Oogenesis: Single cell development and differentiation

Jia L. Song, Julian L. Wong, Gary M. Wessel *

Department of Molecular and Cellular Biology and Biochemistry, Box G, Brown University, Providence RI 02912, USA

Received for publication 10 May 2006; revised 27 July 2006; accepted 28 July 2006
Available online 5 August 2006

Abstract

Oocytes express a unique set of genes that are essential for their growth, for meiotic recombination and division, for storage of nutrients, and for fertilization. We have utilized the newly sequenced genome of *Strongylocentrotus purpuratus* to identify genes that help the oocyte accomplish each of these tasks. This study emphasizes four classes of genes that are specialized for oocyte function: (1) Transcription factors: many of these factors are not significantly expressed in embryos, but are shared by other adult tissues, namely the ovary, testis, and gut. (2) Meiosis: A full set of meiotic genes is present in the sea urchin, including those involved in cohesion, in synaptonemal complex formation, and in meiotic recombination. (3) Yolk uptake and storage: Nutrient storage for use during early embryogenesis is essential to oocyte function in most animals; the sea urchin accomplishes this task by using the major yolk protein and a family of accessory proteins called YP30. Comparison of the YP30 family members across their conserved, tandem fasciclin domains with their intervening introns reveals an incongruence in the evolution of its major clades. (4) Fertilization: This set of genes includes many of the cell surface proteins involved in sperm interaction and in the physical block to polyspermy. The majority of these genes are active only in oocytes, and in many cases, their anatomy reflects the tandem repeating interaction domains essential for the function of these proteins. Together, the expression profile of these four gene classes highlights the transitions of the oocyte from a stem cell precursor, through stages of development, to the clearing and re-programming of gene expression necessary to transition from oocyte to egg to embryo.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Oocytes; Transcription factors; Meiosis; Mitosis; Cortical granules; Vitellogenesis; Fertilization; Gene structure

Introduction

Oogenesis is a prolonged process of cell interactions between the oocyte and its somatic accessory cells. It involves cell growth, new organelle synthesis and storage, cell differentiation, and meiosis. For the sea urchin, an animal capable of storing and shedding millions of eggs at once, it is not surprising that additional oocyte specializations are present. These include the completion of meiosis before fertilization and devotion of large quantities of transcripts to proteins involved in the physical block to polyspermy, which is reflective of this animal’s reproductive strategy of broadcast spawning.

The time required to grow and differentiate a population of oocytes depends on the species, but anywhere from 1 to 3 months is frequently cited (e.g. Harvey, 1956; Walker et al., 2005). During its development, the structure of the oocyte changes dramatically from a mitotically replicating stem cell of 10–15 μm in diameter, to one 10 times greater in size (Walker et al., 2005). Many of the events that result in an oocyte’s growth occur during discrete developmental transitions, such as waves of specific gene activity and synthesis of organelles specialized for the egg. The major developmental changes include: (1) transition from a mitotic stem cell to a meiotically committed egg precursor. (2) A transition in vitellogenesis, from a period of slow growth and organelle replication to a period of rapid nutrient incorporation, storage of glycogen and lipid, and accumulation of yolk. (3) Completion of meiotic divisions, which results in the egg’s haploid genome and parallels a wholesale change in the mRNA composition and translocation of the cortical granules to the egg cortex. Finally, (4)
fertilization, in which a calcium wave initiates from the point of sperm fusion (see Roux et al., 2006) and quickly activates multiple physiological changes in the egg. These include the exocytosis of up to 15,000 cortical granules whose contents merge with the extracellular vitelline layer to form the fertilization envelope, which provides the physical block to polyspermy (Fig. 1).

Here, we survey transcription factors responsible for the unique transcriptional activity of the oocyte. We also examine the genes involved in meiosis, nutrient storage, and fertilization (see Fig. 2). Some of these oocyte-specific genes cluster within genomic loci, and a subset of the genes enlisted for fertilization are organized in a modular fashion—specifically, with stable protein domains encoded on swappable exons whereas diversifying short, tandem repeats are confined to single exons. Together, this genomic organization suggests a regulatory system that enhances proper timing of expression while favoring the divergence of egg-specific genes.

Methods

Gene identification

All genes discussed in this manuscript were identified and analyzed using the Sea Urchin Genome site (http://www.hgsc.bcm.tmc.edu/projects/seaurchin/), the Sea Urchin Annotation site (http://annotation.hgsc.bcm.tmc.edu/Urchin/) and the Genebore site (www.genboree.org), and are tabulated separately (Table 1). When available, experimentally-determined expression profiles beyond the tiling data (Samanta et al., 2006) are also listed with their original citations.

Animals

Strongylocentrotus purpuratus were obtained from Charles Hollahan (Santa Barbara, CA) and housed in aquaria cooled to 16°C in artificial sea water (ASW; Coral Life Scientific Grade Marine Salt; Energy Savers Unlimited, Inc, Carson, CA). Females were shed by intracoelomic KCl (0.5 M) injection, and oocytes were isolated as described (Wessel et al., 2004). Sea urchin embryos were cultured at 16°C and collected at necessary developmental stages.

In situ RNA hybridization

Whole-mount in situ RNA hybridizations were performed as previously described (Minokawa et al., 2004). cDNA products generated from the RT-PCR reactions were cloned into pGEM-T EASY vectors (Promega Corporation, Madison, WI). Linearized plasmids were in vitro transcribed, and the antisense probes were labeled with digoxigenin using the DIG RNA Labeling Kit (Roche Applied Science, Indianapolis, IN). Negative controls were transcribed off plasmid pSPT18-Neo or pSPT19-Neo provided in the DIG RNA Labeling Kit.

Microscopy

Electron microscopic images of the cortical granule and fertilization envelopes were prepared as previously described (Wong and Wessel, 2004). In situ RNA hybridizations were photographed on a Zeiss Axiovert 200 M equipped with a Zeiss color AxioCam MR5 camera (Carl Zeiss Incorporated, Thornwood, NY) or Zeiss Axioplan equipped with a Hamamatsu ORCA CCD digital camera (Advanced Microscopy Technology Corporation software, Danvers, MA).

Fig. 1. Montage of oogenesis and fertilization. Micrographs include: various stages of oogenesis and early development, especially early oocytes (A), the mature egg and fertilized zygote (C). Electron micrographs of the egg cortex (B) and zygotic fertilization envelope (E) are shown for more detail. (D) Photos of S. purpuratus female, showing the eggs shed on the test (lower left) and the coelomic cavity containing ovaries (upper right) (adapted from Wessel et al., 2004).
Phylogenetic analysis

All domains were identified using the Simple Modular Architecture Research Tool (SMART) site (smart.embl-heidelberg.de/). Nucleotide sequences from the YP30 genes or amino acid sequences of the peroxidase genes were aligned using Clustal algorithms within MacVector (Accelrys, Burlington, MA). These alignments were analyzed in PAUP (Swofford, 2002) using parsimony and/or maximum likelihood methods to establish their relationships. Bootstrap scores were determined from 1000 or 100 reiterations for the parsimony or the maximum likelihood method, respectively. These data are represented as phylograms with appropriate distance scales.

Reverse Transcriptase-PCR (RT-PCR)

Embryos and adult tissues were collected, and total RNA was extracted as described previously (Bruskin et al., 1981). RT-PCR of YP30 genes was performed according to the manufacturer’s directions using the Access RT-PCR kit (Promega Corporation, Madison, WI). Primers are listed in Table S1 in the 5’ to 3’ direction. For each sample, a no-RT control was used and indicated no DNA template contamination (data not shown). All primer sets span an intron, except for YP30-1A.1 and GAPDH. The reverse transcription reaction was performed for 45 min at 48°C, followed by denaturation for 2 min at 94°C. PCR amplification was performed for 40 cycles at 94°C for 1 min, annealing for 1 min at 60°C, and extension for 1 min at 68°C. RT-PCR products were run on 1.5% agarose electrophoretic gels and fragments were gel purified using QiaQuick spin columns (Qiagen Inc., Valencia, CA) and cloned into pGEM-T EASY vector (Promega Corporation, Madison, WI) for nucleotide sequencing to test the authenticity of each amplification reaction.

