Hoechst 33342.

Supplemental Figure 3: *Hydra* PIWI proteins are expressed in developing nematoblasts.



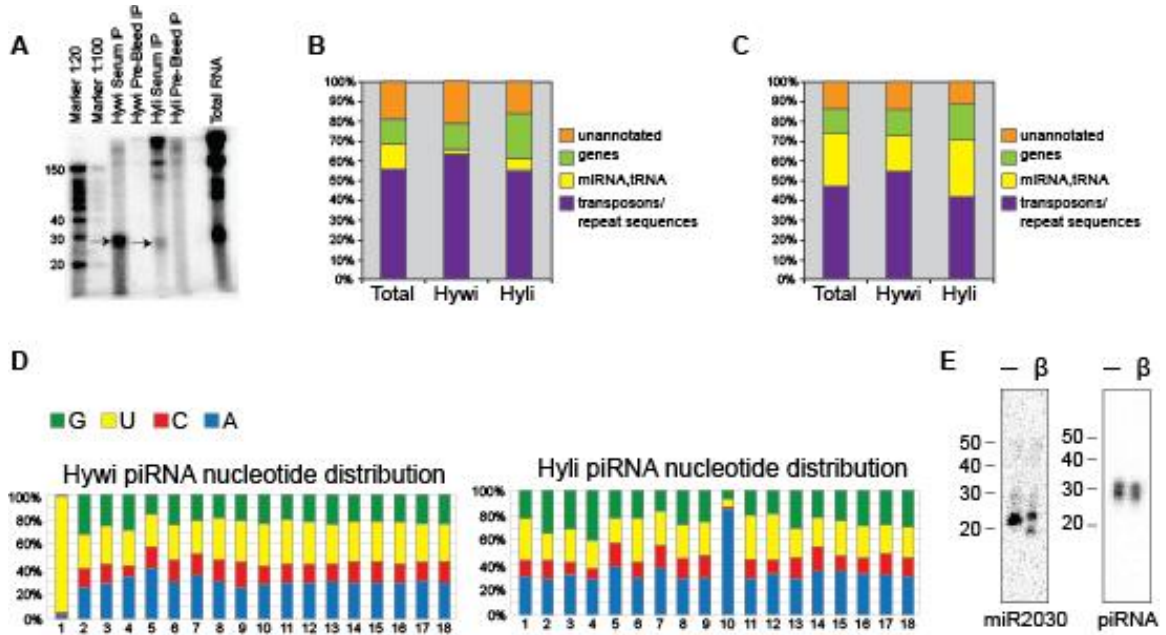
Interstitial stem cells in the process of differentiating into nematocytes, the specialized stinging cells of cnidarians, undergo four divisions with incomplete cytokinesis that give rise to 2-, 4-, 8- and 16-cell nematoblast nests that are distributed throughout the body column. In the final step of differentiation, the cells break apart from the nests and migrate to the tentacles (Slautterback and Fawcett, 1959, David and Gierer, 1974). (A-L) Staining of dissociated *Hydra* cells show that Hywi (red) and Hyli (green) proteins are expressed in 4-cell (A-C), 8-cell (D-F), and 16-cell (G-I) nematoblast nests, but not in differentiated nematocytes (J-L). DNA is labeled with Hoechst 33342.

Supplemental Figure 4: FACS isolation of ectodermal and endodermal cells.



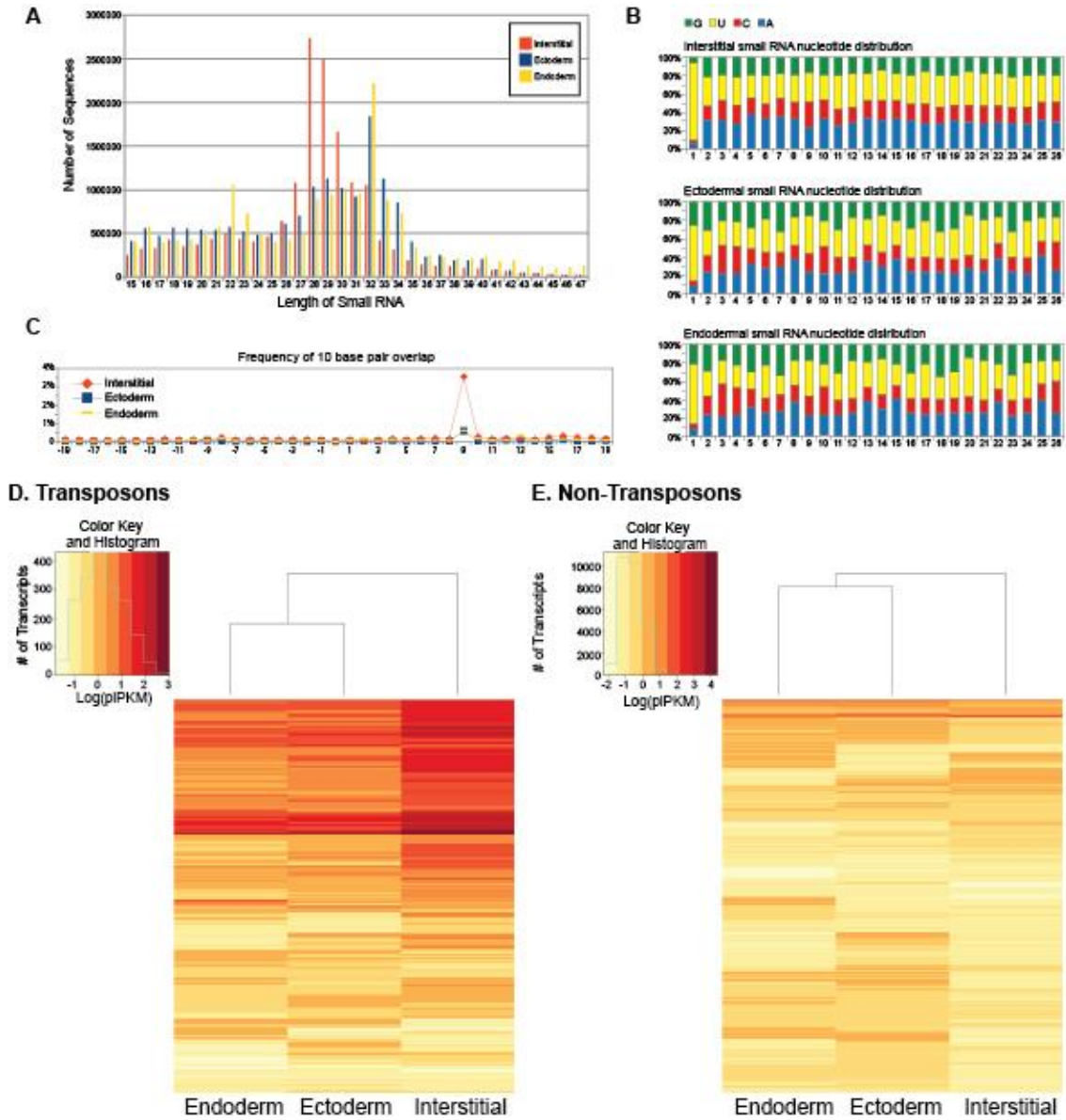
(A) Transgenic *Hydra* expressing GFP in the endoderm and DsRed2 in the ectoderm (see Fig. 2b) (Glauber, et al., 2013) were dissociated into single cells which were then sorted by FACS. Double-negative (DN) cells were also collected as the interstitial lineage cell population (B) Imaging of cells after a sort demonstrates that the GFP-positive cells were successfully separated away from the total cell population.

Supplemental Figure 5: Isolation, deep-sequencing, and mapping of Hywi and Hyli bound piRNAs to the *Hydra* genome and beta elimination assay.



(A) Both total RNA and RNA extracted from Hywi and Hyli immunoprecipitates was 5' end-labeled with [γ - 32 P] ATP. Immunoprecipitated RNA contains piRNAs just below the 30 nucleotide marker (arrows). (B,C) Mapping of *Hydra* piRNAs to the genome: "Total" is small RNAs selected by size and "Hywi" and "Hyli" are the piRNAs isolated from Hywi and Hyli immunoprecipitates respectively. (B) piRNAs that map uniquely to the *Hydra* genome (~50%) showed no enrichment for transposon/repeat sequences (the *Hydra* genome is 57% transposon/repeat sequences) (Chapman, *et al.*, 2010). (C) Analyzing all piRNAs that map to the genome, even those that map in multiple places, also shows no significant enrichment for transposon/repeat sequences. In this analysis, piRNAs that map more than once are weighted. (D) Analysis of nucleotide distribution across the length of Hywi-bound and Hyli-bound piRNAs demonstrates that Hywi-bound piRNAs have a strong preference for uridine at the 5' position and Hyli-bound piRNAs have a strong preference for adenine at the 10th position. (E) Northern blot analysis of total RNA subjected to β -elimination or control total RNA. Anti-sense probes detect *Hydra* miR2030 (Krishna, *et al.*, 2013) or an abundant *Hydra* transposon-derived Hyli-associated piRNA with the following sequence: GGTGATCGAATTCCTGGCGTTTGGATT. The piRNA, but not the miRNA, is protected from nucleotide loss due to β -elimination, thus indicating that the piRNA is 2'-O-methylated at the 3' end similar to piRNAs in *Drosophila* and mice (Horwich, *et al.*, 2007, Kirino and Mourelatos, 2007, Ohara, *et al.*, 2007, Saito, *et al.*, 2007).

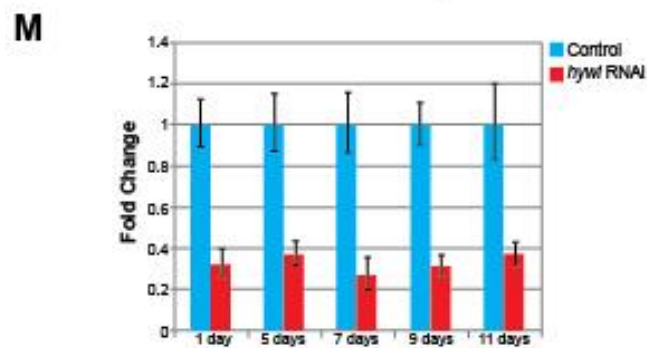
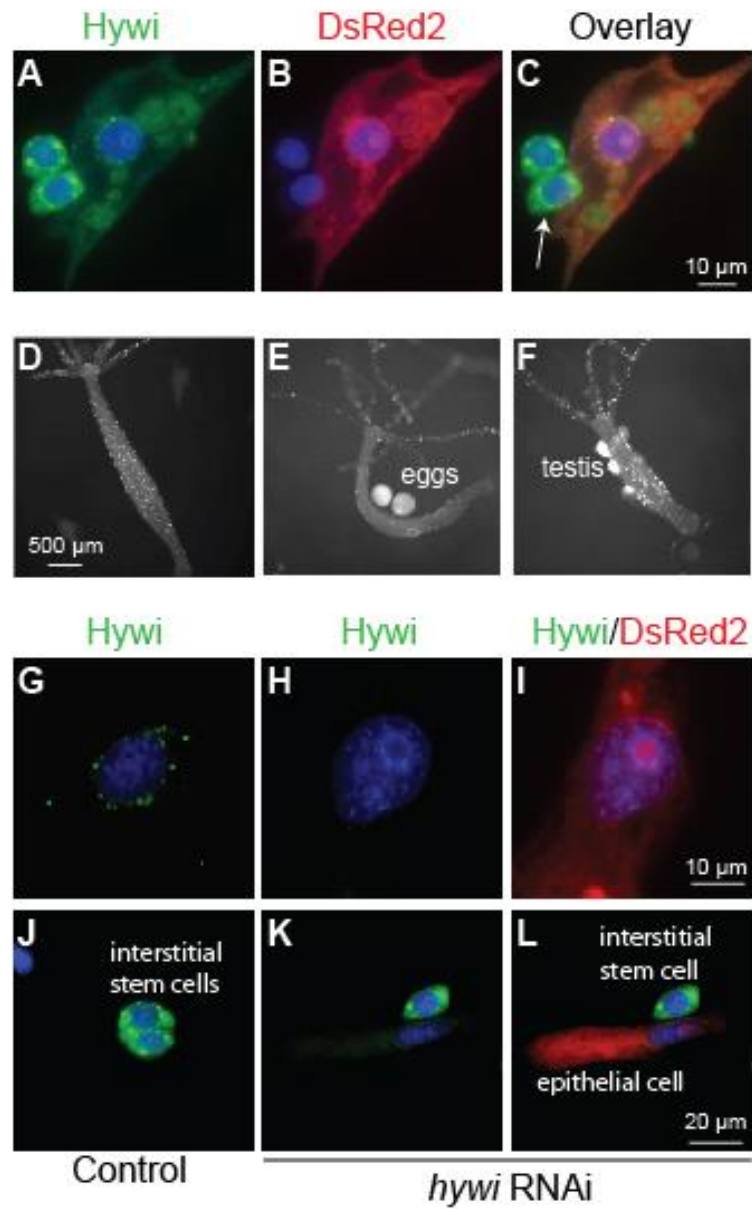
Supplemental Figure 6: Analysis of lineage-specific small RNAs.



(A) Size distribution of lineage specific small RNAs. (B) Analysis of nucleotide distribution across the length of small RNAs between 26 and 34 nucleotides long from each lineage (interstitial, ectoderm, and endoderm) demonstrates a strong preference for uridine at the 5' position for small RNAs from all three lineages. (C) Small RNAs between 26 and 34 nucleotides long isolated from the interstitial lineage have a higher frequency of complementary overlap 10 bases from their 5' end compared to small RNAs of the same length isolated from the ectodermal or endodermal lineages. (D-E) In order to compare the numbers of small RNAs mapping to the same transcript across the three lineage datasets and to compare piRNA mapping numbers between different transcripts, the mapped reads were normalized similarly to RPKM (Mortazavi, *et al.*, 2008) (here piPKM) and log transformed. (D) The interstitial lineage has many more small RNAs mapping to transposons [average log(piPKM) in the interstitial lineage is 0.48] compared with the epithelial lineages [average 0.11, -0.02 in the ectoderm and endoderm lineages

respectively]. (E) Non transposon transcripts have many fewer small RNAs mapping to them and there is no lineage specific difference [average $\log(\text{piPKM})$ is -0.74, -0.51, -0.60 in interstitial, ectoderm and endoderm lineages respectively].

Supplemental Figure 7: Transmission of the *hywi* RNAi-1 transgene through the germline and knockdown of *hywi* in the epithelial cells of F1 hatchlings.



(A-C) A transgenic line was established that uniformly expresses DsRed2 under control of an actin promoter in all three lineages. This was accomplished by sexual transmission of the transgene. Double labeling with antibodies against Hywi and DsRed2 demonstrates that the transgene is expressed in the epithelial cells, but not in the interstitial stem cells (arrow in panel C). (D) Stable lines were created expressing the *hywi* RNAi-1 or *hywi* RNAi-2 transgene (Fig. 4A,B) in the interstitial lineage under the control of the actin promoter, which can be observed by DsRed2 expression in the differentiated cells of the lineage. (E,F) In one line *hywi* RNAi-1 is transmitted through the germline. (G-I) The resulting F1 *Hydra* hatchlings do not express Hywi in the epithelial cells (H,I), but Hywi protein is detected in nontransgenic F1 siblings (G). (J-L) Hywi protein is still detected in the interstitial stem cells of *hywi* knockdown hatchlings because the actin promoter is not active in these cells. Transgenic cells are identified by labeling with an antibody against DsRed2. DNA is labeled with Hoechst 33342. (M) *hywi* mRNA levels were tested by qRT-PCR at several time points after hatching and eating in *hywi* knockdown F1 *Hydra* as compared to wild type F1 sibling controls (normalized to actin).

Supplemental Table 1: piRNA mapping to the *Hydra* transcriptome.

| | 5'UTR | | ORF | | 3'UTR | |
|--------------------------------|-------|-----|-----|-----|-------|-----|
| Average Transcript | 10% | | 76% | | 14% | |
| Hywi piRNAs Transposons | 23% | | 55% | | 22% | |
| | 6% | 17% | 14% | 41% | 6% | 16% |
| Hyli piRNAs Transposons | 22% | | 49% | | 29% | |
| | 15% | 7% | 41% | 8% | 20% | 9% |
| Hywi piRNAs non-Transposons | 11% | | 66% | | 23% | |
| | 4% | 7% | 59% | 7% | 10% | 13% |
| Hyli piRNAs non-Transposons | 18% | | 65% | | 17% | |
| | 15% | 3% | 61% | 4% | 9% | 8% |

The numbers correspond to the bar graph in Figure 3E. For transposon transcripts, the majority of Hywi-bound piRNAs are mapped in the anti-sense orientation (yellow boxes) and the majority of Hyli-bound piRNAs are mapped in the sense orientation (grey boxes). The majority of both Hywi- and Hyli-bound piRNAs that map to non-transposon transcripts map in the sense orientation (yellow boxes).

Supplemental Table 2: Gene ontology analysis of transcripts with greater than 10 Hywi-bound piRNAs mapped.

| Category | Term | Count | Fold Enrichment | P-Value |
|----------|---|-------|-----------------|----------|
| BP | Gastrulation With Mouth Forming First | 2 | 25.4 | 7.70E-02 |
| BP | Nucleosome Assembly | 5 | 6.7 | 5.60E-03 |
| BP | Neuropeptide Signaling Pathway | 12 | 6.6 | 1.10E-06 |
| BP | Chromatin Assembly | 5 | 6.3 | 6.80E-03 |
| BP | Translational Elongation | 9 | 6.3 | 5.80E-05 |
| BP | Nucleosome Organization | 5 | 6 | 8.20E-03 |
| BP | Protein-DNA Complex Assembly | 5 | 5.5 | 1.10E-02 |
| BP | DNA packaging | 6 | 4.8 | 7.50E-03 |
| BP | Cell Growth | 4 | 4.6 | 5.30E-02 |
| BP | Epithelial Cell Differentiation | 4 | 4.4 | 5.90E-02 |
| BP | Induction of Apoptosis by Extracellular Signals | 4 | 3.9 | 8.02E-02 |
| BP | Chromatin Assembly or Disassembly | 5 | 3.7 | 4.30E-02 |
| BP | Cellular Respiration | 7 | 3 | 2.90E-02 |
| BP | Translation | 30 | 3 | 1.70E-07 |
| BP | ATP biosynthetic process | 6 | 2.9 | 5.20E-02 |
| BP | Purine Nucleoside Triphosphate Biosynthetic Process | 6 | 2.7 | 7.20E-02 |
| BP | Purine Ribonucleoside Triphosphate Biosynthetic Process | 6 | 2.7 | 7.20E-02 |
| BP | Ribonucleoside Triphosphate Biosynthetic Process | 6 | 2.7 | 7.20E-02 |
| BP | ATP Metabolic Process | 6 | 2.6 | 8.10E-02 |
| BP | di-, tri-valent inorganic cation transport | 8 | 2.6 | 3.10E-02 |
| BP | Nucleoside Triphosphate Biosynthetic Process | 6 | 2.6 | 7.60E-02 |
| BP | Sensory Perception of Light Stimulus | 6 | 2.6 | 7.60E-02 |
| BP | Visual Perception | 6 | 2.6 | 7.60E-02 |
| BP | Electron Transport Chain | 9 | 2.5 | 2.40E-02 |
| BP | Heart Development | 7 | 2.2 | 9.30E-02 |
| BP | Chromatin Organization | 13 | 2.1 | 2.20E-02 |
| BP | Chromosome Organization | 15 | 1.8 | 3.70E-02 |
| BP | Generation of Precursor Metabolites and Energy | 13 | 1.8 | 5.00E-02 |

| | | | | |
|----|---|----|------|----------|
| BP | GPCR Signaling | 16 | 1.8 | 2.70E-02 |
| BP | Regulation of Cell Proliferation | 12 | 1.7 | 7.90E-02 |
| BP | Cation Transport | 15 | 1.6 | 7.70E-02 |
| BP | Regulation of Biological Quality | 29 | 1.6 | 1.30E-02 |
| BP | Anatomical Structure Development | 41 | 1.4 | 2.90E-02 |
| BP | Cell Surface Receptor Linked Signal Transduction | 25 | 1.4 | 6.30E-02 |
| BP | Cellular Protein Metabolic Process | 66 | 1.4 | 4.30E-03 |
| BP | Protein Metabolic Process | 73 | 1.3 | 1.20E-02 |
| BP | Signal Transduction | 38 | 1.3 | 4.70E-02 |
| BP | System Development | 35 | 1.3 | 9.10E-02 |
| BP | Establishment of Localization | 60 | 1.2 | 4.40E-02 |
| BP | Localization | 65 | 1.2 | 4.20E-02 |
| BP | Transport | 59 | 1.2 | 5.50E-02 |
| MF | Sodium:Potassium-Exchange ATPase Activity | 3 | 24.8 | 4.70E-03 |
| MF | rRNA binding | 4 | 4.5 | 5.60E-02 |
| MF | Structural Constituent of Ribosome | 27 | 4.5 | 9.40E-11 |
| MF | Cytochrome-C Oxidase Activity | 4 | 4.1 | 7.00E-02 |
| MF | Heme-Copper Terminal Oxidase Activity | 4 | 4.1 | 7.00E-02 |
| MF | Oxidoreductase Activity, Acting on Heme Group of Donors | 4 | 4.1 | 7.00E-02 |
| MF | Oxidoreductase Activity, Acting on Heme Group of Donors, Oxygen as Receptor | 4 | 4.1 | 7.00E-02 |
| MF | Structural Molecule Activity | 34 | 3.4 | 4.80E-10 |
| MF | Inorganic Cation Transmembrane Transporter Activity | 12 | 3.2 | 1.10E-03 |
| MF | Monovalent Inorganic Cation Transmembrane Transporter Activity | 8 | 3.1 | 1.30E-02 |
| MF | Cysteine-Type Endopeptidase Activity | 5 | 2.8 | 9.90E-02 |
| MF | Hydrogen Ion Transmembrane Transporter Activity | 6 | 2.5 | 9.20E-02 |
| MF | Cation Transmembrane Transporter Activity | 19 | 1.9 | 9.70E-03 |
| MF | G-protein Coupled Receptor Activity | 14 | 1.9 | 2.80E-02 |
| MF | GTP binding | 14 | 1.9 | 3.70E-02 |
| MF | Guanyl Nucleotide Binding | 14 | 1.8 | 4.00E-02 |
| MF | Guanyl Ribonucleotide Binding | 14 | 1.8 | 4.00E-02 |
| MF | Calcium Ion Binding | 27 | 1.7 | 8.50E-03 |

| | | | | |
|----|---|-----|------|----------|
| MF | Ion Transmembrane Transporter Activity | 21 | 1.7 | 1.80E-02 |
| MF | Transmembrane Receptor Activity | 19 | 1.7 | 3.00E-02 |
| MF | Substrate-Specific Transporter Activity | 26 | 1.6 | 2.20E-02 |
| MF | Substrate-Specific Transmembrane Transporter Activity | 22 | 1.6 | 3.80E-02 |
| MF | Transporter Activity | 32 | 1.6 | 9.70E-03 |
| MF | Transmembrane Transporter Activity | 24 | 1.5 | 3.70E-02 |
| MF | Receptor Activity | 24 | 1.4 | 7.70E-01 |
| CC | Kinesin Complex | 5 | 10.4 | 9.40E-04 |
| CC | Nucleosome | 5 | 9.6 | 1.30E-03 |
| CC | Protein-DNA Complex | 6 | 6.5 | 1.80E-03 |
| CC | Cytosolic Large Ribosomal Subunit | 5 | 5 | 1.60E-02 |
| CC | Cytosolic Ribosome | 8 | 4.5 | 1.60E-03 |
| CC | Ribosome | 28 | 4.1 | 4.10E-10 |
| CC | Respiratory Chain | 8 | 3.6 | 6.40E-03 |
| CC | Cytosolic Part | 9 | 3.3 | 5.80E-03 |
| CC | Large Ribosomal Subunit | 6 | 3.3 | 3.50E-02 |
| CC | Ribosomal Subunit | 10 | 3.3 | 3.00E-03 |
| CC | Clatherin-Coated Vesicle | 7 | 2.5 | 5.60E-02 |
| CC | Ribonucleoprotein Complex | 32 | 2.3 | 2.00E-05 |
| CC | Extracellular Space | 8 | 2.2 | 7.00E-02 |
| CC | Extracellular Region Part | 14 | 1.9 | 3.60E-02 |
| CC | Chromosome | 15 | 1.8 | 3.30E-02 |
| CC | Cytoplasmic Membrane-Bounded Vesicles | 13 | 1.7 | 8.30E-02 |
| CC | Vesicle | 17 | 1.7 | 4.80E-02 |
| CC | Cytoplasmic Vesicle | 16 | 1.6 | 6.70E-02 |
| CC | Cytosol | 28 | 1.5 | 2.40E-02 |
| CC | Intracellular Non-Membrane-Bounded Organelle | 66 | 1.5 | 2.00E-04 |
| CC | Non-Membrane-Bounded Organelle | 66 | 1.5 | 2.00E-04 |
| CC | Plasma Membrane | 59 | 1.4 | 2.20E-03 |
| CC | Cytoplasmic Part | 114 | 1.3 | 7.10E-04 |
| CC | Intrinsic to Membranes | 81 | 1.2 | 8.10E-02 |
| CC | Macromolecular Complex | 66 | 1.2 | 6.00E-02 |
| CC | Cytoplasm | 149 | 1.1 | 1.40E-02 |

Categories highlighted in green are enriched only for transcripts with Hywi-bound mapping piRNAs and categories highlighted in purple are enriched for transcripts with both Hywi- and Hyli-bound mapping piRNAs. BP – Biological Process, MF – Molecular Function, CC – Cellular Component.