Quantitative, real-time PCR (Q-PCR)

Oocytes of different stages were released and collected from ovaries treated with 10 μg/mL collagenase (Sigma-Aldrich Corporation, St. Louis, MO, USA) at 16°C for 4 h. Total RNA from 1000 oocytes was extracted with RNeasy Micro Kit (Qiagen Inc., Valencia, CA). cDNA was prepared from 1000 oocytes or 2 μg of total RNA from embryos and adult tissues by reverse transcription-PCR (TaqMan Reverse Transcription Reagents Kit, Foster City, CA). Q-PCR was performed on the 7300 Real-Time PCR system (Applied Biosystems, Foster City, CA) using SYBR Green chemistry (Applied Biosystems, Foster City, CA). Primer sets were designed to

Fig. 2. Overview of oocyte development and gene expression profiles enhanced during sea urchin oogenesis. (A) Major stages of oogenesis. Dashed circles represent nuclei, either germinal vesicles or pronucleus (as labeled). Polar bodies and their nuclei (grey) are degraded by somatic cells within the ovary. Note, the first polar body does not appear to divide further in sea urchins (Nakashima and Kato, 2001; Walker et al., 2005). (B) Chromatin status within the nucleus of a developing oocyte. Meiotic phases are shown in parallel with the appropriate oocyte stage. Nuclei that segregate with the polar bodies are highlighted as in ‘A’ (grey). Inset shows a detail of DNA synapsis at a chiasma during meiotic DNA recombination. (C) Timing of meiosis-specific gene expression during oogenesis. Genes found to be specifically involved with meiosis in other animals, and their approximate window of expression, are shown (grey). S. purpuratus orthologs identified in the genome are highlighted (lavender), and those expressed in ovary and/or testis are indicated (purple). DNA recombination=meiotic DNA recombination; MC=meiotic complex (D) List of genes expressed primarily in the later stages of oogenesis within the adult sea urchin. Members are grouped according to the localization of their encoded proteins. Note that the major yolk protein (MYP) is predominantly expressed in the gut during oogenesis, and is transported to the developing oocyte (see Brooks and Wessel, 2003b. Embryonic expression of particular genes is noted in the “egg” stage (arrowheads, representing future time points); see Table 1 for references.
### Table 1
Genes related to oocyte biology

<table>
<thead>
<tr>
<th><strong>S. purpuratus gene</strong></th>
<th><strong>Gene family [Function]</strong></th>
<th><strong>Glean ID</strong></th>
<th><strong>RNA</strong></th>
<th><strong>Protein</strong></th>
<th><strong>References</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Meiosis-specific genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAG3/Sec3</td>
<td>Scc3/Rec11 homolog [Core cohesin arm scaffold]</td>
<td>SPU_17052</td>
<td></td>
<td></td>
<td>(Marston and Amon, 2004)</td>
</tr>
<tr>
<td>Rec8</td>
<td>Cohesin Rad21 homolog [DSB repair; cohesin subunit]</td>
<td>SPU_02553</td>
<td></td>
<td></td>
<td>(Marston and Amon, 2004)</td>
</tr>
<tr>
<td>Shugoshin 1 (Sgo1)</td>
<td>Cohesin [Protects centromere cohesin]</td>
<td>None identified</td>
<td></td>
<td></td>
<td>(Marston and Amon, 2004)</td>
</tr>
<tr>
<td>Structural Maintenance of Chromosome-like 1 beta (SMC1β)</td>
<td>Cohesin [Microtubule attachment]</td>
<td>SPU_21629</td>
<td></td>
<td></td>
<td>(Craig and Choo, 2004)</td>
</tr>
<tr>
<td>Spo13</td>
<td>Monopolar Cohesin [Centromere cohesin subunit]</td>
<td>None identified</td>
<td></td>
<td></td>
<td>(Marston and Amon, 2004)</td>
</tr>
<tr>
<td>Spo11 (meiosis-defective 1)</td>
<td>Type II topoisomerase [Initiator of DSB in meiosis]</td>
<td>SPU_17861 Testis (RT)</td>
<td></td>
<td></td>
<td>(Marston and Amon, 2004)</td>
</tr>
<tr>
<td>Mei1 (meiosis-defective 1)</td>
<td>Meiosis defective genes [Synapse formation]</td>
<td>SPU_23500 Ovary (RT)</td>
<td></td>
<td></td>
<td>(Libby et al., 2003)</td>
</tr>
<tr>
<td>Rad51</td>
<td>RecA homolog [Strand invasion]</td>
<td>SPU_09590</td>
<td></td>
<td></td>
<td>(Marston and Amon, 2004)</td>
</tr>
<tr>
<td>Rad51c</td>
<td>RecA homolog [Chromosome pairing]</td>
<td>SPU_22221</td>
<td></td>
<td></td>
<td>(Pawlowski and Cande, 2005)</td>
</tr>
<tr>
<td>Disrupted Meiotic cDNA 1 (DMC1)</td>
<td>RecA homolog [Strand invasion (non-sister)]</td>
<td>SPU_27921</td>
<td></td>
<td></td>
<td>(Marcon and Moens, 2005)</td>
</tr>
<tr>
<td>Msh4</td>
<td>MutS homolog [Repairs synaptic DSBs]</td>
<td>SPU_25763 Ovary (RT)</td>
<td></td>
<td></td>
<td>(Kolas and Cohen, 2004)</td>
</tr>
<tr>
<td>Msh5</td>
<td>MutS homolog [Promotes crossing over]</td>
<td>SPU_04494</td>
<td></td>
<td></td>
<td>(Kolas and Cohen, 2004)</td>
</tr>
<tr>
<td>Him-3 (or Hop1) Red1 and Mek1</td>
<td>Synaptonemal complex [Chromosome nucleation]</td>
<td>None identified</td>
<td></td>
<td></td>
<td>(Pawlowski and Cande, 2005)</td>
</tr>
<tr>
<td>Sycp-3</td>
<td>Synaptonemal complex [Lateral element subunits]</td>
<td>SPU_08527</td>
<td></td>
<td></td>
<td>(Page and Hawley, 2004)</td>
</tr>
<tr>
<td>Sycp-1</td>
<td>Synaptonemal complex [Transverse filament subunit]</td>
<td>SPU_21319</td>
<td></td>
<td></td>
<td>(Page and Hawley, 2004)</td>
</tr>
<tr>
<td>Mam1</td>
<td>Monopolar Complex [Mono-orients kinetochore]</td>
<td>None identified</td>
<td></td>
<td></td>
<td>(Marston and Amon, 2004)</td>
</tr>
<tr>
<td>Mei-1 and Mei-2 katanin ATPase</td>
<td>[Organizes meiotic spindle]</td>
<td>None identified</td>
<td></td>
<td></td>
<td>(Yang et al., 2003)</td>
</tr>
</tbody>
</table>

**Nutrition and storage**

<table>
<thead>
<tr>
<th><strong>Vitellogenin #1</strong></th>
<th>Vitellogenin</th>
<th>SPU_09378</th>
<th>Ovary (RT)</th>
<th>Oocyte (EM)</th>
<th>(Wessel et al., 2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vitellogenin #2</strong></td>
<td>Vitellogenin</td>
<td>SPU_16052</td>
<td>Egg (RT)</td>
<td>Egg (blot)</td>
<td>(Brooks and Wessel, 2002, 2003a,b)</td>
</tr>
<tr>
<td><strong>Vitellogenin #3</strong></td>
<td>Vitellogenin</td>
<td>SPU_28683</td>
<td>Testis (RT)</td>
<td>Embryo (blot)</td>
<td>Song and Wessel (unpublished)</td>
</tr>
<tr>
<td><strong>Major yolk protein (MYP)</strong></td>
<td>Major yolk protein</td>
<td>SPU_13301</td>
<td>Ovary (RT)</td>
<td>Egg (RT)</td>
<td>(Wessel et al., 2000)</td>
</tr>
<tr>
<td><strong>Melanotransferrin</strong></td>
<td>Iron binding and transport</td>
<td>SPU_26949</td>
<td>Ovary (RT)</td>
<td>Egg (RT)</td>
<td>(Wessel et al., 2000)</td>
</tr>
<tr>
<td><strong>Ferritin H-chain</strong></td>
<td>Iron binding (both heavy and light chains)</td>
<td>SPU_24366 SPU_04876 SPU_05198</td>
<td>Ovary (IS, RT, Q) Egg (IS, RT, Q) Embryos (RT, Q)</td>
<td>Ovary (RT) Egg (RT)</td>
<td>(Wessel et al., 2000)</td>
</tr>
<tr>
<td><strong>YP30-1A.1</strong></td>
<td>YP30</td>
<td>SPU_04876 SPU_05198</td>
<td>Ovary (IS, RT, Q) Egg (IS, RT, Q) Embryos (RT, Q)</td>
<td>Ovary (RT) Egg (RT) embryo (RT)</td>
<td>(Wessel et al., 2000)</td>
</tr>
<tr>
<td><strong>YP30-2A.1</strong></td>
<td>YP30</td>
<td>SPU_06345</td>
<td>Ovary (RT) Egg (RT) embryo (RT)</td>
<td>(Wessel et al., 2000)</td>
<td></td>
</tr>
<tr>
<td>S. purpuratus gene</td>
<td>Gene family [Function]</td>
<td>Glean ID</td>
<td>RNA</td>
<td>Protein</td>
<td>References</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------</td>
<td>----------</td>
<td>-----</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>YP30-2A.2 SPU_06346</td>
<td>YP30</td>
<td>Ovary (IS, RT, Q) egg (IS, RT, Q) embryos (RT, Q)</td>
<td>Oocyte (blot, X, EM) egg (blot, X, EM) embryo (blot, X)</td>
<td>(Wessel et al., 2000)</td>
<td></td>
</tr>
<tr>
<td>YP30-2B.1 SPU_25212</td>
<td>YP30</td>
<td>Ovary (blot, IS, RT) egg (RT) embryo (RT)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YP30-2B.2 SPU_10203</td>
<td>YP30</td>
<td>Ovary (blot, IS, RT, Q) egg (IS, RT, Q) embryos (RT, Q)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YP30-2B.3 SPU_16866</td>
<td>YP30</td>
<td>Ovary (blot, IS, RT, Q) egg (RT, Q) embryos (RT, Q)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YP30-3A.1 SPU_03676</td>
<td>YP30</td>
<td>Ovary (blot, IS, RT, Q) egg (IS, RT, Q) embryos (IS, RT, Q)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YP30-3A.2 SPU_20485</td>
<td>YP30</td>
<td>Ovary (RT) egg (RT) embryos (RT)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YP30-3A.3 SPU_03678</td>
<td>YP30</td>
<td>Ovary (RT) egg (RT) embryos (RT)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YP30-4A.1 SPU_15670</td>
<td>YP30</td>
<td>Ovary (RT) egg (RT) embryos (RT)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fertilization

<table>
<thead>
<tr>
<th>Gene</th>
<th>Glean ID</th>
<th>RNA</th>
<th>Protein</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ovoperoxidase CG enzyme</td>
<td>SPU_16914</td>
<td>Ovary (blot, IS, Q) egg (Q)</td>
<td>Oocyte (blot, X) egg (enz, blot, X, WM) FE (enz, blot, X, WM)</td>
<td>(Deits et al., 1984)</td>
</tr>
<tr>
<td>(LaFleur et al., 1998)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cortical granule serine protease (CGSP1) CG enzyme</td>
<td>SPU_19872 SPU_04123 SPU_16163</td>
<td>Ovary (RT)</td>
<td>Egg (enz, blot, X, WM)</td>
<td>(Haley and Wessel, 1999)</td>
</tr>
<tr>
<td>(Haley and Wessel, 2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-glucanase A</td>
<td>SPU_06529</td>
<td>Ovary (blot, RT) egg (blot, RT) embryo (blot)</td>
<td>Egg (enz, blot)</td>
<td>(Glabe et al., 1983; Kopf et al., 1982)</td>
</tr>
<tr>
<td>(Bachman and McClay, 1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-glucanase B</td>
<td>SPU_24075 SPU_24336 SPU_24337 SPU_21062 SPU_00513</td>
<td>Ovary (RT) egg (RT) embryo (RT)</td>
<td>Oocyte (blot, X) egg (enz, blot, X, WM) embryo (blot, X, WM)</td>
<td>(Battaglia and Shapiro, 1988)</td>
</tr>
<tr>
<td>(Wong and Wessel (unpublished)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-glucanase C vertebrate transglutaminase FE enzyme</td>
<td>SPU_25507 SPU_00512</td>
<td>Ovary (RT) egg (RT) embryo (RT)</td>
<td>Oocyte (blot, X) egg (enz, blot, X, WM) embryo (enz, blot, WM)</td>
<td>(Wong and Wessel, 2005)</td>
</tr>
<tr>
<td>Urchin dual oxidase 1 (Udx1) FE enzyme</td>
<td>SPU_28103</td>
<td>Ovary (RT) egg (RT) embryo (RT)</td>
<td>Egg (X, EM) FE (blot, X, WM, EM)</td>
<td>(Somers et al., 1989; Weidman et al., 1985)</td>
</tr>
<tr>
<td>(Wong and Wessel, 2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteolaiasin FE structural protein</td>
<td>SPU_20885</td>
<td>Ovary (RT) egg (RT) embryo (IS)</td>
<td>Oocyte (blot, X) egg (blot, X, WM) embryo (blot, X, WM)</td>
<td>(Haley and Wessel, 2004)</td>
</tr>
<tr>
<td>(Haley and Wessel (unpublished)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFE-1 FE structural protein</td>
<td>SPU_24925</td>
<td>Ovary (blot, IS, RT') egg (blot, X, WM) embryo (blot, X, WM)</td>
<td>Oocyte (X, EM) egg (blot, X, WM) embryo (blot, X, WM)</td>
<td>(Laidlaw and Wessel, 1994)</td>
</tr>
<tr>
<td>(Laidlaw and Wessel, 2000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFE-9 FE structural protein</td>
<td>SPU_17329</td>
<td>Ovary (blot, IS, RT) egg (X, EM)</td>
<td>Oocyte (blot, X) egg (blot, X, WM) embryo (blot, X, WM)</td>
<td>(Wong and Wessel, 2004)</td>
</tr>
<tr>
<td>Rendezvin Egg vitelline layer protein and FE structural protein SPU_19369</td>
<td>Ovary (blot, IS, RT) egg (blot, IS) embryo (IS)</td>
<td>Oocyte (blot, X) egg (blot, X, WM) embryo (blot, X, WM)</td>
<td>(Wessel et al., 1998, 2000)</td>
<td></td>
</tr>
<tr>
<td>(Wessel and Wessel (in press)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyalin Embryonic extracellular matrix protein SPU_01928</td>
<td>Ovary (blot, IS, RT') egg (blot, IS) embryo (blot, IS)</td>
<td>Oocyte (blot, X) egg (blot, X, WM) embryo (blot, X, WM)</td>
<td>(Haley and Wessel, 2004)</td>
<td></td>
</tr>
<tr>
<td>(Haley and Wessel (unpublished)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p160 Egg vitelline layer protein SPU_20885</td>
<td>Ovary (IS, RT') egg (IS) embryo (IS)</td>
<td>Oocyte (blot, X) egg (blot, X, WM) embryo (blot, X, WM)</td>
<td>(Haley and Wessel, 2004)</td>
<td></td>
</tr>
<tr>
<td>(Haley and Wessel (unpublished)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued on next page)
amplify products of 100–180 bp; sets for the YP30 genes are listed in Table S1 whereas sets for transcription factors are listed at http://vanbeneden.caltech.edu/~mlhoward/sutf/ and http://issola.caltech.edu/~materna/znfn/. Forkhead primer sets can be found in Tu et al. (2006). Reactions were run in triplicate, and repeated in at least 2 separate experiments. Q-PCR data were normalized against ubiquitin mRNA, and are presented relative to the egg levels.