Supplemental Table 3: Gene ontology analysis of transcripts with greater than 10 Hyli-bound piRNAs mapped.

| Category | Term | Count | Fold Enrichment | P-Value |
|----------|--|-------|-----------------|----------|
| BP | De Novo Posttranslational Protein Folding | 3 | 10.1 | 3.10E-02 |
| BP | De Novo Protein Folding | 3 | 10.1 | 3.10E-02 |
| BP | Translational Elongation | 18 | 8.4 | 2.90E-12 |
| BP | Mitotic Spindle Elongation | 7 | 7.8 | 1.30E-04 |
| BP | Spindle Elongation | 7 | 7.8 | 1.30E-04 |
| BP | Ribosomal Small Subunit Biogenesis | 3 | 7.2 | 6.10E-02 |
| BP | Nucleosome Assembly | 6 | 5.3 | 4.20E-03 |
| BP | Chromatin Assembly | 6 | 5.0 | 5.30E-03 |
| BP | Mitotic Spindle Organization | 7 | 4.9 | 2.30E-03 |
| BP | Dentrite Morphogenesis | 4 | 4.8 | 4.60E-02 |
| BP | Nucleosome Organization | 6 | 4.8 | 6.60E-03 |
| BP | Imaginal Disc Development | 5 | 4.7 | 1.90E-02 |
| BP | Protein-DNA complex Assembly | 6 | 4.4 | 9.90E-03 |
| BP | ATP Synthesis Coupled Proton Transport | 7 | 4.1 | 6.20E-03 |
| BP | Energy Coupled Proton Transport, Down Electrochemical Gradient | 7 | 4.1 | 6.20E-03 |
| BP | Spindle Organization | 9 | 4.0 | 1.40E-03 |
| BP | Instar Larval or Pupal Morphogenesis | 4 | 3.9 | 7.60E-02 |
| BP | Translation | 59 | 3.9 | 3.40E-20 |
| BP | DNA Packaging | 7 | 3.7 | 1.00E-02 |
| BP | Metamorphosis | 4 | 3.7 | 8.70E-02 |
| BP | Translational Initiation | 6 | 3.7 | 2.00E-02 |
| BP | Ion Transmembrane Transport | 7 | 3.6 | 1.20E-02 |
| BP | Proton Transport | 7 | 3.6 | 1.20E-02 |
| BP | Hydrogen Transport | 7 | 3.5 | 1.40E-02 |
| BP | Oxidative Phosphorylation | 10 | 3.2 | 3.10E-03 |
| BP | Negative Regulation of Transport | 6 | 3.1 | 4.30E-02 |
| BP | Aerobic Respiration | 5 | 3.0 | 8.10E-02 |
| BP | Chromatin Assembly or Dissassembly | 6 | 3.0 | 4.80E-02 |
| BP | ATP biosynthetic process | 8 | 2.6 | 3.30E-02 |
| BP | Nucleoside Triphosphate Biosynthetic Process | 9 | 2.6 | 2.00E-02 |
| BP | Purine Nucleoside Triphosphate Biosynthetic Process | 9 | 2.6 | 1.80E-02 |
| BP | Purine Ribonucleoside Triphosphate Biosynthetic Process | 9 | 2.6 | 1.80E-02 |
| BP | Ribonucleoside Triphosphate Biosynthetic | 9 | 2.6 | 1.80E-02 |

| | Process | | | |
|----|--|-----|-----|----------|
| BP | Cellular Respiration | 9 | 2.5 | 2.40E-02 |
| BP | Ribonucleotide Biosynthetic Process | 10 | 2.4 | 2.40E-02 |
| BP | ATP Metabolic Process | 8 | 2.3 | 5.90E-02 |
| BP | Purine Nucleoside Triphosphate Metabolic Process | 9 | 2.3 | 4.40E-02 |
| BP | Purine Ribonucleoside Triphosphate Metabolic Process | 9 | 2.3 | 3.70E-02 |
| BP | Purine Ribonucleotide Biosynthetic Process | 9 | 2.3 | 4.40E-02 |
| BP | Ribonucleoside Triphosphate Metabolic Process | 9 | 2.3 | 3.70E-02 |
| BP | Electron Transport Chain | 12 | 2.2 | 1.70E-02 |
| BP | Generation of Precursor Metabolic and Energy | 24 | 2.2 | 3.60E-04 |
| BP | Microtubule Cytoskeleton Organization | 10 | 2.2 | 3.80E-02 |
| BP | Nucleotide Biosynthetic Process | 13 | 2.2 | 1.40E-02 |
| BP | Purine Nucleotide Biosynthetic Process | 10 | 2.2 | 3.80E-02 |
| BP | Nucleobase, Nucleoside and Nucleic Acid Biosynthetic Process | 13 | 2.1 | 1.70E-02 |
| BP | Nucleobase, Nucleoside and Nucleotide Biosynthetic Process | 13 | 2.1 | 1.70E-02 |
| BP | Nucleoside Triphosphate Metabolic Process | 9 | 2.1 | 6.60E-02 |
| BP | Ribonucleotide Metabolic Process | 10 | 2.1 | 5.00E-02 |
| BP | Purine Ribonucleotide Metabolic Process | 9 | 2 | 8.50E-02 |
| BP | Chromatin Organization | 17 | 1.8 | 2.60E-02 |
| BP | Cytoskeleton Organization | 17 | 1.8 | 2.70E-02 |
| BP | Microtubule-Based Process | 16 | 1.8 | 2.90E-02 |
| BP | Cellular Macromolecular Complex Assembly | 13 | 1.7 | 8.80E-02 |
| BP | Cellular Macromolecular Complex Subunit Organization | 14 | 1.6 | 8.50E-02 |
| BP | Chromosome Organization | 20 | 1.6 | 4.40E-02 |
| BP | Nitrogen Compound Biosynthetic Process | 18 | 1.6 | 5.10E-02 |
| BP | Cellular Protein Metabolic Process | 111 | 1.5 | 9.30E-07 |
| BP | Mitotic Cell Cycle | 17 | 1.5 | 9.80E-02 |
| BP | Protein Metabolic Process | 121 | 1.4 | 1.50E-05 |
| BP | Cellular Biosynthetic Process | 107 | 1.3 | 2.90E-03 |
| BP | CC Biogenesis | 33 | 1.3 | 7.40E-02 |
| BP | Cellular Macromolecule Biosynthetic Process | 85 | 1.3 | 3.60E-03 |
| BP | Gene Expression | 93 | 1.3 | 1.60E-03 |
| BP | Macromolecule Biosynthetic Process | 85 | 1.3 | 3.90E-03 |
| BP | Organelle Organization | 45 | 1.3 | 5.30E-02 |
| BP | Biosynthetic Process | 107 | 1.2 | 7.40E-03 |

| | | | | |
|----|--|-----|------|----------|
| BP | Cellular Metabolic Process | 196 | 1.1 | 7.30E-02 |
| MF | Structural Constituent of Ribosome | 48 | 5.5 | 3.30E-23 |
| MF | Translation Elongation Factor Activity | 5 | 4.7 | 1.90E-02 |
| MF | rRNA Binding | 6 | 4.6 | 8.10E-03 |
| MF | Cytochrome-C Oxidase Activity | 6 | 4.2 | 1.20E-02 |
| MF | Heme-Copper Terminal Oxidase Activity | 6 | 4.2 | 1.20E-02 |
| MF | Oxidoreductase Activity, Acting on Heme Group of Donors | 6 | 4.2 | 1.20E-02 |
| MF | Oxidoreductase Activity, Acting on Heme Group of Donors, Oxygen as Acceptor | 6 | 4.2 | 1.20E-02 |
| MF | Antioxidant Activity | 4 | 4 | 7.60E-02 |
| MF | Monovalent Inorganic Cation Transmembrane Transporter Activity | 15 | 4.0 | 1.30E-05 |
| MF | Hydrogen Ion Transmembrane Transporter Activity | 14 | 3.9 | 3.60E-05 |
| MF | Structural Molecule Activity | 57 | 3.9 | 1.10E-19 |
| MF | Histone Methyltransferase Activity | 6 | 3.1 | 4.30E-02 |
| MF | Inorganic Cation Transmembrane Transporter Activity | 16 | 2.9 | 2.90E-04 |
| MF | N-methyltransferase Activity | 6 | 2.6 | 7.80E-02 |
| MF | Microtubule Motor Activity | 7 | 2.3 | 7.90E-02 |
| MF | Motor Activity | 10 | 2.2 | 4.00E-02 |
| MF | Cation Transmembrane Transporter Activity | 24 | 1.6 | 1.90E-02 |
| MF | Ion Transmembrane Transporter Activity | 25 | 1.4 | 8.50E-02 |
| CC | Mitochondrial Proton-Transporting ATP Synthase Complex, Coupling Factor F(o) | 3 | 12.5 | 2.00E-02 |
| CC | Polytene Chromosome | 3 | 12.5 | 2.00E-02 |
| CC | Cytosolic Large Ribosomal Subunit | 11 | 7.4 | 7.30E-07 |
| CC | Mitochondrial Proton-Transporting ATP Synthase Complex | 3 | 7.2 | 6.10E-02 |
| CC | Proton-Transporting V-Type ATPase Complex | 3 | 7.2 | 6.10E-02 |
| CC | Kinesin Complex | 5 | 7.0 | 4.20E-03 |
| CC | Cytosolic Ribosome | 17 | 6.5 | 1.60E-09 |
| CC | Cytosolic Small Ribosomal Subunit | 5 | 5.6 | 1.00E-02 |
| CC | Lipid Particle | 5 | 5.6 | 1.00E-02 |
| CC | Proton-Transporting ATP synthesis complex, Coupling Factor F(o) | 5 | 5.6 | 1.00E-02 |
| CC | Nucleosome | 4 | 5.1 | 3.80E-02 |
| CC | Ribosome | 50 | 4.9 | 8.10E-22 |
| CC | Small Ribosomal Subunit | 8 | 4.5 | 1.60E-03 |
| CC | Large Ribosomal Subunit | 12 | 4.4 | 5.80E-05 |
| CC | Proton-Transporting Two-Sector ATPase | 5 | 4.4 | 2.40E-02 |

| | | | | |
|----|--|-----|-----|----------|
| | Complex, Proton-Transporting Domain | | | |
| CC | Ribosomal Subunit | 20 | 4.4 | 5.60E-08 |
| CC | Contractile Fiber | 6 | 3.9 | 1.70E-02 |
| CC | Cytosolic Part | 16 | 3.9 | 9.40E-06 |
| CC | Proton-Transport Two-Sector ATPase complex | 7 | 3.8 | 8.90E-03 |
| CC | Contractile Fiber Part | 5 | 3.6 | 4.50E-02 |
| CC | Protein-DNA Complex | 5 | 3.6 | 4.50E-02 |
| CC | Proton-Transporting ATP Synthase Complex | 5 | 3.5 | 5.10E-02 |
| CC | Microtubule Associated Complex | 11 | 3.2 | 2.00E-03 |
| CC | Ribonucleoprotein Complex | 60 | 2.8 | 8.40E-14 |
| CC | Mitochondrial Membrane Part | 7 | 2.5 | 5.40E-02 |
| CC | Respiratory Chain | 8 | 2.4 | 4.70E-02 |
| CC | Mitochondrial Inner Membrane | 21 | 1.9 | 8.50E-03 |
| CC | Organelle Inner Membrane | 21 | 1.8 | 1.50E-02 |
| CC | Cytosol | 46 | 1.7 | 4.60E-04 |
| CC | Intracellular Non-Membrane-Bounded Organelle | 108 | 1.7 | 8.50E-09 |
| CC | Non-Membrane-Bounded-Organelle | 108 | 1.7 | 8.50E-09 |
| CC | Mitochondrial Envelope | 23 | 1.6 | 2.60E-02 |
| CC | Mitochondrial Membrane | 22 | 1.6 | 2.90E-02 |
| CC | Envelope | 30 | 1.5 | 2.70E-02 |
| CC | Macromolecular Complex | 123 | 1.5 | 3.20E-07 |
| CC | Organelle Envelopes | 28 | 1.4 | 6.20E-01 |
| CC | Cytoplasmic Part | 173 | 1.3 | 9.50E-06 |
| CC | Mitochondrion | 51 | 1.3 | 2.50E-02 |
| CC | Cytoplasm | 235 | 1.2 | 1.70E-05 |
| CC | Intracellular | 297 | 1.1 | 1.60E-02 |
| CC | Intracellular Organelle | 256 | 1.1 | 2.80E-03 |
| CC | Intracellular Part | 295 | 1.1 | 1.80E-02 |
| CC | Organelle | 257 | 1.1 | 2.20E-03 |

Categories highlighted in blue are enriched only for transcripts with Hyli-bound mapping piRNAs and categories highlighted in purple are enriched for transcripts with both Hywi- and Hyli-bound mapping piRNAs. BP – Biological Process, MF – Molecular Function, CC – Cellular Component.

Supplemental Table 4: Gene ontology analysis of putative lineage-specific targets of the PIWI-piRNA pathway.

| Endoderm | | | | |
|-----------------|---|--------------|------------------------|----------------|
| Category | Term | Count | Fold Enrichment | P-Value |
| BP | proteolysis | 6 | 2.9 | 4.10E-02 |
| CC | collagen | 4 | 196.7 | 5.10E-07 |
| CC | proteinaceous extracellular matrix | 7 | 16.9 | 1.80E-06 |
| CC | extracellular matrix | 7 | 16.6 | 2.10E-06 |
| CC | extracellular region part | 8 | 10.5 | 4.70E-06 |
| CC | extracellular region | 10 | 6.5 | 6.00E-06 |
| CC | extracellular matrix part | 5 | 30 | 1.60E-05 |
| CC | collagen type I | 2 | 245.9 | 7.80E-03 |
| CC | fibrillar collagen | 2 | 163.9 | 1.20E-02 |
| CC | cell surface | 3 | 8.4 | 4.50E-02 |
| CC | extracellular space | 3 | 8.1 | 4.80E-02 |
| MF | extracellular matrix structural constituent | 4 | 62.8 | 2.60E-05 |
| MF | peptidase activity, acting on L-amino acid peptides | 6 | 5.7 | 2.50E-03 |
| MF | peptidase activity | 6 | 5.5 | 3.00E-03 |
| MF | endopeptidase activity | 4 | 5.6 | 2.90E-02 |
| MF | structural molecule activity | 4 | 4.4 | 5.50E-02 |
| MF | SMAD binding | 2 | 31.4 | 5.90E-02 |
| MF | metallopeptidase activity | 3 | 6.9 | 6.40E-02 |

| Ectoderm | | | | |
|-----------------|--|--------------|------------------------|----------------|
| Category | Term | Count | Fold Enrichment | P-Value |
| BP | signal transduction | 11 | 1.8 | 7.00E-02 |
| BP | cell surface receptor linked signal transduction | 9 | 2.4 | 2.90E-02 |
| BP | biological adhesion | 5 | 3.3 | 5.90E-02 |
| BP | cell adhesion | 5 | 3.3 | 5.90E-02 |
| BP | G-protein coupled receptor protein signaling pathway | 7 | 3.7 | 1.00E-02 |
| BP | neuropeptide signaling pathway | 4 | 10.2 | 6.50E-03 |
| BP | negative regulation of angiogenesis | 2 | 21.2 | 8.90E-02 |
| CC | membrane | 31 | 1.5 | 2.10E-03 |
| CC | membrane part | 27 | 1.5 | 1.10E-02 |
| CC | intrinsic to membrane | 24 | 1.7 | 4.20E-03 |
| CC | integral to membrane | 23 | 1.7 | 6.30E-03 |
| CC | plasma membrane part | 10 | 2 | 5.30E-02 |
| CC | plasma membrane | 19 | 2.3 | 4.90E-04 |
| CC | extracellular region | 12 | 3.9 | 1.40E-04 |
| CC | intrinsic to plasma membrane | 7 | 3.9 | 8.00E-03 |
| CC | integral to plasma membrane | 7 | 4 | 6.90E-03 |
| CC | extracellular region part | 7 | 4.6 | 3.50E-03 |
| CC | extracellular matrix | 6 | 7.1 | 1.30E-03 |
| CC | proteinaceous extracellular matrix | 6 | 7.2 | 1.20E-03 |
| CC | extracellular matrix part | 4 | 12 | 4.10E-03 |
| CC | fibril | 2 | 61.5 | 3.10E-02 |
| CC | microfibril | 2 | 82 | 2.40E-02 |
| MF | metal ion binding | 22 | 1.5 | 3.30E-02 |
| MF | cation binding | 22 | 1.5 | 3.80E-02 |
| MF | ion binding | 22 | 1.5 | 4.20E-02 |
| MF | molecular transducer activity | 13 | 2.7 | 2.00E-03 |
| MF | signal transducer activity | 13 | 2.7 | 2.00E-03 |
| MF | receptor activity | 13 | 3.5 | 2.00E-04 |
| MF | carbohydrate binding | 4 | 4.2 | 6.70E-02 |
| MF | calcium ion binding | 15 | 4.4 | 3.20E-06 |
| MF | transmembrane receptor activity | 11 | 4.6 | 9.20E-05 |
| MF | G-protein coupled receptor activity | 8 | 5.1 | 7.20E-04 |
| MF | pattern binding | 3 | 7.3 | 6.00E-02 |
| MF | polysaccharide binding | 3 | 7.3 | 6.00E-02 |
| MF | peptide receptor activity, G-protein coupled | 3 | 9.1 | 4.10E-02 |
| MF | peptide receptor activity | 3 | 9.1 | 4.10E-02 |
| MF | extracellular matrix structural constituent | 4 | 27.1 | 3.60E-04 |

| Epithelium | | | | |
|-------------------|--|--------------|------------------------|----------------|
| Category | Term | Count | Fold Enrichment | P-Value |
| BP | positive regulation of DNA binding | 3 | 24 | 6.40E-03 |
| BP | amine metabolic process | 7 | 3.7 | 9.90E-03 |
| BP | positive regulation of binding | 3 | 18.3 | 1.10E-02 |
| BP | transport | 20 | 1.7 | 1.30E-02 |
| BP | establishment of localization | 20 | 1.7 | 1.40E-02 |
| BP | localization | 21 | 1.6 | 1.70E-02 |
| BP | oxidation reduction | 9 | 2.6 | 1.70E-02 |
| BP | regulation of DNA binding | 3 | 12.5 | 2.30E-02 |
| BP | cholesterol metabolic process | 3 | 10.7 | 3.00E-02 |
| BP | Ras protein signal transduction | 3 | 10 | 3.40E-02 |
| BP | gas transport | 2 | 51.9 | 3.70E-02 |
| BP | sterol metabolic process | 3 | 9.4 | 3.80E-02 |
| BP | response to chemical stimulus | 8 | 2.4 | 3.90E-02 |
| BP | biogenic amine metabolic process | 3 | 8 | 5.20E-02 |
| BP | electron transport chain | 4 | 4.6 | 5.30E-02 |
| BP | cellular amine metabolic process | 5 | 3.3 | 6.00E-02 |
| BP | regulation of binding | 3 | 6.9 | 6.70E-02 |
| BP | positive regulation of NF-kappaB transcription factor activity | 2 | 25.9 | 7.30E-02 |
| BP | cellular amino acid and derivative metabolic process | 5 | 2.9 | 8.80E-02 |
| BP | generation of precursor metabolites and energy | 5 | 2.9 | 8.90E-02 |
| BP | regulation of multicellular organismal process | 6 | 2.5 | 9.00E-02 |
| BP | Rho protein signal transduction | 2 | 20.8 | 9.10E-02 |
| BP | positive regulation of transcription factor activity | 2 | 20.8 | 9.10E-02 |
| BP | alcohol metabolic process | 5 | 2.9 | 9.10E-02 |
| BP | amine biosynthetic process | 3 | 5.8 | 9.20E-02 |
| BP | steroid metabolic process | 3 | 5.6 | 9.80E-02 |
| BP | nitrogen compound biosynthetic process | 5 | 2.8 | 9.90E-02 |
| CC | extracellular region | 9 | 2.4 | 2.80E-02 |
| CC | extracellular region part | 6 | 3.2 | 3.40E-02 |
| CC | collagen | 2 | 40.7 | 4.70E-02 |
| CC | vacuole | 5 | 3.5 | 5.20E-02 |
| CC | external side of plasma membrane | 3 | 7.8 | 5.40E-02 |
| CC | proteinaceous extracellular matrix | 4 | 4 | 7.50E-02 |
| CC | extracellular matrix | 4 | 3.9 | 7.90E-02 |
| CC | cytoplasmic part | 28 | 1.3 | 9.00E-02 |

| | | | | |
|----|---|----|------|----------|
| MF | cofactor binding | 6 | 3.9 | 1.70E-02 |
| MF | active transmembrane transporter activity | 6 | 3.7 | 2.00E-02 |
| MF | endopeptidase activity | 6 | 3.2 | 3.80E-02 |
| MF | calcium ion binding | 9 | 2.3 | 3.90E-02 |
| MF | secondary active transmembrane transporter activity | 4 | 5.1 | 4.10E-02 |
| MF | vitamin binding | 4 | 4.7 | 5.10E-02 |
| MF | peptidase activity, acting on L-amino acid peptides | 7 | 2.5 | 5.60E-02 |
| MF | inorganic anion transmembrane transporter activity | 2 | 33.2 | 5.80E-02 |
| MF | peptidase activity | 7 | 2.4 | 6.50E-02 |
| MF | oxidoreductase activity | 8 | 2.2 | 6.70E-02 |
| MF | cysteine-type endopeptidase activity | 3 | 6.8 | 7.00E-02 |
| MF | oxidoreductase activity | 3 | 6.5 | 7.50E-02 |
| MF | cytoskeletal protein binding | 5 | 3 | 8.10E-02 |
| MF | coenzyme binding | 4 | 3.9 | 8.10E-02 |
| MF | cation binding | 23 | 1.3 | 1.00E-01 |

| Interstitial cells | | | | |
|---------------------------|--|--------------|------------------------|----------------|
| Category | Term | Count | Fold Enrichment | P-Value |
| BP | DNA integration | 3 | 33 | 3.40E-03 |
| BP | M phase | 6 | 5 | 5.40E-03 |
| BP | response to stress | 9 | 2.9 | 7.60E-03 |
| BP | cell cycle phase | 6 | 4.5 | 8.40E-03 |
| BP | protein processing | 3 | 15 | 1.60E-02 |
| BP | protein maturation | 3 | 14.5 | 1.70E-02 |
| BP | cell division | 5 | 4.7 | 1.90E-02 |
| BP | cellular response to stress | 6 | 3.7 | 2.00E-02 |
| BP | cell cycle process | 6 | 3.6 | 2.10E-02 |
| BP | cell cycle | 7 | 3 | 2.30E-02 |
| BP | mitosis | 4 | 5.3 | 3.60E-02 |
| BP | nuclear division | 4 | 5.3 | 3.60E-02 |
| BP | M phase of mitotic cell cycle | 4 | 5.2 | 3.70E-02 |
| BP | regulation of BP | 19 | 1.5 | 4.10E-02 |
| BP | organelle fission | 4 | 5 | 4.20E-02 |
| BP | positive regulation of BP | 8 | 2.3 | 4.30E-02 |
| BP | regulation of response to stress | 3 | 8.2 | 4.80E-02 |
| BP | cellular response to stimulus | 6 | 2.9 | 4.90E-02 |
| BP | DNA recombination | 3 | 8.1 | 5.00E-02 |
| BP | signal transduction | 9 | 2.1 | 5.20E-02 |
| BP | MAPKKK cascade | 3 | 7.7 | 5.40E-02 |
| BP | positive regulation of catalytic activity | 4 | 4.3 | 6.00E-02 |
| BP | regulation of multicellular organismal process | 5 | 3.2 | 6.00E-02 |
| BP | response to stimulus | 10 | 1.8 | 6.80E-02 |
| BP | biological regulation | 19 | 1.4 | 7.40E-02 |
| BP | positive regulation of MF | 4 | 3.9 | 7.50E-02 |
| BP | intracellular signaling cascade | 6 | 2.4 | 9.00E-02 |
| BP | DNA metabolic process | 5 | 2.8 | 9.30E-02 |
| BP | regulation of localization | 4 | 3.5 | 9.50E-02 |
| BP | regulation of response to stimulus | 3 | 5.5 | 9.80E-02 |
| BP | mitotic cell cycle | 4 | 3.5 | 1.00E-01 |
| CC | extracellular region | 7 | 3 | 2.20E-02 |
| MF | histone-lysine N-methyltransferase activity | 3 | 15.7 | 1.50E-02 |
| MF | protein-lysine N-methyltransferase activity | 3 | 15.7 | 1.50E-02 |
| MF | lysine N-methyltransferase activity | 3 | 15.7 | 1.50E-02 |
| MF | histone methyltransferase activity | 3 | 13.3 | 2.00E-02 |
| MF | N-methyltransferase activity | 3 | 11.3 | 2.70E-02 |

| | | | | |
|----|---|----|------|----------|
| MF | DNA binding | 10 | 2.2 | 3.00E-02 |
| MF | protein methyltransferase activity | 3 | 10.5 | 3.10E-02 |
| MF | S-adenosylmethionine-dependent methyltransferase activity | 3 | 7.5 | 5.80E-02 |
| MF | magnesium ion binding | 5 | 3.1 | 7.30E-02 |

Lineage-specific small RNAs 23 nucleotides or greater in length were mapped to the *Hydra* transcriptome. Transcripts with at least 10 times more small RNAs mapping from a specific lineage were identified as putative targets specific to that lineage. The category “epithelium” is a combination of both ectodermal and endodermal small RNAs as compared to interstitial small RNAs, thus identifying putative mRNAs that are targeted in both epithelial layers, but not in the interstitial lineage. BP – Biological Process, MF – Molecular Function, CC – Cellular Component.

Supplemental Table 5: Real-time quantitative PCR to test *hywi* knockdown levels

| Gene | Forward | Reverse |
|--------|----------------------|----------------------|
| Actin | AAGCTCAGAGCAAACGTGGT | GGACAGGGTGTTCTTCTGGA |
| GAPDH | GACAACCATTCATGCCACAA | ACAGCTTTTGCAGCTCCAGT |
| Hywi-1 | CCACAACCTCCTGTTGGAGT | TGAGCAGTTTGCTGAGGTTG |
| Hywi-2 | ACCCAAGGACCAATCCTTTT | AAATTTTTCGCACGCATCTC |
| Hyli | GCCCTGGAAACACCTATGAA | GGATGAGTGCCCATCACTT |

**Appendix IV: Deadenylase depletion protects inherited mRNAs in primordial germ
cells**

S. Zachary Swartz, **Adrian Reich**, Nathalie Oulhen, Tal Raz, Patrice M. Milos, Joseph P.
Campanale, Amro Hamdoun, Gary Wessel

Submitted, *Development*

CONTRIBUTION

I sequenced, assembled and annotated the *de novo* transcriptome and conducted large scale bioinformatic analyses. This data is found in Figure 2A,B, which I prepared. I also constructed a database used throughout the project, shown in part, in Supplemental Tables 4, 5.

ABSTRACT

A critical event in animal development is the specification of primordial germ cells (PGCs), which become the stem cells that create sperm and eggs. How PGCs are created provides a valuable paradigm for understanding stem cells in general. We find that the PGCs of the sea urchin *Strongylocentrotus purpuratus* exhibit broad transcriptional repression, yet enrichment for a set of inherited mRNAs. Enrichment of several germ line determinants in the PGCs requires the RNA binding protein Nanos to deplete the transcript encoding CNOT6, a deadenylase, in the PGCs, thereby creating a stable environment for RNA. Misexpression of CNOT6 in the PGCs results in their failure to retain *Seawi* transcripts and Vasa protein. Conversely, broad knockdown of CNOT6 expands the domain of *Seawi* RNA as well as exogenous reporters. Thus, Nanos-dependent spatially restricted CNOT6 differential expression is used to selectively localize germ line RNAs to the PGCs. Our findings support a “time-capsule” model of germ line determination, whereby the PGCs are insulated from differentiation by retaining the molecular characteristics of the totipotent egg and early embryo.

INTRODUCTION

The germ line provides an immortal link between generations by transmitting heritable information from parent to progeny. Specification of the animal germ line typically occurs during embryogenesis, when primordial germ cells (PGCs) fated to become the gamete-producing stem cells of the adult are segregated from somatic lineages. PGCs in the numerous species studied share common molecular signatures, including the RNA helicase Vasa, the translational repressor Nanos, and the argonaute family member Piwi (Ewen-Campen *et al.*, 2010). Surprisingly, despite this conservation of gene expression, the developmental routes that lead to it are remarkably diverse. Strategies for PGC segregation can be considered within a continuum of inherited and inductive mechanisms. An example of the inherited mode is that of *Drosophila melanogaster*, whose PGCs are the first cells to form in the embryo. Their specification involves maternally supplied determinant mRNAs and proteins, collectively called a germ plasm, which is actively transported and inherited by the presumptive PGCs. Conversely, PGCs in the mouse are specified by inductive signaling originating from the extraembryonic ectoderm. Most knowledge regarding these disparate mechanisms comes from studies in *Drosophila*, *C. elegans*, zebrafish, and mice (Ewen-Campen *et al.*, 2010; Pehrson and Cohen, 1986; Seydoux and Braun, 2006; Tanaka and Dan, 1990; Yajima and Wessel, 2011, 2012).

Very little, comparatively, is known outside of these groups. Echinoderms, a large and diverse phylum, form part of the sister group to the chordates. The best-studied examples of this group are the sea urchins, including *Strongylocentrotus purpuratus*. In this animal, four PGCs called small micromeres (sMics) are created by an asymmetric division at the 5th embryonic cleavage (Ewen-Campen *et al.*, 2010; Juliano *et al.*, 2006; Pehrson and Cohen, 1986; Seydoux and Braun, 2006; Tanaka and Dan, 1990; Yajima and Wessel, 2011, 2012) (Fig. 3c). After their formation, the sMics divide once during gastrulation to give 8 descendants. These cells then assort into larval niches called the coelomic pouches, which are the major contributors to the juvenile sea urchin (Pehrson and Cohen, 1986; Tanaka and Dan, 1990). The early creation of the

sMics is perhaps suggestive of inherited specification; however, sea urchins do not possess a classically defined germ plasm of aggregated germ line determinants. Instead, germ line RNAs, including those of Vasa and Seawi (a piwi family member) are maternally deposited and broadly distributed in early embryos, and later refined to the sMics during gastrulation (Juliano *et al.*, 2006; Seydoux and Braun, 2006).

To better understand the specification of the sMics, we used a transcriptomic approach to identify sMic-enriched mRNAs. We learned that the sMics are broadly transcriptionally repressed and identified transcripts, like Vasa and Seawi, which are ubiquitous in the early embryo but later turned over in somatic cells to result in sMic enrichment. The expression dynamics of these discovered transcripts imply a post-transcriptional mechanism by which sMics retain RNA. Transcriptome analysis identified the mRNA encoding the CCR4-related deadenylase CNOT6 as uniquely depleted in the sMics. This depletion is dependent upon the RNA binding protein Nanos and sequence elements in the CNOT6 3'UTR that match the highly conserved binding consensus for Pumilio, the binding partner of Nanos. Depletion of CNOT6 is required for retention of Vasa protein and Seawi RNA in the sMics.

RESULTS

Differential expression analysis and identification of sMic enriched transcripts

We developed a method for isolating PGCs *en masse*. In *S. purpuratus*, the sMics selectively retain the fluorescent dye calcein due to altered multidrug transporter activity (Campanale and Hamdoun, 2012) (Fig. 1a,b), even when dissociated into single cell suspensions (Fig. 1c). We purified sMics by FACS, which comprised approximately 0.5% of the total population of the embryo (Fig. 1d, Supplemental Fig. 1a-f). By qPCR, isolated calcein positive cells were 16-fold enriched for Nanos, a known sMic specific transcript, but not enriched for Spec, an ectodermal transcript (Supplemental Fig. 1g). With this indication of purity, total RNA was then isolated and deep sequenced without amplification from three biological replicates:

isolated sMics, non-sMics, and disaggregated whole embryos. After assessing variation between samples by MDS analysis, we performed differential expression analysis to discover sMic enriched and depleted transcripts (Supplemental Fig. 1h-j and Supplemental Information). In summary, with a significance cutoff of 0.05 (false discovery rate, FDR), we identified a union set of 230 differentially expressed transcripts (both sMic enriched and depleted) between these comparisons (Fig. 1e, Supplemental Fig. 1i,j, and Supplemental Tables 4, 5).

The sMic-enriched transcripts included Nanos and Delta, which were previously identified as sMic-localized, as well as SpG-cadherin, which is required for sMic fate in *S. purpuratus* and is enriched in the sMics of *Lytechinus variegatus* (Juliano *et al.*, 2010; Miller and McClay, 1997; Oliveri *et al.*, 2002; Yajima and Wessel, 2012) The sMic-enriched transcripts fell into diverse functional categories, but transcriptional regulation and RNA binding were overrepresented by gene set enrichment analysis (Supplemental Table 1). Several sMic transcripts lie within the same pathway. For example, we identified the DNA binding factor Baf250 and the ATPase Brg1, which both assemble into the pluripotency-associated esBAF chromatin remodeling complex (Lessard and Crabtree, 2010). In addition to Delta, we also identified MibL, a potential regulator of Notch/Delta signaling (Le Borgne and Schweisguth, 2003).

The sMics are broadly transcriptionally repressed

Surprisingly, the majority of sMic enriched transcripts we discovered appear to be maternally deposited. To examine temporal expression dynamics of sMic enriched genes, we used a microarray dataset containing the whole-embryo expression level of all genes at multiple time points (Wei *et al.*, 2006). sMic transcripts on average are at maximum abundance by at the 2 and 15 h.p.f. time points (Fig. 2a). sMic transcripts then sharply drop in abundance by 30 h.p.f., just after the onset of gastrulation. By calculating the 30/15 hour abundance ratio, we find that sMic transcripts are statistically overrepresented for decreasing abundance compared to the whole-transcriptome average and sMic-depleted transcripts (Fig. 2b). These dynamics suggest maternal loading followed by broad turnover. sMic nuclei are depleted for RNA polymerase II

phosphorylated at serine 2 of the C-terminal domain, a marker of transcriptional elongation (Fig. 2c-d) (Seydoux and Dunn, 1997). This depletion is first apparent at blastula stage, and persists through gastrulation. Furthermore, sMic nuclei are highly enriched for histone 3 lysine 9 trimethylation (H3K9me3), a heterochromatin marker (Fig. 2e-f). Thus, we infer sMics are transcriptionally repressed relative to their somatic neighbors. Furthermore, we did not detect transcripts that accumulate exclusively in the sMics. In this regard, Nanos remains the unique exception (Fig. 3a). Rather, sMic transcripts were broadly detectable in eggs and early embryos until blastula to gastrula stage, when broad turnover occurs in all cells except the sMics. Such genes include Baf250, Ctdsp12/SCP2, an RNA polymerase II (RNAPII) C-terminal domain phosphatase, z62, a C2H2 zinc finger protein, and MibL (Fig. 3b and Supplemental Fig. 2a-f). These observations suggest sMics retain maternally loaded and zygotically transcribed transcripts, which are cleared from somatic cells during the blastula to gastrula transition. Given this repression and the sMic transcript dynamics, we suggest sMics inherit and retain their select mRNAs rather than actively transcribe them (with the important exception of Nanos).

CNOT6 transcript is selectively degraded in the sMics by a Nanos/Pumilio dependent mechanism

Our transcriptome analysis and in situ hybridizations imply a mechanism by which the sMics stably retain inherited RNA. In addition, we previously found that RNA microinjected into the fertilized egg is degraded in resultant somatic cells but retained in sMics, independently of its sequence (Gustafson and Wessel, 2010; Oulhen *et al.*, 2013). Our study offers an explanation for these phenomena: the top sMic-depleted transcript encodes CNOT6 (FDR=2.40E-13), an ortholog of CCR4, a broadly conserved deadenylase subunit of the conserved CCR4/POP2/NOT complex (Collart and Panasenko, 2012). As a major regulator of RNA stability, depletion of CNOT6 could enhance RNA retention. By FISH, CNOT6 transcript is detectable ubiquitously in eggs through 32/60-cell embryos (Fig. 3d-f). But by blastula stage, the transcript is uniquely depleted in the sMics (Fig. 3g-i). Nanos is a strong candidate for mediating this depletion. Nanos

and its binding partner Pumilio recognize highly conserved motifs in the 3'UTRs of target transcripts, termed Pumilio Response Elements (PREs), leading to mRNA degradation (Wreden *et al.*, 1997). The three *S. purpuratus* Nanos paralogs are each expressed in the sMics and required for their survival (Juliano *et al.*, 2010). Furthermore, the 3'UTR of CNOT6 contains two PRE sequences, suggesting it may be a Nanos/Pumilio target (Chen *et al.*, 2012; Gerber *et al.*, 2006; White *et al.*, 2001). We therefore knocked down Nanos with a translation-blocking morpholino antisense oligo (MASO) targeting the two most abundant of the three paralogs. This previously characterized MASO results in the eventual death of the sMics, and its effects were rescued by expression of a MASO-insensitive Nanos construct, demonstrating specificity (Juliano *et al.*, 2010). In morphant embryos, CNOT6 mRNA accumulated in sMics (Fig. 4a,b). We next tested the two putative PREs as sequence-specific targets with MASOs complementary to these two sites, which we predicted would occlude binding of Nanos/Pumilio. Consistent with Nanos knockdown, PRE-protecting MASOs caused retention of CNOT6 mRNA in the sMics (Fig. 4c). To further test the motifs, we used reporter constructs containing either the full-length wild type CNOT6 3'UTR, or with the PREs mutated singly or in combination (Fig. 4d-h). The wild type reporter recapitulated the sMic exclusion of the endogenous transcript (Fig. 4e); however, mutations of the PREs resulted in sMic retention (Fig. 4f-h). In further support of its role, Pumilio co-immunoprecipitates with Nanos (Supplemental Fig. 3a). We attempted to test the effect of Pumilio knockdown on CNOT6 accumulation; however, these embryos were developmentally arrested before blastula stage, likely pointing to pleiotropic effects (data not shown). Indeed, Nanos-independent roles for Pumilio have been identified (Van Etten *et al.*, 2012; Weidmann and Goldstrohm, 2012). Indicative of diverse functions, Pumilio protein is detectable in granules in all cells of the early blastula, and highly enriched in the Veg2 mesodermal precursors of later blastulae (Supplemental Fig. 3c,d). Nanos mediated degradation of CNOT6 transcript is surprising, especially in light of evidence that Nanos functions by recruiting the CCR4-NOT complex itself (Suzuki *et al.*, 2012). One possibility is that the CCR4-NOT complex maintains

functionality with only the Pop2-related nuclease CNOT7, which is present in sMics at the transcript level (Supplemental Fig. 2g). Alternatively, maternal CNOT6 protein may be initially available in the sMics to degrade the mRNA, but then turned over later. Both possibilities are consistent with our conclusion that Nanos/Pumilio directs degradation of maternal CNOT6 mRNA in the sMics via PRE motifs in its 3'UTR.

CNOT6 repression is required for retention of germ line determinants

To test the requirement of CNOT6 repression for germ line fate, we misexpressed CNOT6 in the sMics by multiple approaches. A CNOT6::mCherry fusion construct that expresses in all cells significantly reduced accumulation of a previously characterized sMic reporter, *Vasa::GFP* (Fig. 5a-c) (Gustafson *et al.*, 2011). We next tested the accumulation of endogenous *Vasa* protein. Embryos were fixed at 42 h.p.f., following gastrulation and 1 division of the sMics resulting in approximately 8 descendants on average. Both CNOT6::mCherry, as well as PRE-protecting MASOs, predicted to stabilize endogenous CNOT6 in the sMics, resulted in significantly fewer *Vasa* protein positive sMics (Fig. 5d-g). Endogenous *Vasa* is likely lost later than the reporter construct because of the abundance of maternally supplied *Vasa* protein, whereas injection of reporter RNA requires new translation (Voronina *et al.*, 2008a). Additionally, we tested the transcript abundance of the endogenous Argonaute family member *Seawi* in the sMics with CNOT6 overexpression. *Seawi* transcript is normally present in all cells but highly enriched in the sMics (Yajima *et al.*, 2013). With CNOT6::mCherry expression, the sMics lose enrichment for *Seawi* RNA (Fig. 6a-c). To determine whether the sMics die or lose their inherited determinants, we stably labeled their lineage by EdU incorporation. Due to their slow cell cycle, sMics retain EdU pulsed before first cleavage, compared to other more rapidly dividing cells, and enables definitive lineage analysis (Tanaka and Dan, 1990). Both control and CNOT6 overexpressing embryos possessed similar numbers of sMics, indicating a loss of inherited determinants rather than cell death at this stage (Fig. 6d). We conclude that CNOT6 depletion is necessary for retention of germ line determinants in the sMics.

Repression of the CNOT6 deadenylase may allow for increased background stability for inherited RNAs in the sMics. Therefore, we predicted that global CNOT6 knockdown would expand the domain of RNA retention. CNOT6 is required for normal development of the embryo; strong knockdown leads to profound endomesodermal defects (Supplemental Fig. 3e,f). We therefore tested the effects of weaker CNOT6 knockdown (under which development proceeds relatively normally) on Seawi transcript localization. In control embryos, Seawi transcripts are highly enriched in the sMics. However, when we globally reduce CNOT6 protein with either of two non-overlapping MASOs, Seawi transcripts are more broadly retained throughout the endomesoderm and oral ectoderm (Fig. 6e-h). To test the generality of CNOT6 mediated RNA retention, we used exogenous RNA encoding mCherry with an SV40 3' polyadenylation signal. As reported for other exogenous RNAs, this transcript is retained in the sMics but degraded in somatic cells in a sequence independent manner (Fig. 7a, Gustafson and Wessel, 2010; Oulhen and Wessel, 2013). However, when CNOT6 is globally depleted, mCherry RNA is retained throughout the endomesoderm (Fig. 7b,c). Our results indicate that differential CNOT6 expression is critical for proper accumulation of transcripts within the sMics and is required for normal development of somatic lineages.

DISCUSSION

Our study reveals mechanistic insight into the divergence of germ line from soma. Uniformly dispersed mRNAs in the early embryo become highly asymmetric by the selective Nanos/Pumilio repression of CNOT6 in the germ cell precursors. This paradigm explains the localization of known sMic-enriched mRNAs, including Vasa and Seawi (Juliano *et al.*, 2006), as well as foreign transcripts introduced in the early embryo (Gustafson and Wessel, 2010; Oulhen and Wessel, 2013). The RNA retention mechanism via CNOT6 depletion raises the question: is there specificity in the RNAs that sMics retain, or is it completely nonselective? While our dataset is likely not complete, if retention were completely nonselective, one would expect to

identify more than the 78 sMic enriched transcripts we report. It is possible that the sMics possess mechanisms to exclude RNAs—Nanos/Pumilio is one such example, though there may be others that remain uncharacterized. Indeed, we bioinformatically identified numerous transcripts that are depleted in the sMics, and in the future, it will be important to investigate the mechanisms of their depletion. The fact that sMics generally retain RNA is likely necessary because they are transcriptionally quiescent. Transcriptional repression has also been documented in *Drosophila*, *C. elegans*, ascidians, and mice, indicating it is a fundamental feature of germ line segregation (Nakamura and Seydoux, 2008; Shirae-Kurabayashi *et al.*, 2011). Surprisingly, however, each organism achieves repression via distinct mechanisms. While the precise nature of sMic repression is unknown, we find that the RNAPII phosphatase Ctdspl2 is sMic enriched, pointing to one possible mechanism.

Broad clearance of maternal RNA is a hallmark of the maternal to zygotic transition (MZT), a conserved event when developmental control is passed to the embryo. In *Drosophila*, there are two phases of degradation: the first occurs following egg activation and involves the RNA-binding protein Smaug, which recruits the CCR4-NOT complex to degrade diverse targets (Semotok *et al.*, 2005). The PGCs possess degradation activity, but certain transcripts are protected in the PGCs by motifs in their 3'UTRs, perhaps by Oskar association (Bashirullah *et al.*, 1999; Zaessinger *et al.*, 2006). A second degradation process is driven by the mir-309 cluster at about 2 hours post fertilization (Bushati *et al.*, 2008). The piRNA pathway also contributes to degradation of Nanos transcript in somatic cells and involves the CCR4-NOT complex (Rouget *et al.*, 2010). In zebrafish, a primary effector of maternal RNA degradation at the MZT is mir-430 (Giraldez *et al.*, 2006). It was further observed that some mir-430 targets, such as Nanos, are degraded in the soma but protected in the PGCs (Mishima *et al.*, 2006). In the sea urchin, the zygotic genome activates shortly after fertilization. However, the degradation aspect of the MZT in the future soma may be conserved via CNOT6, and could include small RNA mechanisms (Song *et al.*, 2012).

Prior to this study, the only known mechanisms for stabilizing RNA in the germ line were via Dead end 1 (Dnd1), and Dazl, which work by occluding microRNA binding sites and promoting deadenylation, respectively (Kedde *et al.*, 2007; Mishima *et al.*, 2006; Takeda *et al.*, 2009). However, Dnd1 is not conserved outside of vertebrates. We show here in an early branching deuterostome that the deeply conserved deadenylase CNOT6 is repressed in its PGCs by Nanos/Pumilio, allowing for stable retention of inherited transcripts. Since the PGCs are transcriptionally repressed, their inheritance may represent a “time capsule” of early development; that is, they must subsist solely on the mRNAs they retain from the egg, independently of a germ plasm (Supplemental Fig. 4). A consequence of this strategy may be that the PGCs remain insulated from differentiation into somatic lineages. Furthermore, our model is consistent with an immortal cytoplasm hypothesis for the evolutionary origin of the segregated germ line at the transition from unicellular to multicellular animal life. The ancestral single-celled organism likely possessed the hugely conserved factors found in animal germ lines, which were retained at the transition to multicellularity (Extavour and Akam, 2003). It is the somatic cells that acquired unique characters to diversify from the original, progenitor cell type, while sacrificing reproductive potential (Buss, 1987; Extavour, 2007). The diversification of the soma may have necessitated the evolution of global turnover events (e.g. the MZT) to eliminate RNAs associated with the egg. Instead of acquiring gametogenic capability anew, the embryonic germ cells remain protected and retain the characteristics of the egg. Downregulation of deadenylase activity provides a mechanism for understanding how cytoplasm that confers gametogenic potential is preserved in the segregated germ line.

MATERIALS AND METHODS

Animals

Strongylocentrotus purpuratus were kept in aquaria containing artificial seawater at 16°C. Individuals were induced to shed gametes by shaking or injection of 0.5M KCl. Eggs were

collected in filtered seawater (FSW) and sperm was collected dry. Eggs were fertilized in the presence of 1 mM 3-amino-triazol (3-AT) (Sigma) to prevent crosslinking of fertilization envelopes. Embryos were reared at 15°C at a density of about 0.2% (packed egg volume / seawater volume) in stirring culture vessels.

FACS isolation of sMics

Embryos were collected 15 h.p.f. by straining through 45 micron Nitex® and concentrated to about 0.5% density in 50 ml FSW. PSC833 (Novartis) and Calcein AM (C-AM, Invitrogen) were added to the FSW at 500 nM and 250 nM, respectively. The embryos were then incubated for 90 minutes at 15°C with constant rotation in 50 mL conical tubes. Embryos were pelleted by centrifuging at 250xg for 30 seconds, washed twice in 50 mL of calcium free seawater, and then resuspended in 10 mL 1 M glycine, 25mM EDTA solution. After incubating 5 minutes on ice, the embryos were disaggregated by trituration through a transfer pipette 20 times. The single cell suspension was pelleted by centrifugation at 250xg for 5 minutes at 4°C, washed three times in calcium free seawater to a final sample volume of 4mL. PSC833 was then added to 1uM final concentration. The cell suspension was sorted on an Aria FACS instrument set to 4 degrees. For long sorts, staggered cultures were fertilized at 2-hour intervals, and then labeled and disaggregated at 15 h.p.f. to avoid cell death and changes to their transcriptional profile. Cells were first gated by forward and side scatter to remove debris and aggregates, and then sorted by fluorescence intensity versus forward scatter (Supplemental Fig. 1a-f). Cells were sorted directly into 0.75 mL Trizol LS (Invitrogen) until the total volume reached 1 mL. Sorts, numbers of collected cells, and RNA extraction yields are found in Supplemental Table 2.

Helicos sample preparation and deep sequencing

RNA was extracted using Trizol-LS reagent as described by the manufacturer (Invitrogen). RNA was treated using RQ1 DNase (Promega) for 30 minutes at 37°C, then extracted with acid Phenol:Chloroform (Ambion). Three biological replicates of paired sMic, non-sMic, and whole embryo RNA were collected. Two replicate pairs were pooled from 3

separate sorts, while the third was collected from a single sort. RNA yields are found in Supplemental Table 2. Total RNA was stored in 100% ethanol, processed for RNA-seq without amplification or poly-A selection, and sequenced by Helicos tSMS (Cambridge, MA; Supplemental Table 3; Lipson *et al.*, 2009).