Results and discussion

Transcription factors

The oocyte synthesizes several unique organelles and gene products that enable it to achieve the developmental transitions...
and specializations associated with fertilization and embryo-gensis. Some oocyte-specific transcripts accumulate to high abundance (e.g. SpYP30 species 0.05% of total egg mRNA; JL Song, data not shown). The transcription of these genes ceases at germinal vesicle breakdown – coordinate with the reductive divisions of meiosis – and these genes are not expressed again until the next generation of oocytes. Transcription factors expressed during oogenesis may be expected to have different, but overlapping, profiles compared to embryos and/or other adult tissues. We focused on factors that have putative binding sites within several oocyte gene promoters identified bioinformatically (data not shown), and factors that are in very low abundance or are completely absent in embryos (Howard-Ashby et al., 2006, Materna et al., and Tu et al., 2006). We tested the expression of these factors in oocytes, ovary, testis, and gut from Stronglyocentrotus purpuratus. The transcription factors examined span the major classes of regulatory proteins, including the forkhead, homeobox, zinc-finger, bHLH, and nuclear receptor proteins (Table S2).

We have not found any transcription factors exclusive to one tissue, although several of the factors are remarkably enriched in the gonads (e.g. the forkhead factor FoxL2), in the gut (DsF), or in oocytes (FoxX) (Fig. 3). One candidate factor (Eyg, an ETS factor) is enriched in extra-oogonial cells of the ovary, and thus may be a good marker for somatic cells such as the nutritive phagocytes.

Of the 22 forkhead transcription factors (Fox) identified in S. purpuratus (see Tu et al., 2006), 14 were tested in this study. This class of factors was emphasized, since promoters of oocyte-specific genes (e.g. YP30 and ovoperoxidase) contain multiple, putative forkhead protein binding sites (data not shown). Messenger RNAs for most of these Fox transcription factors are enriched in specific tissues, except for foxJ2, which is evenly expressed in all tissues (Fig. 3).

Fig. 3 (continued).
pathways (Guigon et al., 2005; Ottolenghi et al., 2005). The remaining Fox factors are expressed in a sex-specific manner: foxO, foxJ1, and foxM are enriched in the testis, whereas foxQ2, foxN1/4, and foxX are enriched in oocytes and ovary.

Twelve percent of the transcription factors tested (10/80) have mRNAs enriched in the oocyte and/or egg. Of these, smad1, dsc, pitx3, atonal, and lmo4, are also expressed later in embryogenesis (http://vanbenenden.caltech.edu/~milHoward/sutf/ and http://issola.caltech.edu/~materna/znfn/), suggesting that they play regulatory roles in both oocyte development and embryogenesis. Of particular interest is the oocyte-enriched expression of dsn (a.k.a. dmrt in zebrafish and mab-3 in C. elegans), which is involved in sex determination in several animals (Guo et al., 2004; Pask et al., 2003; Winkler et al., 2004) (see also Juliano et al., this issue). Its re-expression later in embryogenesis may coincide with germ cell development.

While we have identified many transcription factors whose expression is enriched in the oocyte and ovary, we note the absence of orthologs of several factors that are important for murine oocyte and ovarian development, such as Ftgα, the homeobox factors Obox, and Nobox (Rajkovic et al., 2004; Rajkovic et al., 2002; Soyal et al., 2000; Suzumori et al., 2002). Ftgα is a germ cell-specific bHLH factor that regulates the coordinated expression of oocyte-specific zona pellucida (ZP) paralogs ZPA, ZPB, and ZPC (see Wong and Wessel, 2006 for review of ZP proteins). Many genes in the sea urchin have recognizable ZP domains, including one on the cell surface of the sea urchin egg (L. Varghese and ML Leguia, unpublished observation), but none of the ZP-containing proteins predicted in the S. purpuratus genome show similarity to the mammalian oocyte zona pellucida orthologs. The obox genes are preferentially expressed in the gonads, but their role in gonadal development is not known. NOBOX (newborn ovary homeobox), on the other hand, is an oocyte-specific factor essential for early folliculogenesis and, like Ftgα, is critical for specifying oocyte gene expression patterns in mice (Rajkovic et al., 2004). The absence of these orthologs in S. purpuratus may reflect evolutionary divergence of oocyte developmental mechanisms among deuterostomes. For example, unlike mammalian oocytes that share intimate cytoplasmic continuity with their surrounding follicle cells, the sea urchin oocyte develops within nutrient-rich chambers established by specialized somatic epithelial cells called nutritive phagocytes (Walker et al., 2005). A clear, ultrastructural distinction is present between the germ and somatic cells within this chamber, and oocytes instead form pseudopodia that contact the nutritive phagocytes, suggesting that a true intercellular communication mechanism functions during oogenesis (Walker et al., 2005).

Oocytes grow and differentiate in response to signals from the environment and from ovarian somatic cells (reviewed in Eppig, 2001; Walker et al., 2005). The Wnt signaling pathway has been implicated in mammalian ovarian development, oogenesis, and early development (Hsieh et al., 2005; Jeays-Ward et al., 2004; Vainio et al., 1999). Wnt4 is required for gonadal differentiation and maintenance of the female germ line in mice; animals null for Wnt4 have masculinized gonads (Jeays-Ward et al., 2004; Vainio et al., 1999). In addition, mice null for frizzled-4, one of the Wnt signaling receptors, are infertile (Hsieh et al., 2005). Many factors involved in the Wnt signaling pathways are present in the sea urchin (Crocé et al., this issue); enrichment of several target genes of the Wnt pathway, such as foxN1/4, groucho, and atonal (Fig. 3) in the sea urchin oocyte suggests that the Wnt signaling pathway may be important for the growth of oocytes in response to signals released by the environment and/or nutritive phagocytes.

The TGF-β signaling pathway is also important for proper development of the mammalian gonad, as demonstrated by various mouse knockouts. Mammalian oocytes develop within ovarian follicles, and the exchange of growth factors and signaling molecules is important for the overall development of the ovary (reviewed in (Eppig, 2001; Matzuk et al., 2002). TGF-β superfamily members, such as GDF9 and BMP15, are oocyte-secreted growth factors crucial for the differentiation and development of mammalian ovarian follicles (Dong et al., 1996; Juengel et al., 2002; Yan et al., 2001). Mice lacking gdf9 are infertile because their follicular development arrests at the primary stage (Dong et al., 1996). Although not as severe as the gdf9 knockout phenotype, depletion of BMP15 in mice results in a reduction in the number of oocytes, and has negative effects on fertility (Yan et al., 2001). BMP15 and GDF9 act synergistically in mice to ensure proper development of the ovarian follicle (Su et al., 2004; Yan et al., 2001). Ligand binding to TGF-β receptors can activate transcription factors of the Smad family and Smad-interacting proteins, including Smad1 and Sp1, SpSmadIP/zfh1, and FoxO (see Lapraz et al., 2004; reviewed in (Mazerbourg and Hsueh, 2003). smad 1 and 5 knockout mice each exhibit defective primordial germ cell development, indicating that their activity is crucial for proper folliculogenesis (Tremblay et al., 2001; Weinstein et al., 2000); reviewed in (Kaivo-Oja et al., 2006). Transcripts of sea urchin homologs to smad and smad-interacting transcription factors are also enriched in the oocyte and/or ovary (Fig. 3), suggesting that TGF-β signaling pathways may be used to communicate between the developing oocyte and its surroundings.