Illumina sample preparation, deep sequencing, and reference transcriptome assembly

RNA was extracted from several developmental stages, including ovary, 32 cell stage, 15hr blastula, 41 hr gastrula, and 4-day pluteus, using the RNEasy Mini kit (Qiagen) with on column DNase. The isolated RNA was processed using standard procedures using the Illumina mRNA-Seq kit and sequenced on a single lane of a GAIIX using a read length of 105bp, paired-end. The transcriptome was assembled using Velvet (1.0.09) and Oases (0.1.14) with a k-mer of 31 (Schulz *et al.*, 2012). Exemplar sequences were selected from each locus based on abundance with a minimum length cutoff. The exemplar sequences were annotated with Blast2GO and compared by BLAST (minimum score of $1e^{-5}$) with the *S. purpuratus* SPU gene predictions (Conesa *et al.*, 2005; Sea Urchin Genome Sequencing *et al.*, 2006).

Differential expression analysis

Raw Helicos read files were aligned to an Illumina mixed developmental stage *de novo* transcriptome using the Helisphere DGE pipeline. Total and mapped reads for each sample are summarized in Supplemental Table 3. RMS counts for all transcripts were then used for differential expression analysis using the edgeR Bioconductor package (Robinson *et al.*, 2010). Replicates were first filtered to remove low expressing (< 5 total counts summed between sMic, non-sMic, and WE samples) and high expressing (> 10,000 summed counts; primarily rRNA contamination) transcripts. Counts were TMM normalized, and differentially expressed transcripts were identified using tagwise dispersion. To assess variation between replicates, we performed multidimensional scaling (MDS) analysis using edgeR. Replicate 2 and 3 samples were highly related with whole embryo, sMic, and non-sMic transcriptomes clustering separately. However, the replicate 1 sample was an outlier, with poor separation between sMic and non-sMic

samples (Supplemental Fig. 1h). This variation likely points to issues with: sample handling (RNA collection, or RNA-seq sample preparation), less efficient calcein labeling, or the inclusion of 3 crosses rather than 1 or 2. In subsequent differential expression analysis, we found that comparing sMics to non-sMics with Replicate 1 included only yielded 3 significant transcripts (Nanos, and two noncoding RNAs; data not shown). We therefore performed differential expression analysis both with all three replicates, as well as with replicate 1 excluded. Smear plot depictions of differential expression between sMics and whole embryo (replicate 1 included) and sMics and non-sMics (replicate 1 excluded) are provided (Supplemental Fig. 1i,j). Differentially expressed transcripts were annotated with Blast2GO, the corresponding SPU genomic locus identifier (www.spbse.org), and temporal expression levels from a previous microarray study (Wei *et al.*, 2006). Annotations are provided for the comparison between sMics and non-sMics (replicate 1 excluded; Supplemental Table 4 and Supplemental Table 5). Gene set enrichment analysis was performed using the topGO Bioconductor package (Alexa *et al.*, 2006).

Whole mount in situ hybridization (WMISH) and immunofluorescence

WMISH was performed as described previously (Juliano *et al.*, 2006). Approximately 1kb antisense probe templates were PCR amplified from cDNA using a reverse primer tailed with the T7 promoter (Supplemental Table 6). Digoxigenin labeled antisense probes were transcribed using the Roche DIG RNA labeling kit according to the manufacturer's instructions. Embryos of mixed developmental stages were fixed with MOPS buffered PFA and hybridized for at least 5 days at 50°C with 70% formamide and 0.5ng/ul probe. Hybridization was then visualized using either NBT/BCIP chromogenic detection, or tyramide fluorescence amplification (TSA plus system, Perkin Elmer). Vasa immunofluorescence was performed as described previously (Voronina *et al.*, 2008b). Rabbit antibodies to RNAP pSer2 and H3K9me3 were obtained from Abcam (ab5095, ab8898, respectively). Pumilio immunofluorescence was performed with a commercial rabbit polyclonal antibody raised to the conserved PUF domain of Human Pumilio 2 (GTX114172, GeneTex, Irvine, CA). As the Vasa antibody was also raised in rabbit, co-labeling

(e.g. Fig. 2) was performed as follows: first, embryos were incubated with Vasa antibody overnight at room temperature in PBST (0.05% Triton-X 100, pH 7.4). The embryos were washed 4 times, and then incubated 3 hours at room temperature with rhodamine-labeled goat anti-rabbit Fab fragments. This procedure was then repeated sequentially with the pSer2 or H3K9me3 antibodies with FITC-labeled Fab fragments.

Cloning and Reporter constructions

DNA fragments were PCR amplified and cloned by standard methods. For overexpression studies, constructs were built in a modified pCS2 vector containing additional rare 8-base cutting restriction enzyme sites (Gokirmak *et al.*, 2012). PRE mutant reporter constructs were generated by site-directed mutagenesis with the QuikChange II kit (Agilent Technologies; Supplemental Table 7).

Morpholino antisense oligo (MASO) and mRNA microinjection

Custom MASOs were synthesized by Gene-Tools (Philomath, OR; Supplemental Table 8). The control MASO targets a divergent Nanos ortholog in a distantly related sea star species. MASO injection solutions contained 20% glycerol and 100ug/ml 10,000 MW Texas Red-Lysine or FITC dextran. Synthetic mRNAs were transcribed using the mMessage mMachine SP6 or T7 kit (Ambion) from linearized plasmid template. Eggs were dejellied by incubating 10 minutes in pH 5.0 seawater and rowed on protamine sulfate-coated petri dishes. Fertilized eggs were injected with 2 picoliters of MASO or RNA by constant pressure injection in the presence of 1mM 3-AT (Sigma). The injected zygotes were then washed into filtered seawater and incubated at 15°C.

Western blot

After fertilization, at the indicated times, cells were collected and lysed in SDS-PAGE loading buffer. Western blot analyses were performed following electrophoretic transfer of proteins from SDS-PAGE onto 0.22 µm nitrocellulose membranes (Towbin *et al.*, 1979). The single sea urchin Pumilio ortholog was analyzed using a rabbit polyclonal antibody to human Pumilio 2 (GTX114172, GeneTex, Irvine, CA). The 180 kDa Pumilio band was also recognized

by an independently raised antibody to human Pumilio 1 (PA5-30327, Thermo Scientific, Rockford, IL). Actin was used as a loading control (A5060, Sigma Aldrich, St. Louis, MO). Briefly, membranes were incubated one hour in the blocking solution (TBS-Tween, 4% BSA). The anti-pumilio antibody (1/500) or the anti-actin (1/1000) were added overnight at 4C. The antigen-antibody complex was detected by chemiluminescence.

Immunoprecipitation

600 μ l of mesenchyme blastula pellet was lysed in 600 μ l of IP buffer (50mM Tris pH 7.6, 100mM NaCl, 1% NP40) with anti-proteases (Roche) using a dounce homogenizer. Cell lysate was centrifuged for 15 minutes at 15,000g at 4C. 500 μ l of supernatant was used for each immunoprecipitation. The supernatants were diluted by two in the IP buffer and pre-cleared for 1 hour with 60 μ l of protein A magnetic Dynabeads (Life technologies, Carlsbad, CA). The resulting supernatants were incubated for 2 hours at 4C with either the antibody against Sp Nanos (1/500) (Juliano *et al.*, 2010), or control antibody against Sp Vasa (1/500). Then, 80 μ l of pre-washed Dynabeads were added to each tube. After one hour at 4C, the beads were washed three times in the IP buffer, and two times in 50mM Tris pH7.6, 100mM NaCl. After washing, bound proteins were eluted with 30 μ l of SDS-PAGE buffer. 10 μ l of each IP were loaded on an SDS Page gel for either western blot, or silver staining using the Silver Stain for Mass spectrometry kit (Pierce, Rockford, IL, USA).

DATA AVAILABILITY

All Helicos reads, *de novo* transcriptome reads, and the assembled transcriptome are available under BioProject PRJNA188114.

REFERENCES

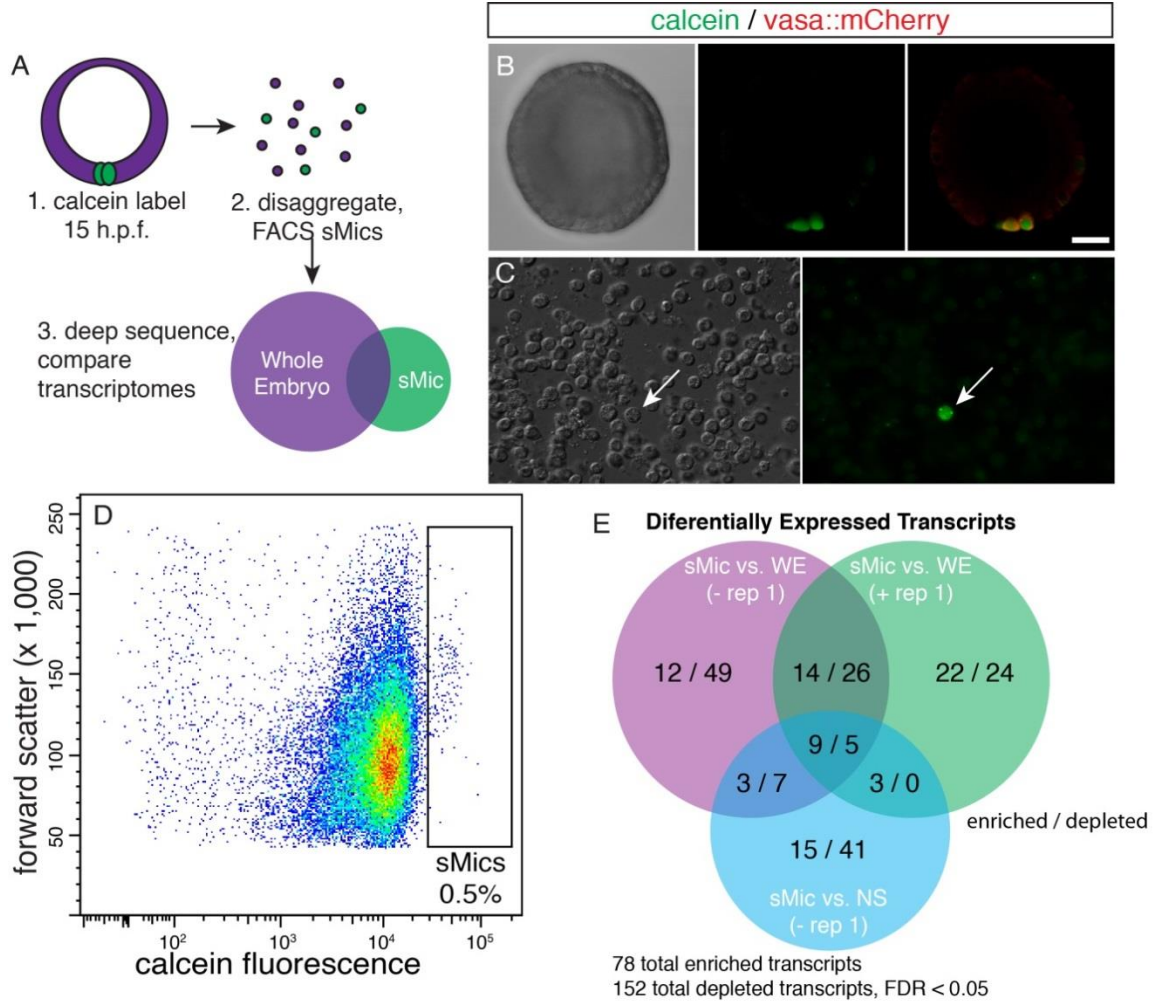
- Alexa, A., Rahnenfuhrer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22, 1600-1607.
- Bashirullah, A., Halsell, S.R., Cooperstock, R.L., Kloc, M., Karaiskakis, A., Fisher, W.W., Fu, W., Hamilton, J.K., Etkin, L.D., and Lipshitz, H.D. (1999). Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in *Drosophila melanogaster*. *EMBO J* 18, 2610-2620.
- Bushati, N., Stark, A., Brennecke, J., and Cohen, S.M. (2008). Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in *Drosophila*. *Curr Biol* 18, 501-506.
- Buss, L.W. (1987). Modern zoophytology: the growth and form of modular organisms. *Science* 237, 1626-1627.
- Campanale, J.P., and Hamdoun, A. (2012). Programmed reduction of ABC transporter activity in sea urchin germline progenitors. *Development* 139, 783-792.
- Chen, D., Zheng, W., Lin, A., Uyhazi, K., Zhao, H., and Lin, H. (2012). Pumilio 1 suppresses multiple activators of p53 to safeguard spermatogenesis. *Curr Biol* 22, 420-425.
- Collart, M.A., and Panasenko, O.O. (2012). The Ccr4--not complex. *Gene* 492, 42-53.
- Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674-3676.
- Ewen-Campen, B., Schwager, E.E., and Extavour, C.G.M. (2010). The molecular machinery of germ line specification. *Mol Reprod Dev* 77, 3-18.
- Extavour, C.G. (2007). Evolution of the bilaterian germ line: lineage origin and modulation of specification mechanisms. *Integr Comp Biol* 47, 770-785.
- Extavour, C.G., and Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* 130, 5869-5884.
- Gerber, A.P., Luschnig, S., Krasnow, M.A., Brown, P.O., and Herschlag, D. (2006). Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 103, 4487-4492.
- Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 312, 75-79.
- Gokirmak, T., Campanale, J.P., Shipp, L.E., Moy, G.W., Tao, H., and Hamdoun, A. (2012). Localization and substrate selectivity of sea urchin multidrug (MDR) efflux transporters. *J Biol Chem* 287, 43876-43883.
- Gustafson, E.A., and Wessel, G.M. (2010). Exogenous RNA is selectively retained in the small micromeres during sea urchin embryogenesis. *Mol Reprod Dev* 77, 836-836.
- Gustafson, E.A., Yajima, M., Juliano, C.E., and Wessel, G.M. (2011). Post-translational regulation by gustavus contributes to selective Vasa protein accumulation in multipotent cells during embryogenesis. *Dev Biol* 349, 440-450.
- Juliano, C.E., Voronina, E., Stack, C., Aldrich, M., Cameron, A.R., and 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., and Wessel, G.M. (2010). Nanos functions to maintain the fate of the small micromere lineage in the sea urchin embryo. *Dev Biol* 337, 220-232.
- Kedde, M., Strasser, M.J., Boldajipour, B., Oude Vrielink, J.A.F., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P.M., van Duijse, J., Orom, U.A., *et al.* (2007). RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* 131, 1273-1286.
- Le Borgne, R., and Schweisguth, F. (2003). Notch signaling: endocytosis makes delta signal better. *Curr Biol* 13, R273-275.

- Lessard, J.A., and Crabtree, G.R. (2010). Chromatin regulatory mechanisms in pluripotency. *Annu Rev Cell Dev Biol* 26, 503-532.
- Lipson, D., Raz, T., Kieu, A., Jones, D.R., Giladi, E., Thayer, E., Thompson, J.F., Letovsky, S., Milos, P., and Causey, M. (2009). Quantification of the yeast transcriptome by single-molecule sequencing. *Nat Biotechnol* 27, 652-658.
- Miller, J.R., and McClay, D.R. (1997). Characterization of the role of cadherin in regulating cell adhesion during sea urchin development. *Dev Biol* 192, 323-339.
- Mishima, Y., Giraldez, A.J., Takeda, Y., Fujiwara, T., Sakamoto, H., Schier, A.F., and Inoue, K. (2006). Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430. *Curr Biol* 16, 2135-2142.
- Nakamura, A., and Seydoux, G. (2008). Less is more: specification of the germline by transcriptional repression. *Development* 135, 3817-3827.
- Oliveri, P., Carrick, D.M., and Davidson, E.H. (2002). A regulatory gene network that directs micromere specification in the sea urchin embryo. *Dev Biol* 246, 209-228.
- Oulhen, N., and Wessel, G.M. (2013). Retention of exogenous mRNAs selectively in the germ cells of the sea urchin requires only a 5'-cap and a 3'-UTR. *Mol Reprod Dev* 80, 561-569.
- Oulhen, N., Yoshida, T., Yajima, M., Song, J.L., Sakuma, T., Sakamoto, N., Yamamoto, T., and Wessel, G.M. (2013). The 3'UTR of nanos2 directs enrichment in the germ cell lineage of the sea urchin. *Dev Biol* 377, 275-283.
- Pehrson, J.R., and Cohen, L.H. (1986). The fate of the small micromeres in sea urchin development. *Dev Biol* 113, 522-526.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.
- Rouget, C., Papin, C., Boureux, A., Meunier, A.-C., Franco, B., Robine, N., Lai, E.C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo. *Nature* 467, 1128-1132.
- Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086-1092.
- Sea Urchin Genome Sequencing, C., Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess, D.R., *et al.* (2006). The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314, 941-952.
- Semotok, J.L., Cooperstock, R.L., Pinder, B.D., Vari, H.K., Lipshitz, H.D., and Smibert, C.A. (2005). Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early *Drosophila* embryo. *Curr Biol* 15, 284-294.
- Seydoux, G., and Braun, R.E. (2006). Pathway to totipotency: lessons from germ cells. *Cell* 127, 891-904.
- Seydoux, G., and Dunn, M.A. (1997). Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of *Caenorhabditis elegans* and *Drosophila melanogaster*. *Development* 124, 2191-2201.
- Shirae-Kurabayashi, M., Matsuda, K., and Nakamura, A. (2011). Ci-Pem-1 localizes to the nucleus and represses somatic gene transcription in the germline of *Ciona intestinalis* embryos. *Development* 138, 2871-2881.
- Song, J.L., Stoeckius, M., Maaskola, J., Friedlander, M., Stepicheva, N., Juliano, C., Lebedeva, S., Thompson, W., Rajewsky, N., and Wessel, G.M. (2012). Select microRNAs are essential for early development in the sea urchin. *Dev Biol* 362, 104-113.
- Suzuki, A., Saba, R., Miyoshi, K., Morita, Y., and Saga, Y. (2012). Interaction between NANOS2 and the CCR4-NOT deadenylation complex is essential for male germ cell development in mouse. *PLoS One* 7.
- Takeda, Y., Mishima, Y., Fujiwara, T., Sakamoto, H., and Inoue, K. (2009). DAZL relieves miRNA-mediated repression of germline mRNAs by controlling poly(A) tail length in zebrafish. *PLoS One* 4.

- Tanaka, S., and Dan, K. (1990). Study of the lineage and cell cycle of small micromeres in embryos of the sea urchin, *Hemicentrotus pulcherrimus*. *Dev Growth Differ* 32, 145-156.
- Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A* 76, 4350-4354.
- Van Etten, J., Schagat, T.L., Hrit, J., Weidmann, C.A., Brumbaugh, J., Coon, J.J., and Goldstrohm, A.C. (2012). Human Pumilio proteins recruit multiple deadenylases to efficiently repress messenger RNAs. *J Biol Chem* 287, 36370-36383.
- Voronina, E., Lopez, M., Juliano, C.E., Gustafson, E., Song, J.L., Extavour, C., George, S., Oliveri, P., McClay, D., and Wessel, G. (2008a). 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.
- Voronina, E., Lopez, M., Juliano, C.E., Gustafson, E., Song, J.L., Extavour, C., George, S., Oliveri, P., McClay, D., and Wessel, G. (2008b). Vasa protein expression is restricted to the small micromeres of the sea urchin, but is inducible in other lineages early in development. *Developmental biology* 314, 276-286.
- Wei, Z., Angerer, R.C., and Angerer, L.M. (2006). A database of mRNA expression patterns for the sea urchin embryo. *Dev Biol* 300, 476-484.
- Weidmann, C.A., and Goldstrohm, A.C. (2012). *Drosophila* Pumilio protein contains multiple autonomous repression domains that regulate mRNAs independently of Nanos and brain tumor. *Mol Cell Biol* 32, 527-540.
- White, E.K., Moore-Jarrett, T., and Ruley, H.E. (2001). PUM2, a novel murine puf protein, and its consensus RNA-binding site. *RNA* 7, 1855-1866.
- Wreden, C., Verrotti, A.C., Schisa, J.A., Lieberfarb, M.E., and Strickland, S. (1997). Nanos and pumilio establish embryonic polarity in *Drosophila* by promoting posterior deadenylation of hunchback mRNA. *Development* 124, 3015-3023.
- Yajima, M., Gustafson, E.A., Song, J.L., and Wessel, G.M. (2013). Piwi regulates Vasa accumulation during embryogenesis in the sea urchin. *Dev Dyn*.
- Yajima, M., and Wessel, G.M. (2011). The DEAD-box RNA helicase Vasa functions in embryonic mitotic progression in the sea urchin. *Development* 138, 2217-2222.
- Yajima, M., and Wessel, G.M. (2012). Autonomy in specification of primordial germ cells and their passive translocation in the sea urchin. *Development* 139, 3786-3794.
- Zaessinger, S., Busseau, I., and Simonelig, M. (2006). Oskar allows nanos mRNA translation in *Drosophila* embryos by preventing its deadenylation by Smaug/CCR4. *Development* 133, 4573-4583.

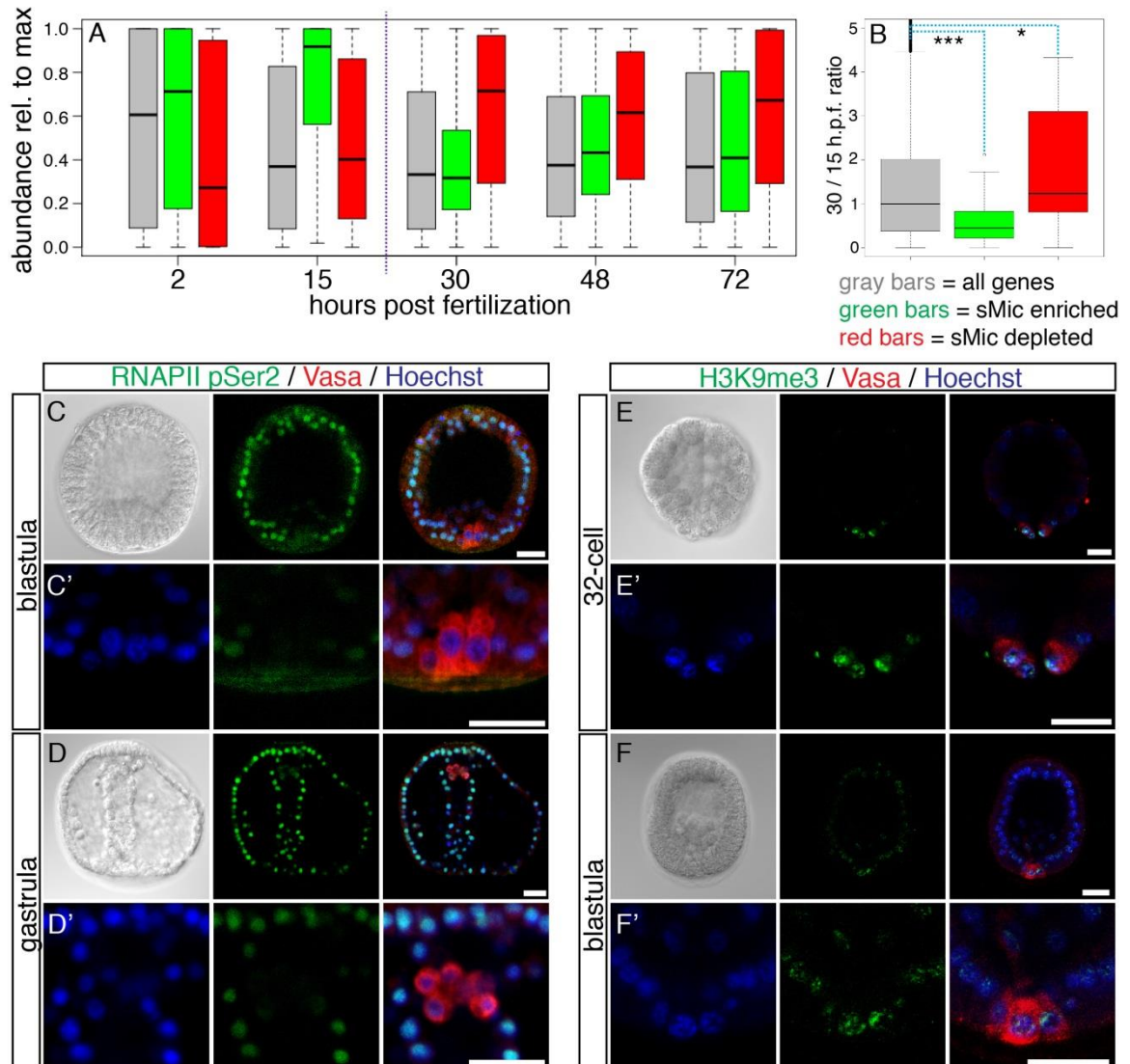
FIGURES

Figure 1. FACS isolation and deep sequencing of sMics.