Overall we find that mRNA of a diverse family of regulatory proteins is present across many adult tissues (Fig. 3). Levels of transcripts are wildly variable, from an estimated 35,000 copies per egg for foxN1/4 to barely detectable for other transcription factors (http://vanbenenden.caltech.edu/~milHoward/sutf/ and http://issola.caltech.edu/~materna/znfn/). Many examples of translational regulation are known in this animal, and transcript abundance is difficult to interpret functionally without knowledge of their cognate protein abundance and activity. However, an examination of the RNA levels of these transcription factors in the oocyte and some adult tissues enables us to identify candidates specifically involved in oogenesis.

**Meiotic and DNA recombination genes**

Only oocytes and spermatocytes undergo the highly conserved process of meiosis. We do not know when during development germ cells commit to enter meiosis in the sea urchin, but presumably this transition occurs well after
metamorphosis. This premise is based simply on the numbers of stem cells necessary to produce the vast population of gametes this animal makes compared to the size of the gonadal rudiment shortly after metamorphosis. It is also clear that continual germline stem cell replication occurs in adults, especially when the seasonal-dependence of gametogenesis is considered (Walker et al., 2005). For example, during inter-gametogenesis (between gravid seasons), most oogonia are amitotic whereas peak gametogenesis periods (the gravid season) corresponds to the highest number of mitotic germ cells, located adjacent to the ovarian capsule (Walker et al., 2005). This propagation of oogonia during the peak spawning season presumably serves both to replenish the population of stem germ cells for the next season and to expand the number of cells progressing through oogenesis.

Hallmarks of meiosis include a single round of DNA replication, followed by two rounds of cell division. During this process, sister chromatids require special handling to facilitate homologous DNA recombination and proper segregation into haploid daughter cells. In oocytes, these events occur in temporally distinct phases that span the duration of oogenesis: DNA replication results in a primary oocyte with its germinall vesicle (4N genome); homologous recombination progresses during the long-lived growing oocyte period that occupies prophase I; and chromosomes separate during anaphases I and II, which correlates with the final maturation of an oocyte to an egg—demarcated in sea urchins by germinal vesicle breakdown (Nakashima and Kato, 2001). Here, we briefly review and summarize the players involved in key meiotic processes in metazoans. Orthologs of genes encoding many of these players are in the S. purpuratus genome, but some important regulators have yet to be identified (Fig. 2; Table 1).

The process of meiosis is distinct from mitosis in how the replicated chromosomes are segregated (reviewed in Craig and Choo, 2005; Marston and Amon, 2004; Pawlowski and Cande, 2005). The initial phase of meiosis is similar to mitosis: DNA synthesis (S phase) is associated with maintenance of sister chromatid pairing through cohesin complexes. These multimeric structures are in the elements and transverse filaments of a synaptonemal complex generally proceeds through sequential recruitment of mechanistic subunits to the chromosome, and may be coordinated with the help of cohesin subunits (Page and Hawley, 2004). The lateral elements and transverse filaments of a synaptonemal complex are assembled in parallel with the progression of meiotic DNA recombination and synapse formation. These structures are composed of a variety of proteins that all contain a central coiled-coil domain for aggregation, but otherwise vary significantly in sequence across phyla. For example, yeast HOP-1, RED1, and MEK1 assemble onto chromosomes first, followed by ZIP-2 and ZIP-3 multimers. Together, these proteins line the chromosome to form lateral elements that can be pulled together (“zipped”) using ZIP-1 polymers as transverse filaments. In nematodes, the HOP1 analog is HIM-3, a sentinel protein that
senses and identifies sites of Spo11-induced DNA double strand breaks, while transverse filaments are established by SYP-1 and SYP-2 heterodimers. Mammals substitute SYCP-2 (or SCP-2) and SYCP-3 to make lateral elements while SYCP-1 is used for transverse filaments (Page and Hawley, 2004). The presence of SYCP orthologs in the genome suggests that *S. purpuratus* utilizes a synaptonemal complex similar to those found in mammals (Fig. 2, Table 1).

As meiotic DNA double strand breaks are repaired, the transient synaptonemal complexes are disassembled. By late prophase I (e.g. approaching germinal vesicle breakdown in the oocyte), however, any DNA crossovers retained on homologous chromosomes are stabilized by synaptonemal complex proteins to maintain the proper pairing at metaphase I—a stage when mono-orientation of kinetochores could otherwise incorrectly segregate the chromosomes (Marston and Amon, 2004; Page and Hawley, 2004). This restricted pairing is aided by the monopolin complex, an assembly of proteins that protects one of the sister chromatids from microtubule attachment during meiosis I, thereby maintaining their pairing during anaphase I. The monopolin complex consists of subunits normally used during mitotic spindle formation as well as meiosis-specific components that build onto a cohesin-like foundation, such as the centromeric Spo13 (Marston and Amon, 2004). Yeast MAM1 is another member of the monopolin complex that functions by suppressing the formation of bi-oriented kinetochores during meiosis I (Marston and Amon, 2004). None of these monopolin complex subunits have been identified yet in *S. purpuratus*.

Progression through meiosis I requires completion of homologous recombination, followed by the loss of all arm cohesins (Craig and Choo, 2005), and finally by separation of sister chromatids during anaphase I (Fig. 2). A second round of division without DNA synthesis or recombination then occurs to complete meiosis II, resulting in haploid gametes. In the second round of chromosome separation, the centromeric cohesins are cleaved to enable bi-oriented kinetochores to completely separate the sister chromatids. In sea urchin oocytes, the asymmetric divisions of meiosis I and II generate a single egg and two unequal polar bodies (the polar body from meiosis I fails to undergo meiosis II; Nakashima and Kato, 2001; Walker et al., 2005) (Fig. 2A). This is distinct from the symmetric divisions of spermatogenesis that result in four equal-sized gametes. Thus, it is not surprising that spindle formation and orientation possess sexually dimorphic mechanisms. One female-specific requirement is the oocyte-specific activity of a katanin homolog consisting of MEI-1 (meiotic spindle formation proteins-distinct from mei1 (meiosis defective 1)), a microtubule-severing ATPase subunit, and its co-conspirator MEI-2 (Clark-Maguire and Mains, 1994a,b; Srayko et al., 2000). Together, the MEI-1/2 heterodimer functions along β-tubulin B-isoform 2-enriched meiotic spindles, where it organizes and translocates the spindle to the oocyte cortex to facilitate chromosomal segregation into the polar body (Lu et al., 2004; Srayko et al., 2000; Yang et al., 2003). Nematode MEI-1/2 was initially identified as a homolog to sea urchin egg katanin based on regions of high sequence identity across the proteins (Clark-Maguire and Mains, 1994b; Hartman et al., 1998; Srayko et al., 2000), but unlike in other metazoans, no homologs have been predicted in the genome beyond the original *S. purpuratus* katanin (p60 subunit A=SPU_01000 {AAC15706}; p80 subunit B=SPU_01360 and 14392 {AAC09329}). Perhaps this original katanin participates in spindle assembly during both meiosis and mitosis. This is possible in sea urchins because oogenesis and meiosis are completed in the ovary well before fertilization (Dale et al., 1999). This temporal separation obviates the mechanisms necessary to distinguish between meiotic and mitotic spindle formation (Srayko et al., 2000; Yang et al., 2003)—a process required by other animals that originate from eggs fertilized during meiosis (Dale et al., 1999; Stricker, 1999; Voronina et al., 2003).