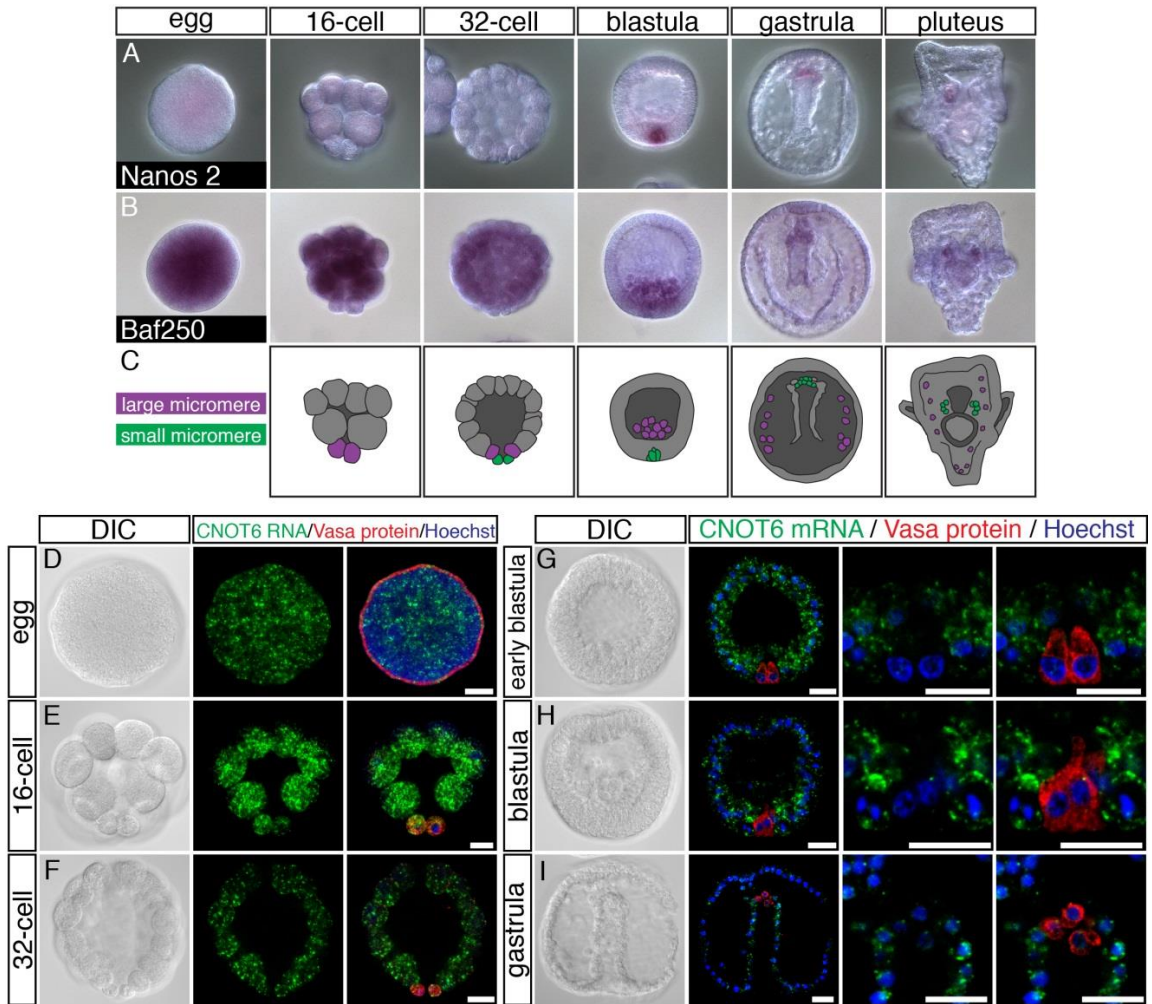
(A) Scheme diagram depicting FACS isolation, RNA deep sequencing, and differential expression analysis. (B) sMics accumulate the fluorescent dye calcein at 15 h.p.f. sMics are co-labeled with an mCherry fusion reporter of Vasa, a conserved germ line RNA helicase that localizes post-translationally. (C) Calcein labeling is retained in blastomeres dissociated at 15 h.p.f. Arrows indicate labeled sMic. (D) Representative scatter plot of sMics collected by FACS, plotted as forward scatter vs. calcein fluorescence. Cells collected as putative sMics are indicated by the box, which represents 0.5% of the total cell sample. (E) Venn diagram summarizing the number of sMic enriched and depleted transcripts discovered in three differential expression comparisons with FDR < 0.05: sMic vs. whole embryo (WE) and sMic vs. non-sMics, with and without replicate 1 included. Transcript totals are given as sMic Enriched / sMic depleted.

Figure 2. Transcriptional repression in sMics.



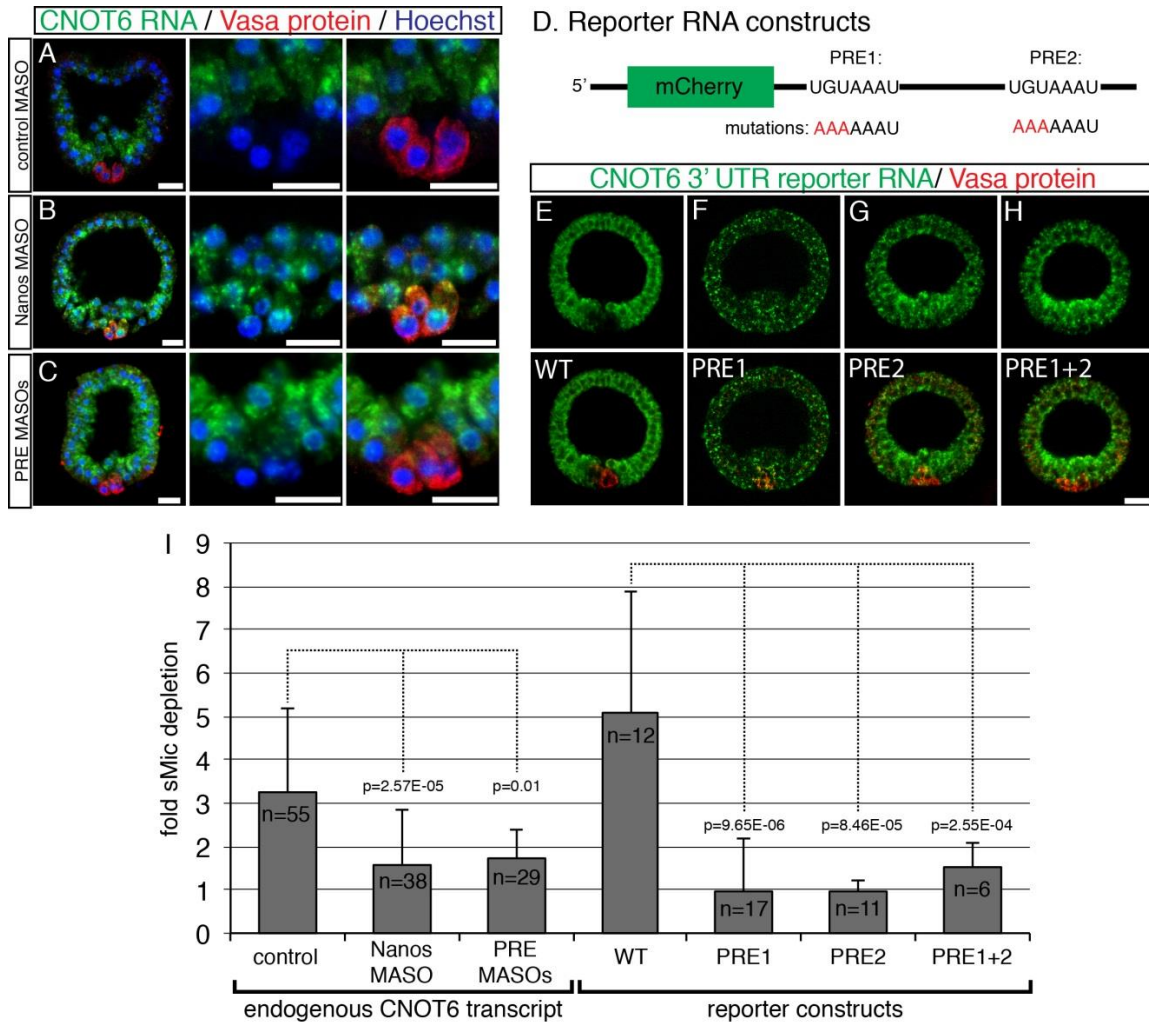
(A) Box and whisker plots summarizing average relative abundance of sMic-enriched, sMic-depleted, and all transcripts in the microarray study at five developmental time points from 2 to 72 h.p.f. sMic-enriched transcripts (green bars) on average reach their maximum abundance before 30hpf, and then decrease. Conversely, sMic-depleted transcripts (red bars) reach their maximum after 30 h.p.f. An average of all genes in the microarray study remains relatively constant through development (gray bars). (B) Ratio of average expression levels from 15 to 30 h.p.f. for sMic-enriched, depleted, and all transcripts. *** $p=1.076e-06$, * $p=0.001357$ (Mann-Whitney-U). (C-D) Immunofluorescence signal for RNAPII pSer2 (green) is depleted in sMics of blastula (C) and gastrulae (D). (E-F) H3K9me3 (green) immunofluorescence is enriched in sMics in 32-cell embryos (C) and blastulae (D). (C'-F') Zoom views of the sMics, labeled with vasa antibody (red) (Voronina *et al.*, 2008). Scale bars = 20 μ m.

Figure 3. Localization of select differentially expressed transcripts.



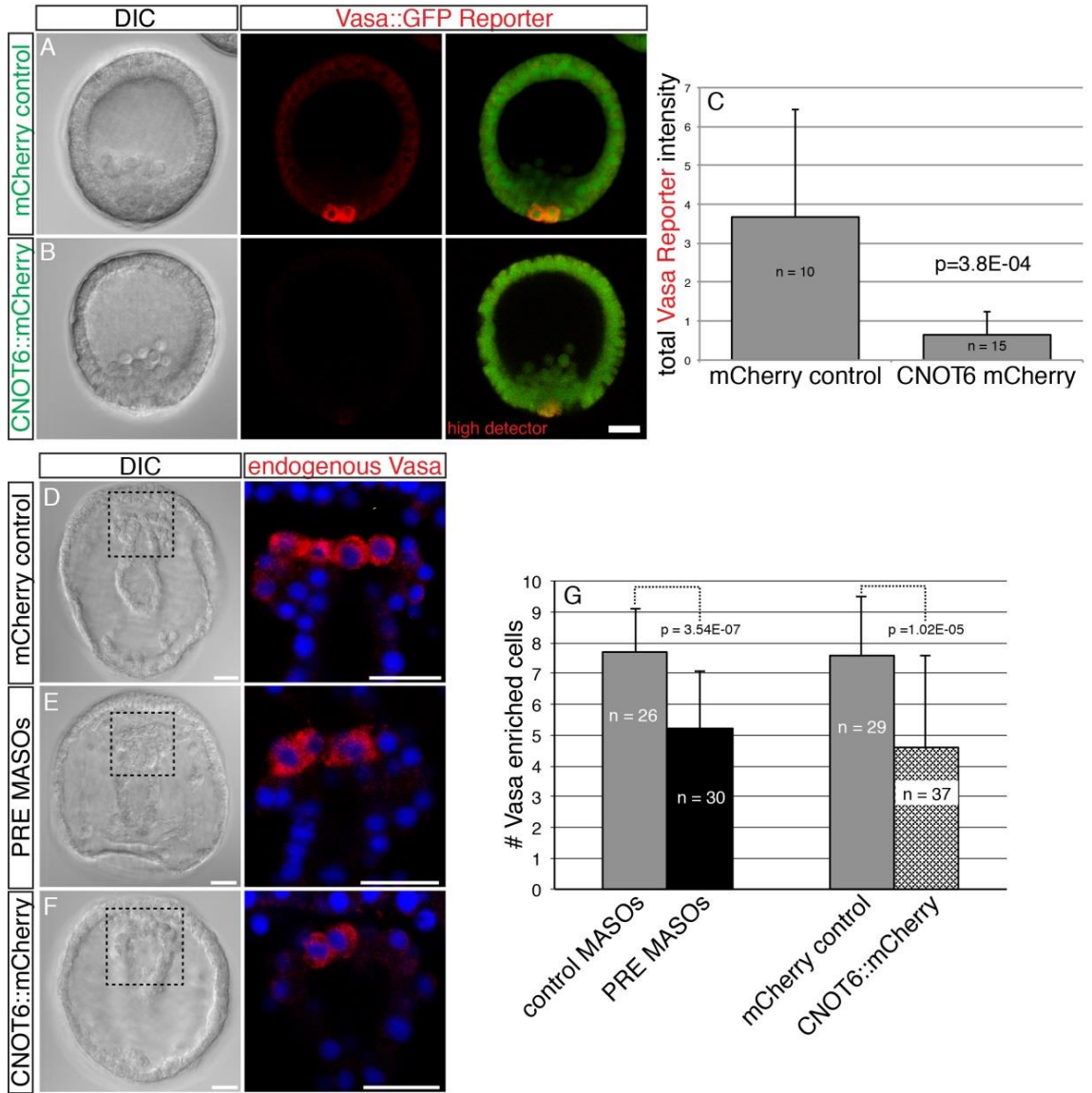
(A,B) In situ hybridizations for sMic-enriched transcripts Nanos2 and Baf250. In contrast to Nanos, Baf250 transcript is detectable in eggs and is broadly distributed until blastula stage when its localization is refined to the sMics. (C) Schematic depiction of the micromere lineage in early development. (D-I) Fluorescent in situ hybridization for CNOT6 transcript (green). CNOT6 transcript is depleted in the sMics (labeled by Vasa antibody, red) by blastula stage and remains depleted through gastrulation, after which it is undetectable. Scale bars = 20µm.

Figure 4. CNOT6 mRNA is depleted in the sMics by Nanos.



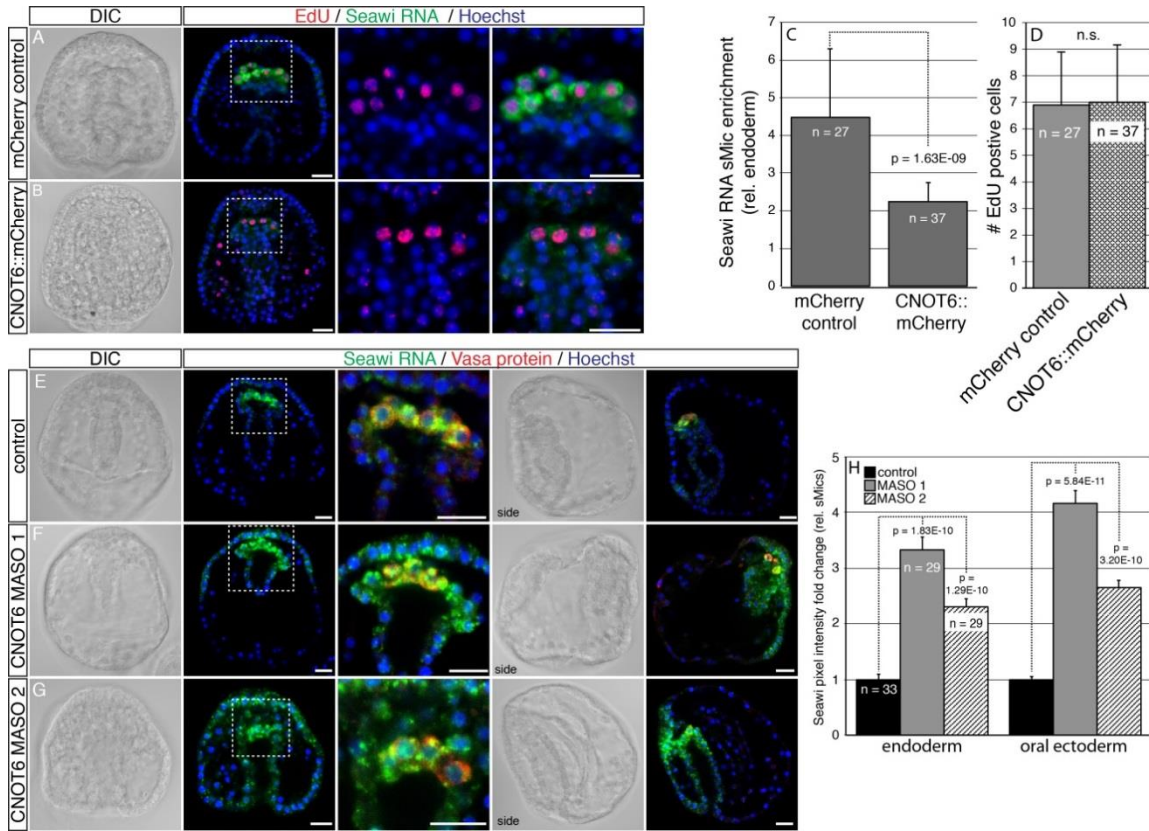
(A,B) MASO knockdown of Nanos derepresses CNOT6 transcript in the sMics compared to control embryos. (C) Co-injection of MASOs targeting the two PREs also derepresses CNOT6 transcript in sMics. (D-H) A synthetic reporter construct (D) containing the full-length CNOT6 3'UTR recapitulates the localization of endogenous CNOT6 transcript (E). Mutating putative PRE sequences singly (F,G) or in combination (H) derepresses reporter accumulation in the sMics. (I) Pixel intensity quantitation of fold depletion in sMics versus the rest of the embryo for endogenous CNOT6 and reporter transcripts. Scale bars = 20 μ m.

Figure 5. CNOT6 depletion is required for sMic Vasa protein expression.



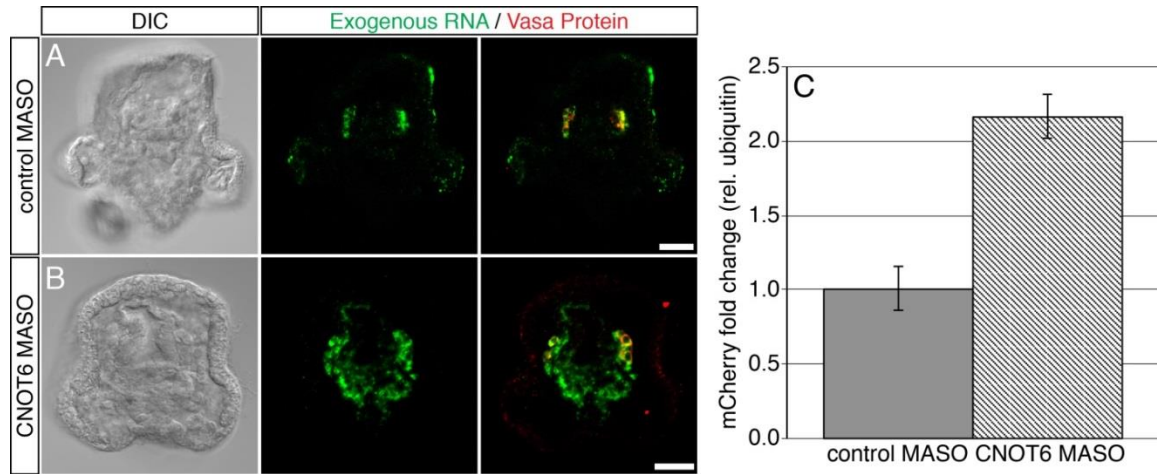
(A-B) Expression of CNOT6::mCherry (green) in all cells significantly reduces Vasa::GFP reporter signal (red) compared to mCherry alone control in 20 h.p.f. embryos. An image at higher detector settings is provided, indicating the sMics are still present and weakly positive for Vasa::GFP. (C) Integrated pixel intensity data for mCherry control and CNOT6::mCherry misexpression embryos. (D) PRE-protecting MASOs result in significantly fewer Vasa positive cells at 42 h.p.f. (E) Expression of CNOT6::mCherry in all cells, including the sMics, similarly reduces the number of Vasa positive cells compared to expression of mCherry alone. (F) Counts of Vasa enriched cells in control, and PRE-protecting MASO, and CNOT6 misexpression embryos. p-values are by 2-way unpaired t-test.

Figure 6. CNOT6 mediates selective enrichment of Seawi transcript in the sMics.



(A,B) Expression of CNOT6::mCherry in all cells reduces the enrichment of Seawi transcript in the sMics, detected by FISH (green), compared to mCherry alone controls. sMic lineage is traced by EdU incorporation (red). (C) Pixel intensity quantitation of Seawi transcript enrichment in sMics relative to the endoderm. (D) Counts of EdU positive cells in mCherry control and CNOT6::mCherry expressing embryos indicate no change in sMic numbers. (E-G) CNOT6 knockdown with either of two non-overlapping MASOs expands the domain of Seawi RNA into the endoderm and oral ectoderm at 42 h.p.f. (H) Fold change of Seawi FISH average pixel intensity in control and knockdown embryos, relative to sMic Seawi intensity. All p-values by unpaired two-tailed T-test. Scale bars = 20µm.

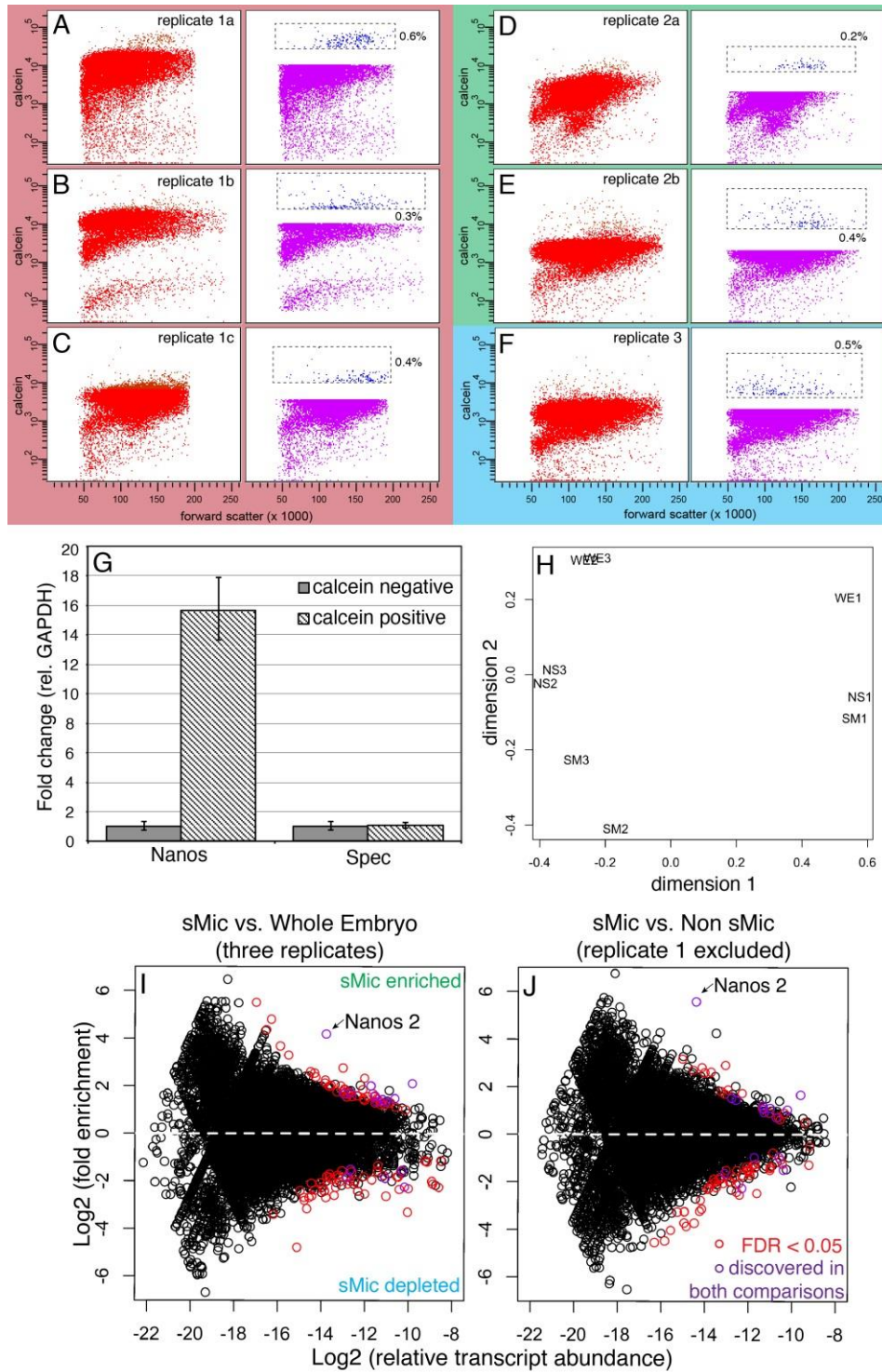
Figure 7. CNOT6 regulates general retention of exogenous RNA in the sMics.



(A,B) FISH for injected RNA containing the mCherry open reading frame and SV40 3' polyadenylation signal. This transcript is normally retained only in sMics at 96 h.p.f (A), but is retained broadly throughout the endomesoderm with CNOT6 knockdown (B). (C) qPCR for exogenous mCherry RNA at 96 h.p.f. With CNOT6 knockdown, mCherry levels increase by 2-fold. Scale bars = 20 μ m.

SUPPLEMENTAL INFORMATION

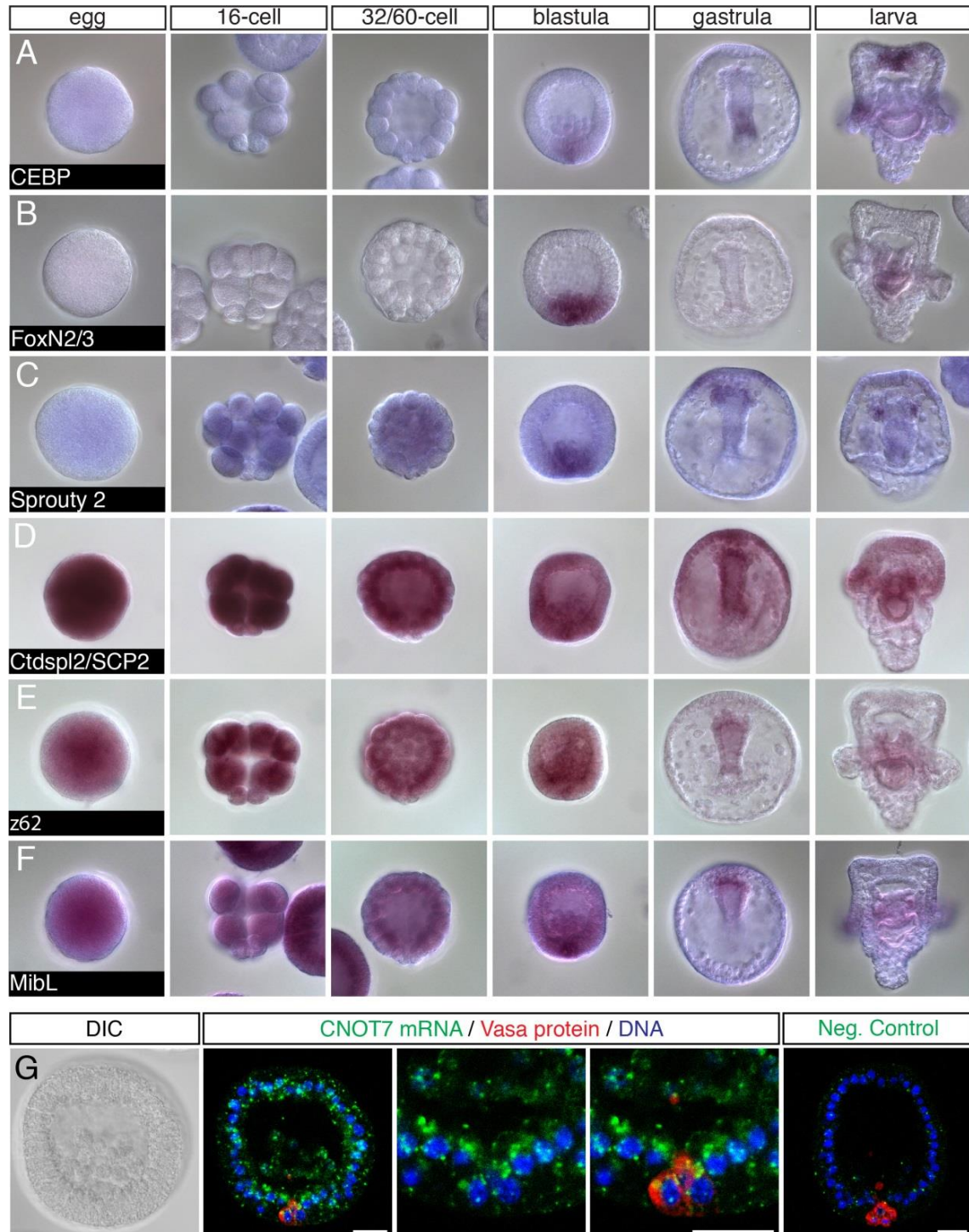
Supplemental Figure 1: FACS isolation of sMics and transcriptomic analysis



(A-F) Scatter plot representations of FACS sorts for three biological replicates. Replicate 1 (A-C, red background) consisted of three sorts of embryos derived from distinct parental crosses.

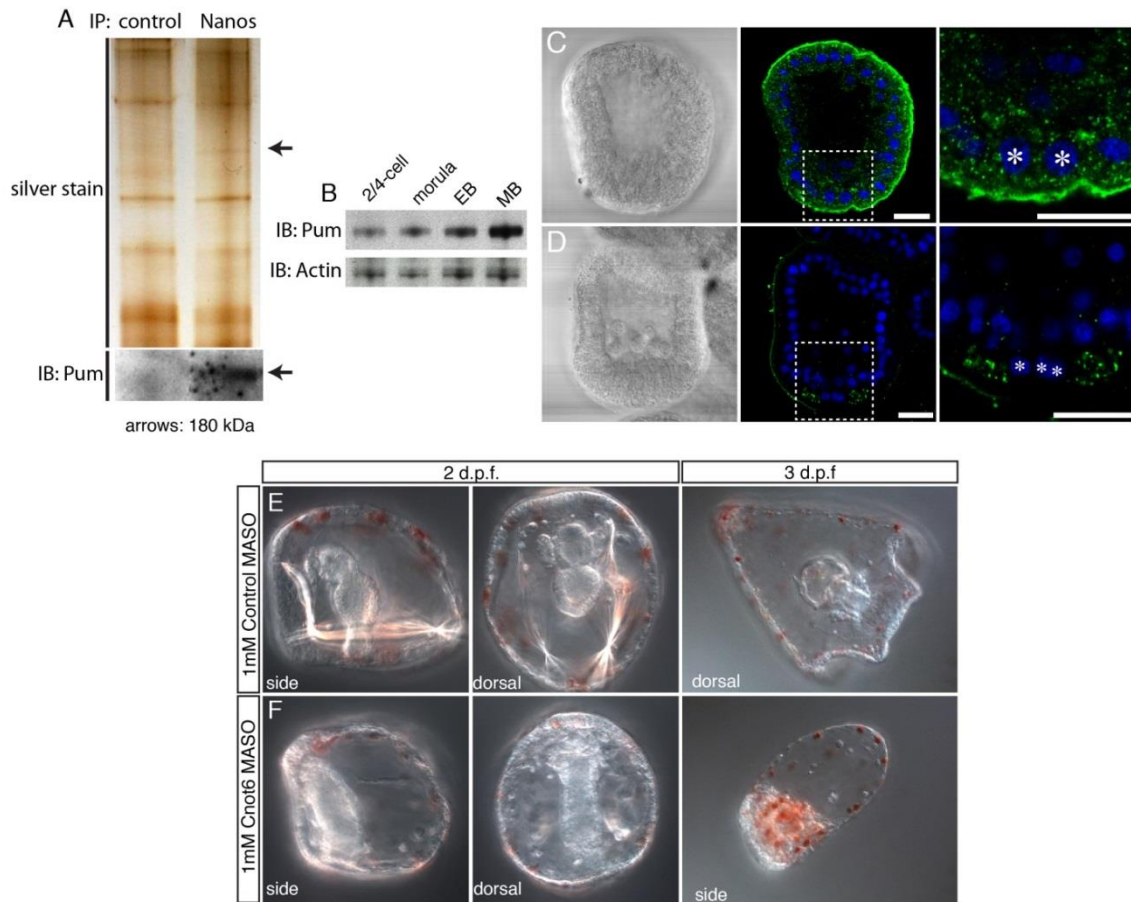
Replicate 2 (D-E, green background) consisted of two sorts from two parental crosses. Replicate 3 (F, blue background) was sorted from a single parental cross. Red points indicate the total cell population, blue points indicate the collected calcein⁺ cells, and purple points indicate collected calcein⁻ cells. Percentages represent fraction of collected calcein⁺ cells relative to the whole population. (G) qPCR for Nanos indicates enrichment of sMics in the calcein⁺ fraction, as opposed for Spec, an ectodermal negative control. (H) Multidimensional scaling (MDS) analysis of replicate transcriptomes for isolated sMics (SM1-3), non-sMics (NS1-3), and whole embryo (WE1-3). Replicate 2 and 3 samples were highly related with whole embryo, sMic, and non-sMic transcriptomes clustering separately. Replicate 1 samples are outliers, showing low relatedness to the others in both dimensions, and poor separation between SM and NS samples. This variation likely points to issues with sample handling (RNA collection, or RNA-seq sample preparation) or less efficient calcein labeling. In subsequent differential expression analysis, we therefore made comparisons both with and without including Replicate 1 samples. (I,J) Smear plot depiction of differentially expressed genes. (I) sMic vs. Whole Embryo with replicate 1 included, and (J) sMic vs. non-sMic with replicate 1 excluded are considered here. Transcripts are represented by open circles, with fold enrichment in the sMics on the y-axis and relative transcript abundance on the x-axis. Transcripts meeting a significance threshold of 0.05 (false discovery rate, FDR) are in red; transcripts identified in both comparisons are in purple. Arrows indicate the previously identified sMic factor Nanos2.

Supplemental Figure 2: WMISH for selected transcripts.



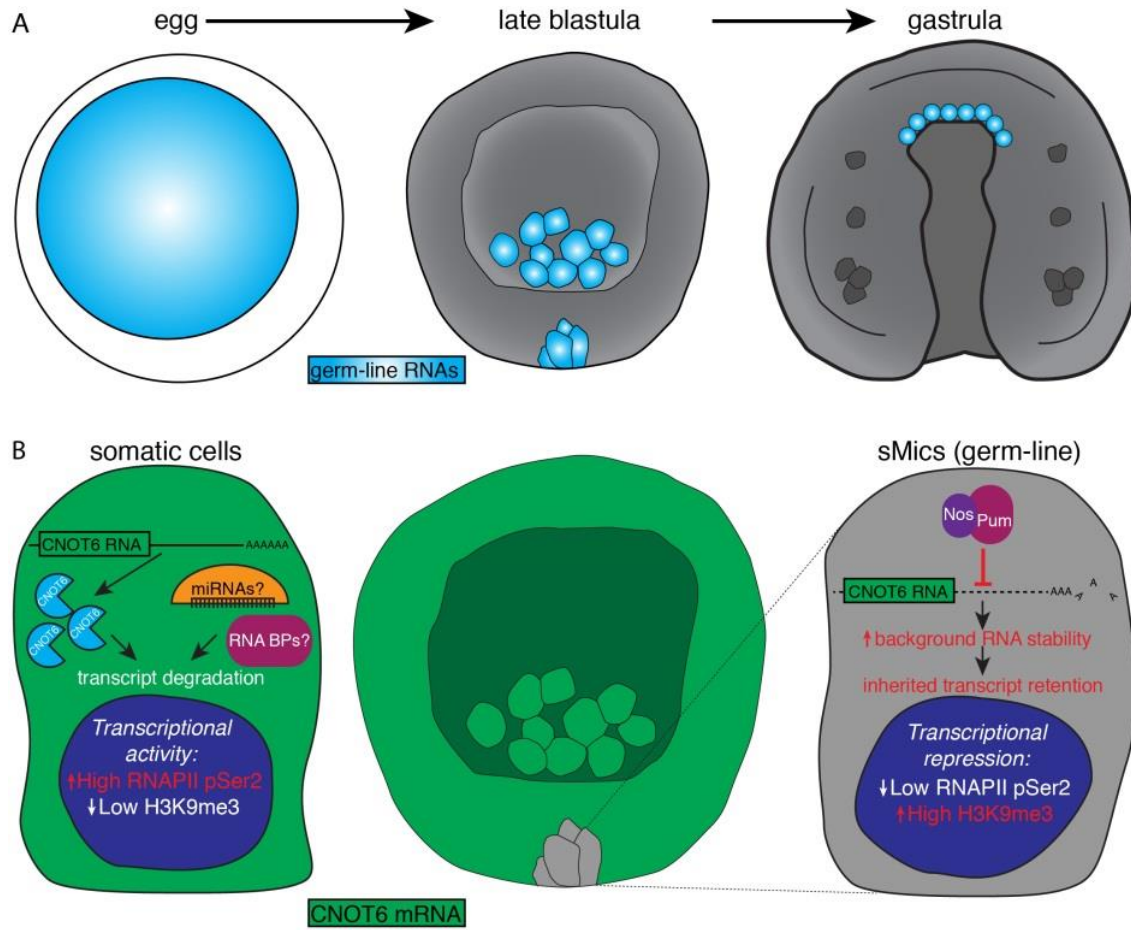
(A-F) WMISH for sMic-enriched transcripts identified through differential expression analysis. (G) The *S. purpuratus* genome contains two CNOT-related nucleases: CNOT6 and CNOT7. Unlike CNOT6, which is specifically depleted in the sMics, CNOT7 transcript (green) is ubiquitously distributed. sMics are labeled in red with vasa antibody.

Supplemental Figure 3: Interaction between Nanos and Pumilio and CNOT6 knockdown.



(A) Association of Nanos and Pumilio. Compared to Vasa antibody control, Nanos antibody specifically pulls down a 180 kDa band recognized by silver stain and Pumilio immunoblot. (B) Temporal expression of Pumilio. Pumilio is detectable in early embryos, and increases in abundance through early blastulae (EB) and mesenchyme blastulae (MB) stages. (C,D) Spatial localization of Pumilio protein. Pumilio is detectable in granule structures in all cells, including the sMics, of early blastula (C), but becomes highly enriched in the Veg2 mesodermal precursors after primary mesenchyme ingression (D). Asterisks indicate sMic nuclei. (E,F) Injection of 1mM CNOT6 MASO 1 results in failure to produce skeletons or a tripartite gut. By 3 days, the endoderm degenerates into mesenchyme-like cells.

Supplemental Figure 4: Time capsule model for germ line development.



(A) Maternally deposited mRNAs are coordinately degraded in somatic cells at blastula/gastrula stage but stabilized in the sMics. (B) Mechanism for sMic segregation. CNOT6 transcript is in green, and is present in all cells except sMics (gray), where Nanos/Pumilio (Nos/Pum) represses it. This repression creates a stable environment for inherited RNA. Conversely, CNOT6 protein accumulates in somatic cells, where it may act in parallel with and/or downstream of miRNA and RNA binding protein (BP) mediated degradation.

Supplemental Table 1: Gene set enrichment analysis of the union sets of 78 sMic enriched transcripts, and 152 sMic-depleted transcripts.

| sMic-enriched categories | | | sMic-depleted categories | | |
|--------------------------|---|----------|--------------------------|---|----------|
| GO ID | Term | p-val | GO ID | Term | p-val |
| 3676 | nucleic acid binding | 4.40E-05 | 5198 | structural molecule activity | 4.50E-08 |
| 988 | protein binding transcription factor act... | 0.0015 | 32561 | guanyl ribonucleotide binding | 0.0042 |
| 989 | transcription factor binding transcripti... | 0.0015 | 19001 | guanyl nucleotide binding | 0.0045 |
| 3712 | transcription cofactor activity | 0.0015 | 5201 | extracellular matrix structural constitu... | 0.0053 |
| 43565 | sequence-specific DNA binding | 0.0049 | 15399 | primary active transmembrane transporter... | 0.01 |
| 8134 | transcription factor binding | 0.006 | 15405 | P-P-bond-hydrolysis-driven transmembrane... | 0.01 |
| 8349 | MAP kinase kinase kinase kinase activity | 0.0067 | 4683 | calmodulin-dependent protein kinase acti... | 0.016 |
| 43425 | bHLH transcription factor binding | 0.0067 | 4197 | cysteine-type endopeptidase activity | 0.0212 |
| 43426 | MRF binding | 0.0067 | 22804 | active transmembrane transporter activit... | 0.0228 |
| 5488 | binding | 0.0086 | 4674 | protein serine/threonine kinase activity | 0.0237 |
| 35258 | steroid hormone receptor binding | 0.0178 | 3723 | RNA binding | 0.0287 |
| 3729 | mRNA binding | 0.0189 | 16289 | CoA hydrolase activity | 0.0368 |
| 35257 | nuclear hormone receptor binding | 0.0224 | 50839 | cell adhesion molecule binding | 0.0368 |
| 51427 | hormone receptor binding | 0.0275 | | | |
| 90079 | translation regulator activity, nucleic ... | 0.0328 | | | |
| 3677 | DNA binding | 0.0389 | | | |
| 42974 | retinoic acid receptor binding | 0.0393 | | | |
| 45182 | translation regulator activity | 0.0393 | | | |
| 8092 | cytoskeletal protein binding | 0.0433 | | | |
| 32561 | guanyl ribonucleotide binding | 0.0446 | | | |
| 46332 | SMAD binding | 0.0457 | | | |
| 19001 | guanyl nucleotide binding | 0.0467 | | | |

Analysis was performed using the topGO Bioconductor package. P-values are by Fisher's exact test.

Supplemental Table 2: Yields of cell sorting runs.

| | sMic | non-sMic | whole embryo |
|---|-------------------------------|-------------------------------|---------------------|
| Replicate 1 (3 pooled parental crosses) | 700 ng RNA 113,191 cells | 1,800 ng RNA 618,511 cells | 3,800 ng RNA |
| Replicate 2 (2 pooled parental crosses) | 1,080 ng RNA 213,174 cells | 1,820 ng RNA 892,329 cells | 6,700 ng RNA |
| Replicate 3 (1 parental cross) | 350 ng RNA 55,571 cells | 1,278 ng RNA 438,938 cells | 1,508 ng RNA |

Supplemental Table 3: Run statistics for Helicos sequencing runs.

| Sample Name | Total Filtered Reads | Mean Read Length (nt) | Percent aligned reads | Sequencing Error Rate |
|--------------------|-----------------------------|------------------------------|------------------------------|------------------------------|
| WE 1 | 27,035,746 | 31.21 | 49.08% | 6.95% |
| sMic 1 | 25,331,121 | 31.15 | 44.29% | 6.86% |
| non-sMic 1 | 30,889,594 | 31.27 | 42.25% | 7.02% |
| WE 2 | 14,670,195 | 33.75 | 60.12% | 5.44% |
| sMic 2 | 15,096,355 | 33.28 | 60.84% | 5.49% |
| non-sMic 2 | 12,043,418 | 34.24 | 57.30% | 5.63% |
| WE 3 | 14,450,706 | 33.84 | 54.68% | 5.89% |
| sMic 3 | 8,300,956 | 33.66 | 56.06% | 5.62% |
| non-sMic 3 | 10,630,234 | 34.15 | 58.47% | 5.69% |

Supplemental Table 4: Transcripts that are differentially enriched in sMics compared to non-sMics (replicate 1 excluded).

| Oases TransID | Oases GOname | SPU identifier | SPU GOname | Helicos LogConc | Helicos FC | Helicos FDR | Microarray Annotation | 2hpf | 15hpf | 30hpf | 48hpf | 72hpf |
|---------------|--|----------------|--|-----------------|------------|-------------|----------------------------------|-------|-------|-------|-------|-------|
| 1733T41 | creb-binding protein | SPU_019024 | creb binding protein | -10.206 | 1.033 | 2.04E-10 | Sp-CBP | 21445 | 13363 | 4065 | 7377 | 8550 |
| 3161T7 | ccat enhancer binding protein (c ebp)gamma | SPU_001657 | ccat enhancer binding protein (c ebp)gamma | -13.047 | 2.9084 | 2.35E-08 | Sp-Cebpa | 114 | 400 | 481 | 2445 | 8431 |
| 3106T4 | nanos | SPU_003591 | nanos | -14.394 | 5.5768 | 5.37E-05 | Sp-nanos2 | 4 | 1188 | 247 | 113 | 50 |
| 3955T11 | Na | SPU_018326 | centrosome-associated protein 350 | -11.338 | 1.2012 | 0.00010711 | Sp-CLIP170 putative homolog | 573 | 1445 | 503 | 579 | 350 |
| 34941T2 | Na | None | Na | -31.806 | 36.421 | 0.00039785 | Na | Na | Na | Na | Na | Na |
| 2557T6 | sry | SPU_004217 | sry | -13.708 | 2.6199 | 0.00045609 | Sp-SoxD1 | 1807 | 1017 | 675 | 806 | 1107 |
| 7602T5 | Na | SPU_022170 | histone h3 | -10.349 | 0.88947 | 0.00053661 | Sp-early-histone-H3 | 5841 | 8098 | 5682 | 2561 | 1135 |
| 6476T24 | protein | SPU_003509 | protein | -12.28 | 1.7277 | 0.0007431 | Sp-Ncor2 | 435 | 402 | 319 | 145 | 135 |
| 1347T18 | protein | SPU_027389 | laminin a | -9.3954 | 0.51088 | 0.0011283 | Sp-laminin alpha 5-like fragment | 676 | 1261 | 398 | 976 | 487 |
| 1785T14 | novel protein | SPU_023530 | novel protein | -10.991 | 1.1034 | 0.0013652 | Sp-swi-like | 8976 | 3764 | 985 | 2084 | 3534 |
| 1047T39 | Na | SPU_009086 | protein | -10.625 | 0.73416 | 0.0014856 | Sp-QKI | 15673 | 32994 | 22304 | 18903 | 19123 |
| 3461T18 | ubiquitin-conjugating enzyme | SPU_026270 | ubiquitin-conjugating enzyme | -14.121 | 2.8141 | 0.0017235 | Sp-Ube2j2 | 4552 | 2347 | 717 | 458 | 617 |
| 1359T35 | histone h2b | SPU_001312 | histone h2b | -14.001 | 2.5942 | 0.0030715 | Sp-cleavage histone H2b | 6634 | 6003 | 599 | 284 | 177 |
| 853T4 | g protein alpha subunit | SPU_003898 | g protein alpha subunit | -10.715 | 0.71243 | 0.0041543 | Sp-Gq | 3082 | 173 | 177 | 100 | 156 |
| 187T32 | Na | SPU_021973 | Na | -11.352 | 1.0974 | 0.0052696 | | 0 | 35 | 0 | 0 | 0 |
| 48999T1 | Na | SPU_026204 | Na | -31.985 | 36.062 | 0.0052696 | | 0 | 42 | 4 | 0 | 0 |
| 29309T1 | Na | None | Na | -11.03 | 0.90928 | 0.0057965 | Na | Na | Na | Na | Na | Na |
| 13945T8 | Na | SPU_019217 | Na | -12.557 | 1.4463 | 0.0065427 | | 45 | 2455 | 4244 | 4112 | 2655 |
| 3402T18 | saps domainmember 2 | SPU_016394 | saps domainmember 2 | -11.291 | 1.0746 | 0.0070351 | | 3427 | 3142 | 1829 | 3012 | 2770 |

| | | | | | | | | | | | | |
|---------|------------------------------|------------|------------------------------------|---------|---------|-----------|----------------------|-------|-------|-------|-------|-------|
| 848T27 | cadherin | SPU_010840 | protein | -11.269 | 0.93239 | 0.0094185 | Sp-G-cadherin-like1 | 77 | 518 | 94 | 285 | 224 |
| 539T9 | novel protein | SPU_011640 | novel protein | -12.929 | 1.8694 | 0.012626 | Sp-Arid5b | 1102 | 1127 | 646 | 1194 | 749 |
| 68957T1 | Na | None | Na | -10.536 | 0.62006 | 0.014336 | Na | Na | Na | Na | Na | Na |
| 192T428 | Na | SPU_015457 | Na | -9.598 | 1.6591 | 0.01659 | | 7051 | 50736 | 51385 | 51608 | 49517 |
| 5557T1 | Na | SPU_018356 | Na | -32.11 | 35.813 | 0.016608 | Sp-cub/hyalin-like3 | 70 | 276 | 2406 | 1933 | 652 |
| 1052T6 | chimerin1 | SPU_005298 | chimerin1 | -12.765 | 1.5152 | 0.02001 | Sp-RacGAP1 | 3948 | 3238 | 944 | 956 | 1071 |
| 9387T6 | Na | SPU_013698 | Na | -14.52 | 2.696 | 0.023081 | | 0 | 23 | 204 | 914 | 1189 |
| 3725T7 | Na | SPU_023854 | large homolog-associated protein 4 | -13.245 | 1.8072 | 0.03318 | | 14977 | 7600 | 1504 | 1519 | 1801 |
| 5006T6 | protein | SPU_007379 | protein | -15.019 | 3.1651 | 0.038055 | Sp-Ddb1_1 | 1968 | 1321 | 142 | 1163 | 883 |
| 3260T3 | egg bindin receptor 1partial | SPU_017620 | egg bindin receptor 1partial | -13.074 | 1.7014 | 0.042748 | Sp-EGF/hyalin-like10 | 0 | 655 | 93 | 74 | 60 |
| 14414T1 | meis homeobox 2 | SPU_023739 | meis homeobox 2 | -32.21 | 35.612 | 0.048433 | Sp-Pbx | 5276 | 2927 | 2007 | 3462 | 2551 |

OasesTransID: Transcript identifier from Oases de novo transcriptome.

OasesGOname: Name assigned to Oases transcript by Blast2GO.

SPU Identifier: Top blast hit of Oases transcript to SPU gene predictions.

SPUGOname: Name assigned to SPU transcript by Blast2GO.

HelicosLogConc: Transcript abundance calculated by EdgeR.

HelicosFC: Fold change enrichment of the transcript; positive values are sMic enriched, negative values are sMic depleted.

HelicosFDR: Statistical significance (false discovery rate) for differentially expressed transcript.

Microarray Annotation: Annotation given in temporal gene expression database.

2hpf-72hpf: Relative transcript abundance by temporal microarray data.