Our current level of understanding of meiosis in the sea urchin is limited to a list of putative *S. purpuratus* orthologs predicted from the genome assemblies (Table 1). Most of these genes are involved with meiotic cohesion and DNA recombination, implying that these mechanisms are conserved among metazoans. Our identification of some genes in the sea urchin, such as the “vertebrate-specific” mei1 (Libby et al., 2003; Reinholdt and Schimenti, 2005), implies that the specialization of some meiotic genes originated prior to the split of deuterostomes from other animals. Conversely, the apparent absence of other genes found to be essential in eukaryotes, including sgo1 and spo13, suggests that sea urchins either do not require this set of genes if their expression can be timed appropriately, or may use alternative splicing of a gene shared with other somatic processes (such as DNA repair or mitosis) to accomplish the same goal. An example of the first scenario lies in the mitotic kinetochore checkpoint protein BUB1 (budding uninhibited by benimidazoles 1; SPU_04518) (Craig and Choo, 2005), a protein that may serve a meiotic role analogous to Sgo1 by regulating cohesin degrading enzymes along the chromosome arms during meiosis I, followed by a more general release during meiosis II. Like katanin, the clear temporal separation between meiosis and zygotic mitosis in the sea urchin allows for the overlapping use of *S. purpuratus* BUB1 in both processes without detrimentally affecting either.

Based on the seasonal cycling of gametogenesis in sea urchins, we believe the set of meiotic genes is regulated by environmental cues. Expression of meiotic genes may be controlled by the availability of nutrients in the environment in coordination with vitellogenesis (Walker et al., 2005), and/or may be stimulated by light cycles as perceived through photodetection devices (Raible et al., this issue). Perhaps some of the transcription factors enriched in the oocyte and testis (Fig. 3) are key regulators of the dynamic gonadal response to the environment.

**Vitellogenesis**

Accumulation and storage of nutrients (e.g. yolk) for embryogenesis is an essential function of the oocyte, and sea urchins invest enormous energy into this process. The primary role of these nutrients is to act as a reserve for early embryonic development, although they also appear to serve as currency for the adult. The nutritional state of the sea urchin dictates the
progress of gametogenesis (Walker et al., 2005). When the adult is in a nutrient rich environment, it builds and maintains large yolk reserves by expanding its oocyte and egg populations, but when food is scarce, the gametes appear to serve as the adult’s nutritional resource through reabsorption by nutritive phagocytes (Reunov et al., 2004a,b; Walker et al., 2005). During the pre-gametogenesis (summer in S. drobachiensis) and gametogenesis phases (autumn to winter), these nutritive phagocytes envelop developing gametes as incubation chambers, providing them the nutrients necessary to complete gametogenesis (Walker et al., 2005).

In females, yolk proteins are internalized by the oocytes only during a major transition in the vitellogenic phase during oogenesis (Harrington and Ozaki, 1986; Ozaki et al., 1986; Unuma et al., 1998). Yolk initially accumulated in vacuoles of the nutritive phagocytes is transferred to eggs wherein it is packaged into membrane-bound platelets, uniformly dispersed throughout the cytoplasm. Yolk comprises nearly one third of the volume of the egg and occupies more than 10–15% of the total egg protein (Harrington and Easton, 1982; Ichio et al., 1978; Kari and Rottman, 1980; Ozaki et al., 1986; Walker et al., 2005). The function of this large, vitellogenic resource has been controversial over the years, with several models proposed. The genome encodes three S. purpuratus vitellogenin homologs (Table 1) based on the positional conservation of a lipoprotein N-terminal domain (LPD_N), a motif common to all characterized animal vitellogenins (Schneider, 1996). Vitellogenin 1 is the paralog most closely related to the vertebrate vitellogenin, whereas vitellogenins 2 and 3 are homologs of apolipoprotein B (data not shown). No additional characterization of these putative vitellogenins has been described, leaving their functions unknown. On the other hand, the life history of the 180-kDa major yolk protein (MYP) isolated from yolk platelets is well understood (Brooks and Wessel, 2002,2003a,b; Harrington and Easton, 1982; Kari and Rottman, 1985; Shyu et al., 1986). MYP is encoded by a single gene containing 22 exons spanning over 20 kb. It belongs to the transferrin family of iron binding proteins, including the granulosa cells of the ovarian follicle (Briggs et al., 1999) and the Sertoli cells of the seminiferous tubules (Skinner and Griswold, 1980; reviewed in Sylvester and Griswold, 1994). The localized synthesis of transferrin in granulosa cells has been implicated in follicle maturation (Briggs et al., 1999), whereas male mice lacking normal transferrin have abnormal spermatogenesis and a decreased number of germ cells (reviewed in Sylvester and Griswold, 1994).

Consistent with mammalian transferrin expression, MYP may be a major regulator of iron homeostasis in the adult sea urchin, and could play an essential role in gametogenesis. MYP contains two transferrin-like iron-binding domains and chelates iron (Brooks and Wessel, 2002). The protein is largely expressed in the adult intestine (Shyu et al., 1986) and is transported via the coelomic fluid to be taken up by the ovarian nutritive phagocytes enveloping the oocyte (Ozaki et al., 1986; Unuma et al., 1998). MYP is ultimately transferred to the oocyte where it is endocytosed in a dynamin-dependent manner and stored in yolk platelets (Brooks and Wessel, 2003b). Like its mammalian homologs, MYP is equally abundant in testes (Brooks and Wessel, 2002; Harrington and Easton, 1982; Harrington and Ozaki, 1986; Kari and Rottman, 1985; Ozaki et al., 1986), suggesting that its primary function in the adult is the trafficking of iron to the gametes—possibly for use in assembly of heme-containing enzymes such as oocyte peroxidases and mitochondrial cytochromes.

MYP is proteolized during embryogenesis into characteristic cleavage products that disapper by the blastula stage (Armant et al., 1986; Kari and Rottman, 1985; Lee et al., 1989; Scott and Lennarz, 1989). Yolk granules are considered “lysosome-like” particles, because they contain typical acid hydrolases, including a cysteine proteinase effective in digesting yolk polypeptides (Fagotto, 1990a,b; Mallya et al., 1992). Two important factors for the regulation of yolk degradation appear to be pH and enzymatic latency (reviewed in Fagotto, 1995). The pH of the yolk granule is initially neutral, which is unfavorable for normal enzyme function. Regulated acidification of the yolk granules is the main trigger of proteolysis, a process that occurs during embryogenesis in several species such as the tick Ornithodoros moubata (Fagotto, 1990a), the cockroach Blattella germanica (Nordin et al., 1991), the frog Xenopus laevis (Fagotto and Maxfield, 1994a,b), and the sea urchin S. purpuratus and Lytechinus pictus (Mallya et al., 1992). Yolk degradation does not occur in the presence of bafilomycin, suggesting that acidification of the yolk granules is achieved with a vacuolar proton pump (Fagotto and Maxfield, 1994b). During sea urchin embryogenesis, lowering of yolk granule pH coincides with cysteine protease-mediated degradation of MYP (Mallya et al., 1992).

Various members of the cathepsin family of cysteine proteinases are thought to be involved in yolk proteolysis (reviewed in Carnevali et al., 2006), specifically vitellogenin degradation in tetrapods (De Stasio et al., 1999; Opresko and Karpf, 1987; Yamamura et al., 1995; Yoshizaki and Yonezawa, 1994), chicken oocytes (Retzek et al., 1992), and fish (Brooks et al., 2001; Sire et al., 1994). Annotation of the sea urchin genome revealed at least 14 cathepsins (http://annotation.hgsc.bcm.tmc.edu/Urchin/). One of the yolk cysteine proteinases was partially purified from sea urchin eggs (Okada and Yokota, 1990). This cathepsin B-like enzyme is present in all developmental stages from egg to larvae, but its enzymatic activity is only detectable at low pH, concomitant with acidification of the yolk granules during early embryogenesis. The significance of the proteolytic processing of MYP,
however, is unclear since this protein is not significantly depleted during early embryogenesis: all of its fragments remain long after gastrulation (Armant et al., 1986). Thus, in addition to functioning as an iron transporter, MYP may serve as an energy reserve to be used if food is scarce when feeding begins.