Supplemental Table 5: Transcripts that are differentially depleted in sMics compared to non-sMics (replicate 1 excluded).

| Oases TransID | Oases GName | SPU identifier | SPU GName | Helicos LogConc | Helicos FC | Helicos FDR | Microarray Annotation | 2hpf | 15hpf | 30hpf | 48hpf | 72hpf |
|---------------|---------------------|----------------|---------------------|-----------------|------------|-------------|-----------------------------|------|-------|-------|-------|-------|
| 62424T1 | Na | None | Na | -11.003 | -1.4974 | 6.09E-10 | Na | Na | Na | Na | Na | Na |
| 10589T2 | Na | None | Na | -10.997 | -1.4096 | 1.28E-08 | Na | Na | Na | Na | Na | Na |
| 10791T1 | Na | None | Na | -12.532 | -2.4235 | 1.23E-07 | Na | Na | Na | Na | Na | Na |
| 72017T1 | Na | None | Na | -12.405 | -2.5464 | 2.32E-07 | Na | Na | Na | Na | Na | Na |
| 30688T1 | Na | None | Na | -12.268 | -1.7188 | 2.87E-05 | Na | Na | Na | Na | Na | Na |
| 5257T7 | Na | None | Na | -14.13 | -3.1432 | 6.82E-05 | Na | Na | Na | Na | Na | Na |
| 45391T1 | Na | None | Na | -14.472 | -3.7312 | 0.00012088 | Na | Na | Na | Na | Na | Na |
| 15885T8 | Na | None | Na | -10.435 | -1.2533 | 0.00016613 | Na | Na | Na | Na | Na | Na |
| 9T29 | beta-tubulin | SPU_003894 | beta-tubulin | -10.579 | -0.97541 | 0.00016613 | Sp-beta-tubulin-5 | 0 | 0 | 19 | 112 | 106 |
| 16434T4 | Na | None | Na | -12.036 | -1.4949 | 0.00018003 | Na | Na | Na | Na | Na | Na |
| 20238T1 | Na | None | Na | -14.84 | -3.8709 | 0.00021686 | Na | Na | Na | Na | Na | Na |
| 22818T1 | Na | SPU_000646 | Na | -9.1771 | -0.50125 | 0.0003601 | Sp-SRCR-4 | 0 | 0 | 0 | 0 | 0 |
| 78153T1 | Na | None | Na | -14.226 | -3.2329 | 0.00045609 | Na | Na | Na | Na | Na | Na |
| 2994T3 | Na | SPU_021385 | Na | -12.29 | -2.2667 | 0.00045609 | | 0 | 2155 | 1820 | 1674 | 1446 |
| 4495T7 | Na | SPU_007513 | Na | -11.912 | -1.7192 | 0.00052755 | | 630 | 229 | 450 | 348 | 423 |
| 11678T4 | Na | None | Na | -11.946 | -1.5863 | 0.00057657 | Na | Na | Na | Na | Na | Na |
| 52423T1 | Na | SPU_002844 | Na | -12.819 | -1.9799 | 0.00057782 | | 0 | 0 | 0 | 0 | 0 |
| 150T4 | Na | None | Na | -9.2151 | -1.279 | 0.00081858 | Na | Na | Na | Na | Na | Na |
| 30603T2 | chaperone protein | SPU_018707 | chaperone protein | -12.063 | -1.362 | 0.0015241 | | 0 | 0 | 215 | 54 | 55 |
| 7491T5 | polyketide synthase | SPU_028395 | polyketide synthase | -11.682 | -1.4083 | 0.0028286 | Sp-Polyketide-synthase-like | 0 | 414 | 54 | 0 | 0 |
| 14179T1 | protein | SPU_014053 | protein | -15.556 | -4.0331 | 0.0039548 | | 2241 | 2368 | 1165 | 1783 | 2213 |
| 17872T1 | Na | None | Na | -14.931 | -3.2587 | 0.0039548 | Na | Na | Na | Na | Na | Na |
| 52812T1 | Na | None | Na | -14.921 | -3.2471 | 0.0052696 | Na | Na | Na | Na | Na | Na |
| 632T1 | Na | None | Na | -11.6 | -2.477 | 0.0052696 | Na | Na | Na | Na | Na | Na |

| | | | | | | | | | | | | |
|--------------|--|------------|--|---------|----------|-----------|-------------------------------------|------|------|------|-------|-------|
| 17506T1 | Na | None | Na | -13.79 | -2.2529 | 0.0052696 | Na | Na | Na | Na | Na | Na |
| 32764T1 | Na | None | Na | -13.145 | -2.058 | 0.0065427 | Na | Na | Na | Na | Na | Na |
| 36674T1 | Na | None | Na | -13.695 | -2.1051 | 0.0093462 | Na | Na | Na | Na | Na | Na |
| 3599T6 | bacterial ig-like domain protein | SPU_000439 | bacterial ig-like domain protein | -13.523 | -1.9966 | 0.0094185 | | 0 | 1927 | 7595 | 6399 | 6062 |
| 27434T1 | Na | None | Na | -13.246 | -1.8712 | 0.010277 | Na | Na | Na | Na | Na | Na |
| 70215T1 | Na | None | Na | -15.829 | -4.4808 | 0.011622 | Na | Na | Na | Na | Na | Na |
| 631T3 | typealpha 1 | SPU_003768 | Na | -11.789 | -1.2997 | 0.011622 | Sp-3 alpha procollagen; collagen IV | 3 | 579 | 1047 | 929 | 882 |
| 498T10 | novel krab box and zincc2h2 type domain containing protein | SPU_004148 | novel krab box and zincc2h2 type domain containing protein | -13.083 | -1.6444 | 0.012286 | Sp-krl | 0 | 1449 | 288 | 139 | 21 |
| 11467T2 | Na | SPU_004557 | Na | -12.843 | -2.0027 | 0.012626 | Sp-Sarm-r11 | 3481 | 1818 | 599 | 697 | 866 |
| 67248T1 | Na | None | Na | -14.184 | -2.7711 | 0.012628 | Na | Na | Na | Na | Na | Na |
| 11987T1 | alpha tubulin | SPU_024615 | alpha-tubulin | -10.385 | -1.4938 | 0.01389 | Sp-alpha-tubulin-6 | 3545 | 9150 | 9318 | 10749 | 11113 |
| 56498T1 | Na | None | Na | -12.639 | -1.8298 | 0.014336 | Na | Na | Na | Na | Na | Na |
| 20463T5 | Na | SPU_000512 | nadph oxidase | -13.331 | -1.8627 | 0.014721 | Sp-urchin dual oxidase 2 | 0 | 0 | 0 | 0 | 0 |
| 37817T1 | Na | None | Na | -14.939 | -3.7227 | 0.017104 | Na | Na | Na | Na | Na | Na |
| 16759T1 | Na | None | Na | -13.854 | -2.3038 | 0.018869 | Na | Na | Na | Na | Na | Na |
| 8613T1 | wd-40 repeat protein | SPU_028036 | protein | -12.528 | -1.5251 | 0.021366 | Sp-RasGRF1 | 1431 | 1493 | 1816 | 1205 | 1334 |
| 4718T2 | Na | None | Na | -11.721 | -0.97429 | 0.024226 | Na | Na | Na | Na | Na | Na |
| 70995T1 | Na | None | Na | -16.296 | -4.549 | 0.024319 | Na | Na | Na | Na | Na | Na |
| 23700T1 | Na | None | Na | -10.963 | -1.3578 | 0.024319 | Na | Na | Na | Na | Na | Na |
| 41842T1 | Na | None | Na | -12.009 | -1.5425 | 0.031411 | Na | Na | Na | Na | Na | Na |
| 41839T1 | Na | None | Na | -32.134 | -35.763 | 0.032898 | Na | Na | Na | Na | Na | Na |
| 664T1 | Na | None | Na | -10.861 | -0.77699 | 0.033167 | Na | Na | Na | Na | Na | Na |
| 22213T1 3 | protein | SPU_026357 | protein | -12.425 | -2.061 | 0.035541 | Sp-Cog4 | 1700 | 35 | 84 | 97 | 90 |

| | | | | | | | | | | | | |
|---------|----|------------|----|---------|---------|----------|---------------------|------|-------|------|------|------|
| 4926T3 | Na | SPU_013076 | Na | -11.68 | -1.2581 | 0.039754 | Sp-AJPX1 | 3887 | 22665 | 8782 | 5301 | 6635 |
| 12505T1 | Na | None | Na | -13.58 | -1.8855 | 0.04085 | Na | Na | Na | Na | Na | Na |
| 65813T1 | Na | None | Na | -10.512 | -1.0764 | 0.041658 | Na | Na | Na | Na | Na | Na |
| 4098T1 | Na | SPU_013822 | Na | -13.466 | -2.2351 | 0.046102 | Sp-MSP130-related-1 | 0 | 312 | 275 | 98 | 28 |
| 35026T1 | Na | None | Na | -15.251 | -3.2698 | 0.048046 | Na | Na | Na | Na | Na | Na |
| 26684T1 | Na | None | Na | -13.004 | -1.512 | 0.048046 | Na | Na | Na | Na | Na | Na |

OasesTransID: Transcript identifier from Oases de novo transcriptome.

OasesGOname: Name assigned to Oases transcript by Blast2GO.

SPU Identifier: Top blast hit of Oases transcript to SPU gene predictions.

SPUGOname: Name assigned to SPU transcript by Blast2GO.

HelicosLogConc: Transcript abundance calculated by EdgeR.

HelicosFC: Fold change enrichment of the transcript; positive values are sMic enriched, negative values are sMic depleted.

HelicosFDR: Statistical significance (false discovery rate) for differentially expressed transcript.

Microarray Annotation: Annotation given in temporal gene expression database.

2hpf-72hpf: Relative transcript abundance by temporal microarray data.

Supplemental Table 6: PCR primers for WMISH probes.

| Primer Name | Sequence (5'→ 3') |
|--------------------|---|
| Baf250 | F: TCGACAACCACCACTACCAA R: taatagactcactatagggTGTTGTTCACTCCACCCGTA |
| CEBP | F: AGGGCTGAGGTACAGCAGAA R: taatagactcactatagggCCCTCGACACGTTTCTTTA |
| FoxN2/3 | F: CGAATGGACAAAGGACCACT R: taatagactcactatagggCTTGGTGATGGGGTACACT |
| Sprouty2 | F: GCTCTGTTCCCTTGAGCAAC R: taatagactcactatagggGCAGGGATCATCCGTACAGT |
| Ctdspl2/SCP2 | F: AAGCCACCAATCCTGTGTTC R: taatagactcactatagggCAGAAAGGCACAAGCAATCA |
| z62 | F: AGTCTAGCAAATGGCGTCGT R: taatagactcactatagggATGTAACCACATTCGCAGCA |
| MibL | F: CTGGACAGACACCAGAGCAA R: taatagactcactatagggCATCTGCTCCGTGCATAAGA |
| CNOT6 | F: GCAGGTGCTAGGTCTGAAGG R: taatagactcactatagggCGAGTTGGAGGAGAAGTTGC |
| CNOT7 | F: TGCCAACTCAAACCAATGAA R: taatagactcactatagggGCACCCTGGTTAAAAGGTCA |

Supplemental Table 7: Primers used for generating constructs.

| Primer Name (description) | Sequence (5'→ 3') |
|--|---|
| CNOT6 3'UTR (reporter construct) | F: GGGACTCAGGGTGGTGTTC R: ACAGAGAATTGCACATTGGTTGG |
| PRE1 (site-directed mutagenesis) | F: TTCGTA CTTTGTACAGTGaaaAAATTGATGCTATTTTGC R: GCAAAATAGCATCAATTTtttCACTGTACAAAGTACGAA |
| PRE2 (site-directed mutagenesis) | F: ATTTGAGACCATGGGTTGaaaAAATGAGGATTTGAACCA R: TGGTTCAAATCCTCATTtttCAACCCATGGTCTCAAAT |

Supplemental Table 8: Custom morpholino sequences.

| MASO Name | Sequence (5' → 3') | Concentration of injection solution |
|------------------|---------------------------|--|
| Nanos | GTGACTAAAGTGCGTGGAAACTCGA | 500 μ M |
| CNOT6 1 | ATTTATCTTTGGGCATCCTGGTGGC | 500 μ M |
| CNOT6 2 | GTCGGTTTTACACAGTTCAGGAGGC | 500 μ M |
| PRE1 | AATAGCATCAATTTACACACTGTAC | 1,000 μ M |
| PRE2 | GTTCAAATCCTCATTACACAACCC | 1,000 μ M |

**Appendix V: A computational approach for the identification of the sex
chromosomes in *S. purpuratus***

Adrian Reich, Zhijin Wu, and Gary Wessel

Unpublished

CONTRIBUTION

I conducted all experiments and analyses except for the copy number variation analysis and fold coverage modeling.

ABSTRACT

Many organisms have two distinct sexes that are determined by the presence, absence or ratio of different chromosomes. There exist a number of chromosomal based sex determination systems, but outside of several model organisms, the mechanisms of sex determination are largely unknown. In this study we attempt to identify the sex determination system of the purple sea urchin, *Stronglyocentrotus purpuratus*, a deuterostome closely related to chordates, using a novel, bioinformatic based approach.

INTRODUCTION

For animals that have two distinct sexes (as opposed to hermaphroditic or asexual animals), sex determination in animals falls broadly into two categories: strictly genetic or requiring an external environmental cue (Uzzell, 1984). The sex determination of a majority of animals is genetic or chromosomally specified, however, numerous systems have evolved. Often one sex is homochromatic, having one or two copies of the same sex chromosome, while the other sex is heterochromatic, having one copy of two different sex chromosomes. Most organisms that have the X and Y sex chromosomes (e.g. mammals), the female is homochromatic (XX) and the male is heterochromatic (XY). In contrast, organisms that have the Z and W sex chromosomes (e.g. birds), the male is homochromatic (ZZ) and the female is heterochromatic (ZW). Often in chromosome derived sex determination systems, the chromosome that is shared between the two sexes (e.g. X in mammals, Z in birds) has a much greater proportion of gene coding products compared to the non-obligate chromosome.

Two lines of evidence suggest that the purple sea urchin uses a chromosomal based sex determination strategy. If individual blastomeres from a two cell embryos that are raised to sexual maturity, the twinned adults are of the same sex (Cameron *et al.*, 1996). Even though there were a small number of twinned embryos raised until sexual maturity, there was only a 0.2% cumulative chance that the results were due to random chance (Cameron *et al.*, 1996). The authors concluded that the results supported a chromosomal based sex determination as opposed to an unknown environmental influence. The second line of evidence supporting genetic sex determination in the purple sea urchin comes from karyotyping experiments. The authors performed chromosomal squashes on 34 *S. purpuratus* embryos and observed two distinct karyotypes (Eno *et al.*, 2009). In both karyotypes, 20 invariant chromosomal pairs were identified as well as a 21st pair which the authors described as presumptive sex chromosomes (Eno *et al.*, 2009). In 19 of the 34 squashes, the 21st pair consisted of a large chromosome paired with a diminutive chromosomal body; the remaining 15 of the 34 squashes had two large chromosomes paired together (Eno *et al.*, 2009).

Due to the use of embryos in the sample preparation, the authors were unable to determine which sex was the homochromatic karyotype.

The genome of *S. purpuratus* was assembled from the sperm of a single male (Sea Urchin Genome Sequencing *et al.*, 2006). If sex determination in *S. purpuratus* was determined with an X and Y chromosome as in many mammals, then the urchin genome would contain scaffolds from all 20 autosomal chromosomes and both the X and the Y chromosomes, because each haploid sperm would contain one copy of each autosome and either an X or a Y chromosome. In an XY sex determination system, therefore, the average sequencing coverage of an autosome would be twice that of a sex chromosome. If on the other hand, sex determination was determined with a ZW system, each sperm would contain one autosome and one Z chromosome and the genome would lack all W chromosome derived sequences. The average sequencing coverage of an autosome and a sex chromosome would be equal in the case of a ZW sex determination system with genomic reads derived from sperm DNA.

We hypothesized that we could identify the method of sex determination in the purple sea urchin *S. purpuratus* by mapping the short reads used in the assembly of the genome and analyzing the fold coverage of each scaffold. Furthermore, we sought to identify genetic scaffolds that belong to the sex chromosomes.

RESULTS AND DISCUSSION

The genome of *S. purpuratus* is estimated to be 815Mb in size (Sea Urchin Genome Sequencing *et al.*, 2006) and the karyotype data suggests that the genome is contained on 21 pairs of chromosomes that are nearly equal in size (Eno *et al.*, 2009). From this data, we estimate that the amount of DNA found on the sex chromosomes is 4-5% of the genome, or approximately 40MB in size.

Testing XY sex determination

If *S. purpuratus* uses an XY sex determination system, then both the X and Y chromosomes are represented in the sequenced genome of sperm. Mapping the genomic reads to the genome, should therefore yield twice the coverage on autosome derived scaffolds compared to sex chromosome scaffolds. We performed a modeling series in order to test what the minimum fold coverage would be necessary to distinguish a twofold difference between sex chromosomes and autosomes that were normally distributed. We modeled sex chromosome derived scaffolds accounting for 5% of the genome and would on average have half the sequencing coverage (Fig. 1). Tallying the total number of raw sequence reads, we calculated that autosomes would be covered at 50 fold coverage and sex chromosomes at 25 fold coverage (similar to modeling run Fig. 1c). However due to the low quality reads and mapping efficiency (Table 1) the calculated fold coverage of autosomes and sex chromosomes was only 25 fold and 12 fold, respectively.

The low overall fold coverage of autosomes and sex chromosomes did not allow for clear separation between the two populations (Fig. 2). The data fell into a single distribution, with no obvious shoulder on the left side of the graph which would hint at two overlapping distributions. The two scaffolds with very few reads mapping to them did not have any SPU gene predictions (Fig.2 and data not shown).

When we examined the scaffolds on a case by case basis, we detected significant copy number variation (CNV) within individual scaffolds (Fig. 3). One potential explanation of this data could be due to the very high heterozygosity between two different haplotypes in the urchin (Britten *et al.*, 1978). The sequenced urchin for the genome project was a wild type individual and therefore the genome was assembled as two separate haplotypes, and during the assembly of the urchin genome, every effort was made to collapse the sites from the two different haplotypes into a single hybrid region (Sea Urchin Genome Sequencing *et al.*, 2006). Therefore, in certain regions of the scaffold, both genome haplotypes are represented as hybrid region (therefore receive a full complement of mapping reads), while in other regions of the same scaffold, only a single haplotype is represented (the reads from only that haplotype map and therefore half the

coverage is observed). In individual scaffolds, significant CNVs (that are two fold in nature) within each scaffold greatly complicates the ability to assign a scaffold as derived from an autosome or a sex chromosome (also a twofold difference).

A second explanation for the CNVs observed, is that the majority of the differences between two haplotypes in sea urchins are due to indels as opposed to SNPs (Britten *et al.*, 2003). If a read maps to an insertion found within one of the haplotypes, the read will not be able to map to the other haplotype and if that region is found in the hybrid genome assembly, then that region of the genome will only have half the coverage of another region.

Testing ZW sex determination

In a ZW sex determination system, the genome derived from sperm would not contain any sequences from the W chromosome and the scaffolds originating from the Z chromosome would have the same fold coverage of the autosomes. In order to test for ZW sex determination, we assembled a *de novo* transcriptome of a developmental series of embryos. It is likely that during development, transcripts originating from both the Z and W chromosomes would be expressed. Therefore, we mapped genomic DNA read fragments to the transcriptome, with the intention of identifying transcripts that do not have any reads mapping to them. This might identify transcripts that are derived from the W chromosome. We were unable to detect any transcripts that did not have any genomic read fragments mapping to them (data not shown). If any transcripts were identified in this manner it would be important to test for the transcript in eggs by nuclear FISH.

FUTURE DIRECTIONS

The initial analyses using short reads were inconclusive in determining the sex determination system of *S. purpuratus*. The current fold coverage of the genomic reads is insufficient to separate autosomes from sex chromosomes in the case of XY sex determination. This could be overcome with more reads, but it would also require a robust method to accurately

identify the fold coverage of a scaffold in the presence of CNVs on every scaffold. The present strategy mapped reads to the genome uniquely, meaning that if reads mapped to multiple locations, they were discarded. This minimized any effects of reads mapping to repetitive elements in the genome. Mapping genomic reads to the *de novo* transcriptome is still a viable method but would require a more nuanced approach. For example, it is unlikely that a full length transcript from the W chromosome would have little to no sequence similarity to any other region of the genome. It will be important to examine the distributions of read fragments on each transcript because if a W transcript shares a conserved motif then that region of the transcript will have reads mapping to it, but the rest of the transcript will be devoid of mapping reads. This level of detail was beyond the scope of the first pass of the analysis.

Another method that was begun but has not yet been completed was building a *de novo* restriction map of the entire genome. Using an optical mapping technique, greater than 30 fold coverage of the *S. purpuratus* genome was obtained in single DNA molecules longer than 150kb each. This data can be assembled into restriction maps up to the length of chromosomal arms. The existing genome scaffolds can then be mapped to this *de novo* restriction map to create super scaffolds. The benefit of this approach is that instead of analyzing differences between tens of thousands of scaffolds, the linkage information from the super scaffolding will allow for testing differences in several hundred super-scaffolds. Furthermore, we hypothesize that the long DNA molecules are quantitative and that we can use the pileups in the same way that we tried to use the short read data. In the case of an XY sex determination system, autosomes should have twice the fold coverage of sex chromosomes but because the individual molecules are at least 150kb long and the super scaffolds are on the order of megabases in size, ambiguities of fold coverage due to CNVs should be minimized.

MATERIALS AND METHODS

Read processing

All Illumina reads (SRR446979, SRR446980 and SRR446981) that contributed to the assembly of the genome (SpBase version 3.1; NCBI BioProject PRJNA10736) were processed as following. Original reads were trimmed from 150bp to 98bp due to low quality (1st base and the last 51 bases removed); any reads after trimming with an average phred quality score less than 20 were also removed.

Genome mapping

The purple sea urchin genome scaffolds were sampled so that all scaffolds longer than 1kb were retained (16,110 scaffolds); yielding a total sequence length of 926Mb. All paired end reads were treated as single end reads for the purpose of mapping to the genome. The reads were mapped to the genome as read fragments as opposed to whole reads using bowtie (version 0.12.7) with the following command: “bowtie --trim5 X --trim3 Y AllReads.fastq -v 2 -k 1 --best”. Three separate mapping experiments were run where X was 0, 32, or 64 and Y was 66, 34, or 2, respectively (Table 1). This resulted in read fragments 32nt long that mapped uniquely to the genome with a maximum mismatch of 2nt per read fragment. For each of the 3 mapping experiments there were total of 232,740,105 read fragments, and the number of mapping read fragments and mapping efficiency can be seen in Table 1.

Each genomic scaffold was divided into 50nt bins and tallied with the number of read fragments mapping to each bin. Maximum fold coverage of each scaffold and the average read fragments mapping to each scaffold was analyzed.

***S. purpuratus* de novo transcriptome sequencing and assembly**

RNA was extracted from a developmental series of *S. purpuratus* embryos: ovary, 32 cell stage, 15hr blastula, 41 hr gastrula, and 4-day pluteus. The RNA was cleaned, using the RNEasy Mini kit (Qiagen) with on column DNA digestion. Using standard procedures, the purified RNA was processed with the Illumina mRNA-Seq kit and sequenced on a single lane of a GAIIx with a

paired-end read length of 105bp. The transcriptome was assembled using Velvet (1.0.09) and Oases (0.1.14) with a k-mer of 31 (Schulz *et al.*, 2012). Each locus generates several similar sequences; therefore a single exemplar sequence was selected for each locus. The exemplar selected had the highest expression transcript that was also above a minimum length threshold, which was then annotated with Blast2GO. The exemplar transcript was compared by BLAST (minimum score of 1e-5) with the *S. purpuratus* SPU gene predictions for further annotation (Conesa *et al.*, 2005; Sea Urchin Genome Sequencing *et al.*, 2006).

Long single molecule DNA visualization

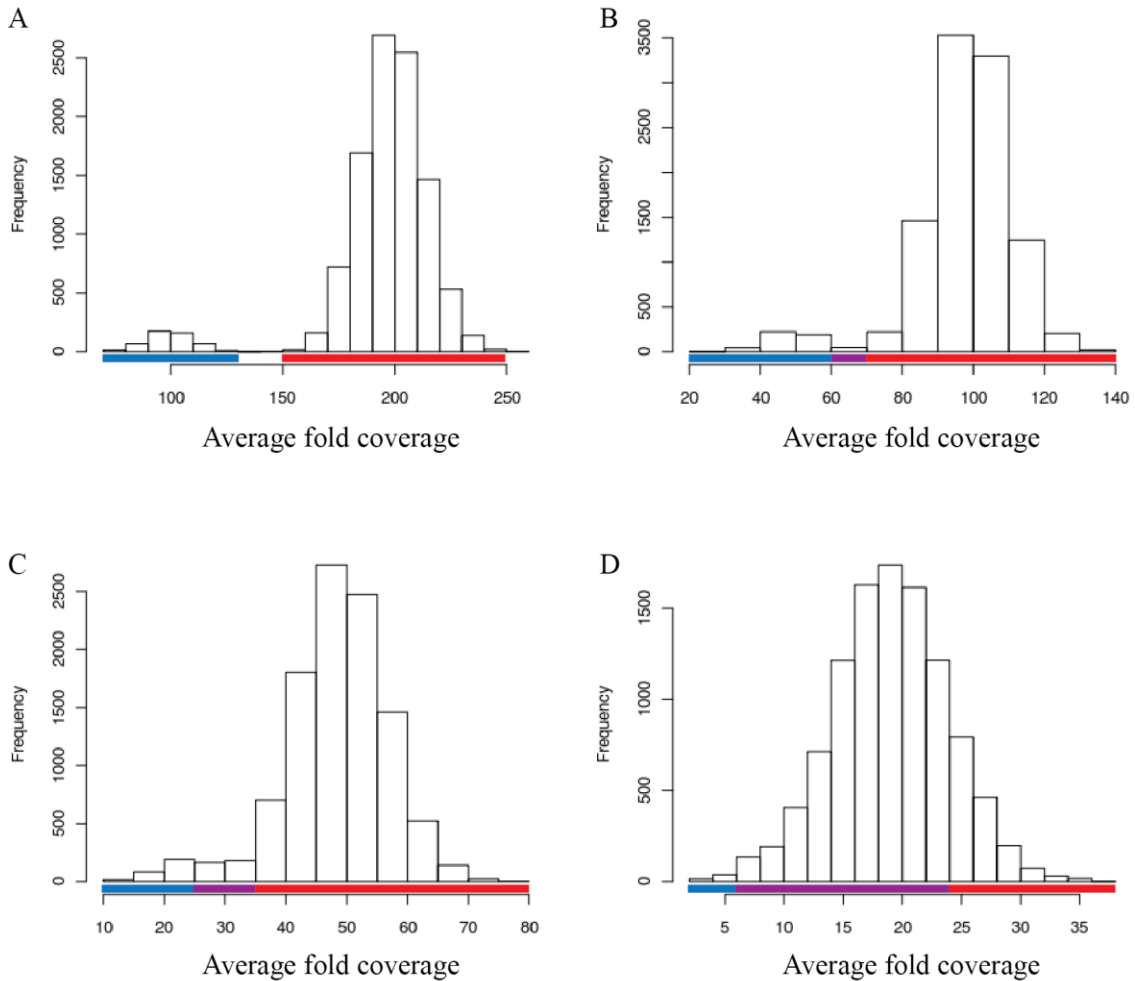
Megabase sized DNA was prepared as previously described (Zhang *et al.*, 2012). Briefly, 10ug of *S. purpuratus* sperm was embedded in an agarose plug and processed using the CHEF Mammalian Genomic DNA Plug Kit (Bio-Rad, Hercules, CA). The DNA was run on an Irys instrument using the IrysChip V1 (BioNano Genomics, San Diego, CA). Thirty fold coverage of the *S. purpuratus* genome was obtained.

REFERENCES

- Britten, R.J., Cetta, A., and Davidson, E.H. (1978). The single-copy DNA sequence polymorphism of the sea urchin *Strongylocentrotus purpuratus*. *Cell* *15*, 1175-1186.
- Britten, R.J., Rowen, L., Williams, J., and Cameron, R.A. (2003). Majority of divergence between closely related DNA samples is due to indels. *Proc Natl Acad Sci U S A* *100*, 4661-4665.
- Cameron, R.A., Leahy, P.S., and Davidson, E.H. (1996). Twins raised from separated blastomeres develop into sexually mature *Strongylocentrotus purpuratus*. *Dev Biol* *178*, 514-519.
- Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* *21*, 3674-3676.
- Eno, C.C., Bottger, S.A., and Walker, C.W. (2009). Methods for karyotyping and for localization of developmentally relevant genes on the chromosomes of the purple sea urchin, *Strongylocentrotus purpuratus*. *Biol Bull* *217*, 306-312.
- Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* *28*, 1086-1092.
- Sea Urchin Genome Sequencing, C., Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess, D.R., *et al.* (2006). The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* *314*, 941-952.
- Uzzell, T. (1984). Sex determination: evolution of sex determining mechanisms. *Science* *224*, 733-734.
- Zhang, M., Zhang, Y., Scheuring, C.F., Wu, C.C., Dong, J.J., and Zhang, H.B. (2012). Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *Nat Protoc* *7*, 467-478.

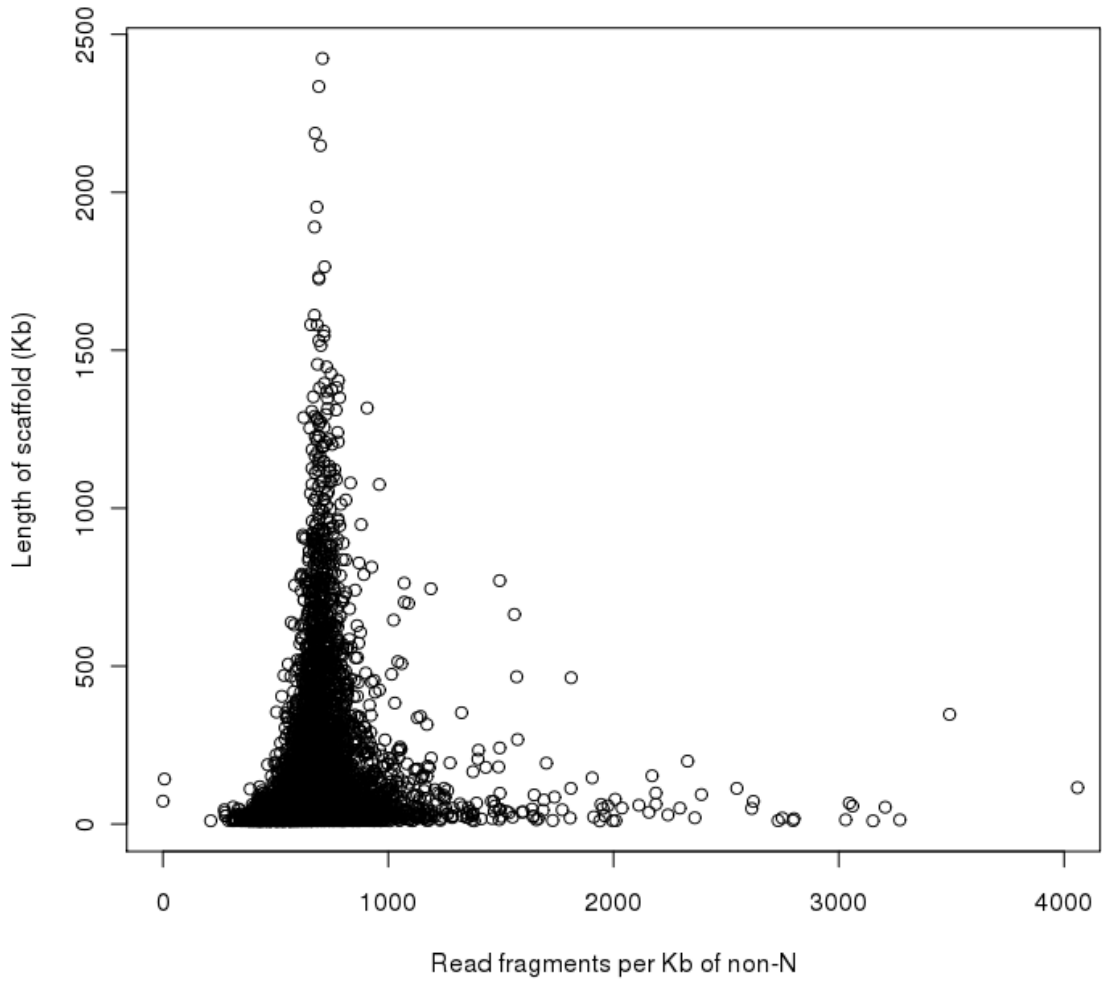
FIGURES AND TABLES

Figure 1: Modeling run of fold coverage difference between sex chromosomes and autosomes.



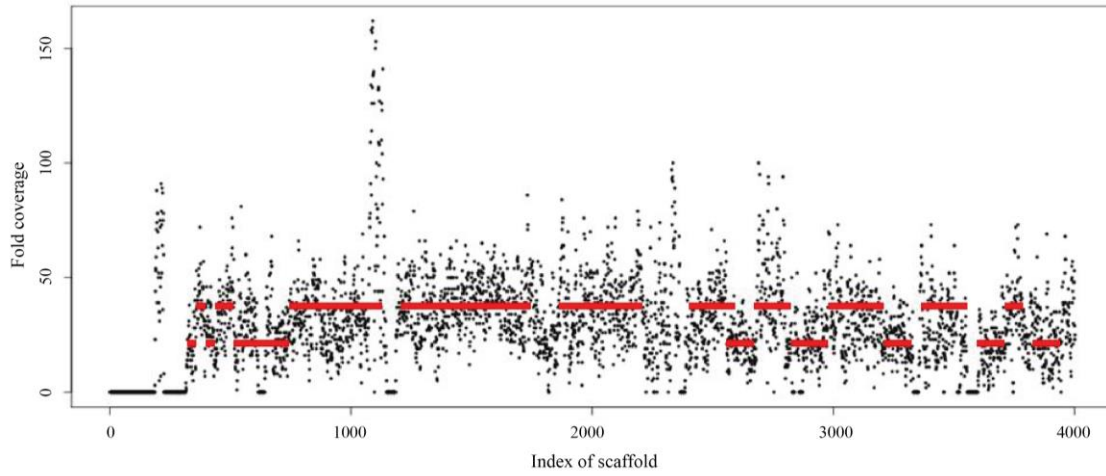
A) With an average of 200 fold sequencing coverage for 95% of the genome (autosomes, red) and 100 fold coverage for 5% of the genome (sex chromosomes, blue), there is clear separation between the two normally distributed populations. B) There is again, clear separation between the autosomes (100 fold coverage) and sex chromosomes (50 fold coverage), though some scaffolds are unknown (purple). C) The two normally distributed populations begin to overlap significantly, but a number of sex chromosome scaffolds are still clearly distinguishable. D) With only 20 fold coverage of autosomes it becomes very difficult to distinguish the two populations even under ideal modeling conditions.

Figure 2: Number of mapped read fragments per Kb of known sequence of each scaffold.



The number of mapped read fragments for each scaffold was normalized to the length of unambiguous nucleotides in each scaffold.

Figure 3: High copy number variation is seen in an individual scaffold.



The copy number variation seen in a single 200kb scaffold; each point is a 50bp window. The red line shows the average copy number for that region of the scaffold.

Table 1: Mapping efficiency of read fragments to genome.

| | bowtie -trim5 | bowtie -trim3 | Range of read fragment | Reads mapping | Mapping percentage |
|---------|--------------------------|--------------------------|-----------------------------------|--------------------------|-------------------------------|
| Round 1 | 0 | 66 | 1-32 | 193,397,724 | 83.10% |
| Round 2 | 32 | 34 | 33-64 | 185,443,818 | 79.68% |
| Round 3 | 64 | 2 | 65-96 | 176,579,912 | 75.87% |

Every round of read fragment mapping used the same set of reads (98bp long after trimming), but used a different range within the read.

Appendix VI: Assembly of the genome of an early branching echinoderm,
Oxycomanthus japonicus

Adrian Reich, Mariko Kondo, Koji Akasaka, and Gary Wessel

Unpublished

CONTRIBUTION

I conducted all experiments and analyses.

ABSTRACT

Echinoderms are a diverse group of organisms with a rich evolutionary history. Of the extant echinoderms, the crinoids are the earliest branching member and are sister group to all other echinoderms. As such, we sequenced and assembled a draft genome of the sea lily *Oxycomanthus japonicus* in order to test evolutionary transitions with echinoderms, with a particular emphasis on the hox gene cluster.

INTRODUCTION

Echinoderms arose nearly 570 million years ago (Pisani *et al.*, 2012) and rapidly diversified over the following 10-15 million years into the five extant groups of echinoderms as well as several extinct groups (Smith *et al.*, 2013). Echinoderms are important model organisms in developmental biology and as a closely related group to chordates; as such, echinoderms occupy an important evolutionary node.

Hox genes, which belong to a group of homeodomain transcription factors, are important factors for development and regulate patterning along the posterior/anterior axis in bilaterians (McGinnis *et al.*, 1984). Interestingly, the Hox genes and ParaHox genes are sometimes expressed temporally and/or spatially colinear; the expression of these genes follows the relative position of these genes on the chromosome. The echinoderm Hox cluster that has been studied in the greatest depth is that of the purple sea urchin, *Strongylocentrotus purpuratus*. The Hox cluster in this organism has a unique rearrangement; a translocation of Hox 1, 2, and 3 and loss of Hox 4, along with several secondary losses, duplications and inversions of other Hox genes (Cameron *et al.*, 2006). Interestingly, many of the Hox genes in *S. purpuratus* still maintain the colinear expression during embryonic development (Arenas-Mena *et al.*, 2000). It is unknown if the Hox gene cluster rearrangement observed in *S. purpuratus* is specific to this species, to Echinoidea, or to all of echinoderms (Pascual-Anaya *et al.*, 2013). There is strong support for Crinoids as the sister group to all other echinoderms (Janies *et al.*, 2011; Pisani *et al.*, 2012, and Chapter 2) and Chapter 2). This critical evolutionary node may resolve some of the conflicting data surrounding the origin of the Hox genes in deuterostomes that pattern the posterior of the embryo (Pascual-Anaya *et al.*, 2013).

Not only are Hox genes critical during development, but they are also important in regeneration in a number of species, including axolotl (Carlson *et al.*, 2001; Torok *et al.*, 1998), as well as in multiple groups of echinoderms (Ben Khadra *et al.*, 2014). Crinoids have a remarkable

ability to regenerate (Candia Carnevali and Bonasoro, 2001), and have retained that regenerative capacity, in all likelihood, through evolutionary history (Gahn and Baumiller, 2010).

Because of the critical evolutionary node that crinoids occupy and the remarkable regenerative capacity of these organisms, we sequenced, assembled and annotated a *de novo* genome of *O. japonicus* to study the arrangement of the Hox gene cluster and expression of these genes.

Numerous genome assemblers are available, including: Celera (Myers *et al.*, 2000), Velvet (Zerbino and Birney, 2008), ALLPATHS-LG (Butler *et al.*, 2008), ABySS (Simpson *et al.*, 2009), SOAPdenovo (Li *et al.*, 2010), Ray (Boisvert *et al.*, 2010), SGA (Simpson and Durbin, 2012), among others. It is time consuming to assemble a *de novo* genome with even only a single set of parameters using every available assembler and then compare the assemblies to identify the “best” assembly. This is made even more difficult by the fact that there is no, one metric that can define a good assembly, though work is progressing in this field (Bradnam *et al.*, 2013; Hunt *et al.*, 2013; Parra *et al.*, 2007). One valuable resource is the Assemblathon competition, where groups of researchers assemble the same set of read data using their assembler of choice (Bradnam *et al.*, 2013; Earl *et al.*, 2011). Using many different metrics for assembly quality, some assemblers do better than others. However, most assemblers had very different successes using different datasets and read compositions, which was one of the main conclusions of Assemblathon2 (Bradnam *et al.*, 2013). The sample that was most similar to the *O. japonicus* dataset in terms read composition (exclusively Illumina paired-end and mate pair reads), and to a lesser extent, genome size, was the boa constrictor from Assemblathon2 (Bradnam *et al.*, 2013). As such, we selected the SGA assembler to assemble the *O. japonicus* dataset because it performed the best in almost all metrics of quality on the snake dataset (Bradnam *et al.*, 2013).

RESULTS AND DISCUSSION

The genome of *O. japonicus* is estimated to be 650Mb as estimated by DAPI staining of nuclei (data not shown). The combined read coverage is approximately 60 fold coverage, with a combination of paired-end, mate pair, single end and BAC end sequencing reads, and in excess of 350 fold physical coverage. Due to the nature of mate pair library construction, many reads in mate pair experiments cannot be definitively classified as a mate pair because the junction marking the circularization of the linear molecule was not sequenced in either of the reads. After filtering the mate pair data using NextClip (Leggett *et al.*, 2014), all reads that were not classified as mate pair were classified as single end sequencing. The true fold coverage distribution was therefore: 30 fold coverage of paired-end reads, 15 fold coverage of single end reads, and 12 fold coverage of mate pair reads (Table 1).

FUTURE DIRECTIONS

The genome and subsequent annotation are ready to be assembled but has not yet begun. Prior to annotation, the genome must be screened for repetitive elements; once complete the *O. japonicus de novo* transcriptome will be used to guide the gene prediction pipeline (Chapter 2, and Yandell and Ence, 2012). Depending on the quality of the genome assembly, I could also use the *de novo* transcriptome to scaffold a fragmented genome (Li and Copley, 2013).

MATERIALS AND METHODS

Paired-end sequencing

Isolated DNA was processed with the Paired-End Sample Preparation Kit (Illumina Inc., San Diego, CA) using standard procedures. Briefly, the maximum recommended 5ug of genomic DNA was sheared using a nebulizer. Following end repair and adapter ligation, the sample was run on a 2% agarose gel and size selected by cutting a thin slice of gel of about 2mm thick at approximately 450bp as estimated by adjacent ladders. After PCR amplification, the sample was

run on a second gel to purify the final library and exclude the adapter sequences. The library was run on three lanes of a GAIIIX (Table 1; Illumina Inc., San Diego, CA).

Mate pair sequencing

DNA was isolated from the same individual as the paired-end sequencing data, and processed using the Nextera Mate Pair Sample Prep Kit (Illumina Inc., San Diego, CA). Three separate libraries were processed and sequenced: one gel-free sample, and two agarose gel size-selection samples. The input DNA for the gel-free sample and agarose size-selection samples were 1ug and 4ug, respectively. The gel-free sample was processed using standard procedures and the size selection sample was run on a 0.6% agarose gel and several gel slices were cut from the gel. Prior to circularization, all samples were run on a 12 capillary Fragment Analyzer (Advanced Analytical Technologies, Inc., Ames, Iowa) to measure insert size (Supplemental Fig. 1). Samples were sheared on a Covaris S220 (Covaris, Inc., Woburn, MA) using the recommended settings; following end-repair and adapter ligation, all libraries were PCR amplified for 15 rounds. The three multiplexed samples were run on a single lane of HiSeq 2500 (non rapid-run cycle; Illumina Inc., San Diego, CA). To identify true mate-pair reads, the reads were processed using NextClip (Leggett *et al.*, 2014); reads not classified as mate pair were trimmed of adapter sequences and treated as single end sequences during assembly (Table 1).

BAC end sequencing

A BAC library was prepared using standard techniques from the same individual from which the Illumina libraries were prepared. The library was screened for BACs that contained a Hox gene homolog and sequenced from either end (Table 1).

Genome assembly

The reads will be assembled with the SGA genome assembler (Simpson and Durbin, 2012).

De novo transcriptome sequence and assembly

Ovary dissected from a single gravid female was put in Trizol (Invitrogen). RNA was cleaned and purified with on-column DNA digestion using a Qiagen RNeasy Micro column. The sequencing library was constructed using the Illumina mRNA-Seq Sample Prep Kit with the maximum recommended 10ug RNA input. The protocol was followed exactly except for an agarose gel size selection step prior to PCR enrichment of the library. The transcriptome was assembled using the agalma pipeline (ver. 0.3.5; Dunn *et al.*, 2013; Howison *et al.*, 2012) in conjunction with the trinity *de novo* transcriptome assembler (ver. r2013_08_14; Haas *et al.*, 2013), using default settings.

DATA AVAILABILITY

The genomic reads and assembly have been deposited in the GenBank database (NCBI BioProject no. PRJNA236227). The *de novo* transcriptome and reads therein can be found under BioProject no. PRJNA236087.

REFERENCES

- Arenas-Mena, C., Cameron, A.R., and Davidson, E.H. (2000). Spatial expression of Hox cluster genes in the ontogeny of a sea urchin. *Development* 127, 4631-4643.
- Ben Khadra, Y., Said, K., Thorndyke, M., and Martinez, P. (2014). Homeobox genes expressed during echinoderm arm regeneration. *Biochem Genet* 52, 166-180.
- Boisvert, S., Laviolette, F., and Corbeil, J. (2010). Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol* 17, 1519-1533.
- Bradnam, K.R., Fass, J.N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., Boisvert, S., Chapman, J.A., Chapuis, G., Chikhi, R., *et al.* (2013). Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience* 2, 10.
- Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I.A., Belmonte, M.K., Lander, E.S., Nusbaum, C., and Jaffe, D.B. (2008). ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Res* 18, 810-820.
- Cameron, R.A., Rowen, L., Nesbitt, R., Bloom, S., Rast, J.P., Berney, K., Arenas-Mena, C., Martinez, P., Lucas, S., Richardson, P.M., *et al.* (2006). Unusual gene order and organization of the sea urchin hox cluster. *J Exp Zool B Mol Dev Evol* 306, 45-58.
- Candia Carnevali, M.D., and Bonasoro, F. (2001). Microscopic overview of crinoid regeneration. *Microsc Res Tech* 55, 403-426.
- Carlson, M.R., Komine, Y., Bryant, S.V., and Gardiner, D.M. (2001). Expression of Hoxb13 and Hoxc10 in developing and regenerating Axolotl limbs and tails. *Dev Biol* 229, 396-406.
- Dunn, C.W., Howison, M., and Zapata, F. (2013). Agalma: an automated phylogenomics workflow. *BMC Bioinformatics* 14, 330.
- Earl, D., Bradnam, K., St John, J., Darling, A., Lin, D., Fass, J., Yu, H.O., Buffalo, V., Zerbino, D.R., Diekhans, M., *et al.* (2011). Assemblathon 1: a competitive assessment of de novo short read assembly methods. *Genome Res* 21, 2224-2241.
- Gahn, F.J., and Baumiller, T.K. (2010). Evolutionary history of regeneration in crinoids (Echinodermata). *Integr Comp Biol* 50, 514a-514m.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., *et al.* (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8, 1494-1512.
- Howison, M., Sinnott-Armstrong, N., and Dunn, C.W. (2012). BioLite, a lightweight bioinformatics framework with automated tracking of diagnostics and provenance. *Proceedings of the 4th USENIX Workshop on the Theory and Practice of Provenance (TaPP '12)*.
- Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M., and Otto, T.D. (2013). REAPR: a universal tool for genome assembly evaluation. *Genome Biol* 14, R47.
- Janies, D.A., Voight, J.R., and Daly, M. (2011). Echinoderm phylogeny including *Xyloplax*, a progenetic asteroid. *Syst Biol* 60, 420-438.
- Leggett, R.M., Clavijo, B.J., Clissold, L., Clark, M.D., and Caccamo, M. (2014). NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* 30, 566-568.
- Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y., *et al.* (2010). The sequence and de novo assembly of the giant panda genome. *Nature* 463, 311-317.
- Li, Y.I., and Copley, R.R. (2013). Scaffolding low quality genomes using orthologous protein sequences. *Bioinformatics* 29, 160-165.
- McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A., and Gehring, W.J. (1984). A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell* 37, 403-408.

- Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A., *et al.* (2000). A whole-genome assembly of *Drosophila*. *Science* 287, 2196-2204.
- Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061-1067.
- Pascual-Anaya, J., D'Aniello, S., Kuratani, S., and Garcia-Fernandez, J. (2013). Evolution of Hox gene clusters in deuterostomes. *BMC Dev Biol* 13, 26.
- Pisani, D., Feuda, R., Peterson, K.J., and Smith, A.B. (2012). Resolving phylogenetic signal from noise when divergence is rapid: a new look at the old problem of echinoderm class relationships. *Mol Phylogenet Evol* 62, 27-34.
- Simpson, J.T., and Durbin, R. (2012). Efficient de novo assembly of large genomes using compressed data structures. *Genome Res* 22, 549-556.
- Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Res* 19, 1117-1123.
- Smith, A.B., Zamora, S., and Alvaro, J.J. (2013). The oldest echinoderm faunas from Gondwana show that echinoderm body plan diversification was rapid. *Nat Commun* 4, 1385.
- Torok, M.A., Gardiner, D.M., Shubin, N.H., and Bryant, S.V. (1998). Expression of HoxD genes in developing and regenerating axolotl limbs. *Dev Biol* 200, 225-233.
- Yandell, M., and Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet* 13, 329-342.
- Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18, 821-829.

FIGURES AND TABLES

Table 1: Total read breakdown and coverage used to assemble genome.

| Read type (insert size) | Number of reads | Read length | Read coverage | Physical coverage |
|--------------------------------|------------------------|--------------------|----------------------|--------------------------|
| Single End | 102,226,516 | varies | 15.72 | 15.72 |
| Paired-End (200bp) | 98,193,184 | 105 | 31.72 | 61.94 |
| Mate Pair (2.55kb) | 14,005,314 | 100 | 3.04 | 54.94 |
| Mate Pair (3.6kb) | 32,783,606 | 100 | 7.11 | 154.39 |
| Mate Pair (4.7kb) | 9,816,399 | 100 | 2.13 | 71.04 |
| BAC end (varies) | | varies | | |

SUPPLEMENTAL INFORMATION

Supplemental Figure 1: Insert sizes of genome mate pair libraries.