While a significant amount of endocytosis and transcytosis of MYP from the adult gut is necessary for enriching yolk organelles during vitellogenesis, a subset of yolk proteins is synthesized in situ by the oocyte. These oocyte-specific proteins are members of the YP30 family, a group of genes that encode proteins with two carboxyl-terminal fasciclin (FAS1) domains. The FAS1 domain was first identified in fasciclin 1, a Drosophila neural cell adhesion molecule consisting of four repeated FAS1 domains (Zinn et al., 1988). Generally, FAS1 domains are thought to represent ancient cell adhesion domains since its superfamily includes members from all phyla (Clout et al., 2003). FAS1 domains fold into a seven-stranded β wedge with several helices that may homopolymerize through polar interactions (Clout et al., 2003). Yeast 2-hybrid screens using *S. purpuratus* YP30 as the bait indicated its predominant interacting protein is another YP30 family member (Wessel et al., 2000). Yet, further analysis of some YP30 family members in protein–protein interactions indicated that YP30 dimerizes via its more divergent amino-terminal region, not via its conserved FAS1 domains (A Howell and ML Leguia, unpublished).

The YP30 family consists of 10 members that fall into four major clades (Fig. 4A). Most of the *S. purpuratus* YP30 members have similar gene structures (Fig. 4A), suggesting that gene duplications may have been maintained to enhance production of this accessory protein. The divergence in amino-terminal sequences between different clades suggests that some members have specialized functions. This is consistent with the differential expression profiles of the genes in each clade (Figs. 4B, C). For example, members of clades 1 and 2 have highest

---

**Fig. 4.** The YP30 family of fasciclin-containing proteins. (A) Parsimonous phylogeny of full length cDNA (left) and genomic organization of the ten *S. purpuratus* YP30 paralogs (right). Exons (boxes) and introns (carat, not to scale) are shown, mapped over the shared signal sequence (sig seq) and tandem fasciclin (FAS1) domains. See Table 1 for annotation SPU designations. Colors indicate members of each clade that were analyzed further as representatives for the expression of their respective clade. All bootstrap values are 100, as calculated from 1000 iterations. Grey scale bar embedded in the YP30-15670 branch equals 100 nucleotide changes. Expression of each member is shown by semi-quantitative (B) and quantitative (C) reverse transcriptase PCR detection in eggs, ovary, and embryos. Quantitative PCR results are normalized to ubiquitin, and shown relative to egg. Standard deviations of triplicates are all less than 1. (D) Maximum likelihood phylograms of cDNA encoding the tandem *fasciclin1* domain exons (#2–5; black) or introns (#2–4; grey) are overlaid to show incongruency of the exon–intron relationships. Analysis was performed over 100 bootstrap iterations. Branches point to their respective gene, as indicated by GLEAN3 assignment. Scale bar and dashed lines represent 100 nucleotide changes (dashed line is 3 times longer than solid bar). (E) Parsimony phylogram of the coding sequence for each YP30 fasciclin domain, as grouped by type (FAS1 A or B). Clades are indicated as in ‘A’ (grey). Differences in rooting at the distal nodes are highlighted (arrowheads). Bootstrap values are shown if less than 100, based on 1000 iterations. Bar and dashed lines represent 100 nucleotide changes. (F) Representative RNA in situ hybridizations of select YP30 members from each subclade in egg, ovary, and plutei. For all, colors correspond to YP30 paralog as in ‘A’.
expression in the egg and ovary (Figs. 4B, C) whereas members of clade 3 are expressed in highest abundance in larvae. Comparison of each of the conserved, tandem FAS1 domains distributes the YP30 members into the same four major clades (compare Fig. 4A with Fig. 4D). We find no evidence of YP30 FAS1 domain swapping between the FAS1 A and FAS1 B isoforms (Fig. 4E), implying that the paired FAS1 domains resulted from a single duplication event, and have been maintained in tandem since. Together, these phylogenies imply that the entire YP30 protein is diverging from its siblings rather than a more focused region of the protein.

Comparison of intron versus exon phylogenies encoded in the tandem FAS1 region shows incongruence in the divergence profiles of the individual members (Fig. 4D). While the final clustering of members within their respective clades (as determined by coding region) does not change, the divergent branching pattern of the members from a hypothetical ancestor differs between the intron and exon phylogenies. Although we cannot exclude possible divergence via the intron sequences, these differential paths of divergence likely reflect the selective pressures applied to the exons and may be related to specialized functions of individual members.

The exact function of YP30 is still not known, although its co-localization with MYP in the yolk platelets and the concurrent decrease in maternal stores of YP30 and MYP suggest that YP30 could be involved in the utilization of MYP. Although we do not know the contents of cortical granules in most animals, those of the sea urchin are well characterized (Fig. 5). Some of the proteins are enzymes activated by secretion, including: ovoperoxidase (Deits et al., 1984; Foerder and Shapiro, 1977; LaFleur et al., 1998), the cortical granule serine protease CGP1 (Haley and Wessel, 1999), and a β-1,3-glucanohydrolase (Bachman and McClay, 1996; Glabe et al., 1983). At least one of the transglutaminases found in the genome (Fig. 5) is thought to act on the fertilization envelope, but its exact location in the egg is still not certain (Battaglia and Shapiro, 1988; IL Wong, unpublished). Other contents are structural proteins that interact with each other and with members of the vitelline layer to form the fertilization envelope, a mechanical barrier to sperm penetration, including: SFE 1, SFE 9, proteolaisin, and rendezvin (submitted for publication). Another protein, hyalin, functions in blastomere adhesion (Wessel et al., 1998).

S. purpuratus has a small collection of heme-dependent peroxidases encoded in the genome (Fig. 5A). Ovoperoxidase is a major enzyme of the fertilization envelope (Mozingo et al., 1994; Weidman et al., 1985); two of the peroxidases are part of dual oxidases Udx1 (Wong et al., 2004) and Udx2; and catalase is the peroxisomal enzyme involved in reduction of hydrogen peroxide. Ovoperoxidase is most closely related to vertebrate myeloperoxidase. It functions to cross link adjacent tyrosine residues of fertilization envelope structural proteins, producing a stabilized molecular structure impervious to additional sperm and microbes, and protective of harsh environmental conditions (Foerder and Shapiro, 1977; Hall, 1978). Udx1 synthesizes hydrogen peroxide essential for this ovoperoxidase-dependent reaction via its NADPH oxidase domain, and is positively regulated by free calcium binding to its tandemly repeated, cytoplasmic EF-hand regions (Wong et al., 2004). The catalase-like
peroxidases of Udx paralogs are distinct from the other heme-dependent peroxidases found within the genome, and their physiologic function is currently not understood.

The gene encoding the cortical granule serine protease (CGSP1) contains both a typical serine protease domain as well as a protein binding motif prevalent in other proteins of the cortical granule (Haley and Wessel, 1999), the low density lipoprotein type A (LDLrA) repeat first identified in the human receptor for low density lipoproteins (Fig. 5A). Unlike Udx1, no CGSP1 paralogs appear to be present in the S. purpuratus genome. Conversely, the β-1,3-glucanohydrolase isoform B found in the cortical granules has two paralogs with identical gene structures (Fig. 5A), and each appears to be expressed in early embryos (Samanta et al., 2006). The presence of a mucin domain in the B isoform suggests that, like CGSP1, this enzyme may directly associate with its target proteins, thereby favoring hydrolysis of glucose polymers on specific substrates. The modularity of both the enzymes’ respective LDLrA and mucin domains in exon units reflect a simple exon insertion event, one that may have promoted the specialization of the respective recipient enzymes CGSP1 and β-1,3-glucanohydrolase (Patthy, 1996; Tordai et al., 2005).

Four S. purpuratus genes encode the five cortical granule proteins that constitute the majority of the fertilization envelope...
structure. One of these proteins is generated by the post-translational proteolysis of one splice variant of rendezvin (submitted for publication). Along with hyalin, mRNAs for these cortical granule proteins make up a considerable proportion of transcripts found in the oocyte (Laidlaw and Wessel, 1994). They are also present throughout the vitellogenic phase of oogenesis, from the smallest of oocytes detectable (about 15 μm) until the last stages of meiosis II. The transcripts of each gene are turned over en mass within the several hours it takes to complete oocyte maturation (Berg and Wessel, 1997; Nakashima and Kato, 2001) and a dramatic decrease in these mRNAs is observed in the egg (Laidlaw and Wessel, 1994; Wessel, 1995; Wessel et al., 2000, 2004). Based on the similar expression profiles of these cortical granule genes, it is likely they are transcriptionally regulated by shared cis-regulatory elements and trans-acting factors, and/or post-transcriptionally regulated by selective and timely degradation of transcripts at oocyte maturation.

Some notable characteristics of these genes include their positions within the genome relative to one another and the modular genetic organization of their repetitive structural domains. For example, the proteoliaisin and SFE9 genes are adjacent to each other, and share a 2.2 kb bi-directional promoter (Fig. 6). This juxtaposition is consistent with a “neighborhood” model of fertilization envelope structural genes also seen in the SFE9 and rendezvin genes of Lytechinus variegatus (data not shown). The size of the four S. purpuratus cortical granule genes, encoded by 10–36 exons spanning some 14–40 kb, precludes a rigorous analysis of their neighborhood at this time, but we believe that the clustering of these structural genes at the same locus may reflect their regulation by chromatin remodeling. For example, opening this single neighborhood of chromatin early in oogenesis would permit coordinated expression of these genes to enhance stoichiometric production of cortical granule gene contents. The molecular mechanism that regulates oocyte-specific gene transcription likely involves the interaction between the chromatin structure, shared cis-regulatory elements in their promoters, and/or common trans-acting factors (reviewed in DeJong, 2006; Song and Wessel, 2005).

Each of the cortical granule structural proteins contains protein interaction motifs common in eukaryotic extracellular proteins. Across phyla, each of these motifs is encoded in the genome in a manner that reflects their versatility and use in a broad range of multi-domain proteins (Tordai et al., 2005). One
family of such ubiquitous motifs can be classified as minimal units for folding linked extracellular repeats (MUFFERs). These single-exon elements encode a motif capable of folding autonomously within the cell. This feature permits it to be very mobile and likely to be retained at new sites within the genome. For example, the hyalin repeat (HYR) is consistently encoded by a single exon, making it a simple structural building block. One HYR consists of about 80 amino acids, and is thought to take on a fold of β-strands belonging to the immunoglobulin superfamily (Callebaut et al., 2000). The 157 predicted S. purpuratus proteins containing this repeat is a gross underestimation of its prevalence, since the HYR is often concatamerized within a protein—as exemplified by its organization in EBR1 and hyalin (Fig. 5B). It has not yet been determined if the context of HYR reflects diversifying function as a protein-binding motif involved in extracellular matrix interactions (Callebaut et al., 2000).

CUB domains represent yet another level of protein motif complexity. This beta-sheet sandwich is generally encoded by 110 amino acids (Bork and Beckmann, 1993; Varela et al., 1997), and is found 422 times throughout the annotated S. purpuratus genome in 257 predicted genes (data not shown). Within the sea urchin genome, 47% of the CUBs are encoded by single exons (199/257), 45% are encoded by two exons (196/257), and the remaining 6% are encoded by three or more exons (27/257).

Fig. 6. Preliminary analysis of genome evolution and utilization. Oocyte-enriched S. purpuratus genes (white boxes, representing exons and introns) analyzed herein, and mapped onto their respective scaffolds (black lines), if known (e.g. ovoperoxidase 5′ end has not been mapped). Genes are grouped as enzymes (A), structural proteins (B), and YP30 paralogs (C). Genes encoded on multiple scaffolds are linked by a dashed line. Orientation of the coding strand is indicated by an arrow within the gene. Numbers in parentheses below scaffold identifications represent the top and bottom nucleotide positions, respectively, of each diagram. Additional SPU annotations/predictions (grey boxes) are identified if they fall within the neighborhood of the oocyte-expressed genes. Scale bar equals 1000 base pairs.
Domains encoded by two or more exons may exhibit additional diversity in the cognate protein structure, consistent with the high frequency (72–82%) of intronic disruptions falling in a flexible loop or linker region of a domain (Barik, 2004; de Souza et al., 1996). Within S. purpuratus, the number of multi-exon CUB domains suggests that about half of these domains contain flexible loops that may participate in activities independent of the CUB fold itself. For example, the presence or absence of such a loop could differentiate between the two types of quaternary folds that CUB dimers form (Romero et al., 1997; Varela et al., 1997); perhaps single-exon CUBs, maintained as MUFFLERs, preferentially dimerize into carbohydrate binding pockets whereas those derived from multiple exons, and containing additional looping sequence, polymerize into sheets that allow the divergent loops to specify protein–protein interactions. This is consistent with our current model that the role of the multi-exon CUB domains of rendezvin (Fig. 5B) is mediating specific protein–protein polymerization within the fertilization envelope, perhaps by forming a platform for protein adherence (submitted for publication). The multi-exon CUB domain is found in other oocyte-specific genes such as the egg binding receptor ebr1 and hyalin, suggesting that their corresponding products also participate in specific protein–protein interactions. ebr1 also exhibits an additional linkage between thrombospondin type 1 motifs and CUB domains, where a heterogenous pair is distributed over three exons (Fig. 5B). This array appears to have replicated as a set since introns of various lengths separate this pair of motifs (data not shown).

The cell surface protein p160 (Haley and Wessel, 2004), which is encoded by a gene with 5 consecutive CUB domains in a single exon (Fig. 5B), is a rare exception (1/257) to the common multi-exon encoding of CUB domains found within the S. purpuratus genome. Other cases of multi-CUB exons can be found in S. purpuratus cubulin (predicted partial sequence, SPU_21921) and a protein similar to ovocyhmase-2 (predicted, SPU_11883). The binding partner of p160 in the vitelline layer has not been identified. One possibility, however, is that the individual CUB domains within the multi-CUB-encoding exon of p160 take on a specific quaternary structure that favors carbohydrate binding, allowing p160 to associate with the glycoproteins in the vitelline layer. Regardless, it is clear that selective pressures are maintaining the p160 CUB-containing exon unit over the more abundant multi-exon CUBs in the genome. Thus, the organization of these MUFFLER-like CUB domains within this gene may be linked to the function of p160.

Although a common mechanism of protein domain movement involves exact duplication of gene fragments, such shuffling of mobile domains is not restricted to sites with pre-existing introns (de Roos, 2005; de Souza et al., 1996; Fedorov et al., 2003; Patthy, 1996). Many genes contain single exons with consecutive domains, which may have evolved through recombination-based expansion (Patthy, 1996). This is likely the case for the LDLrA repeat (Patthy, 1996), a 40-residue MUFFLER capable of multimerizing, as seen in the S. purpuratus cortical granule proteins (Fig. 5B). The classic MUFFLER stoichiometry of one-exon-to-one-LDLrA is utilized for all 5 motifs in the cgsp1 gene. However, the other LDLrA-containing cortical granule genes, proteoliaisin, sfe1, and sfe9, possess exons encoding tandem LDLrA repeats. The highly conserved nucleotide sequence of these intragenic domains (Wong and Wessel, 2004) strongly suggests that the tandem replications evolved through imperfect chromatid pairing during meiotic recombination (Patthy, 1996). This is consistent with the variability in quantity of near-identical LDLrA motifs between S. purpuratus and Lytechinus variegatus orthologs (Wong and Wessel, 2004). The final number of concatemerized LDLrA repeats in each exon, however, may also be indicative of allelic variance among individuals within a species. A similar, recombination-based concatamerization may have given rise to the imperfect repetitive sequences encoded by the ultimate exon of sfe1, sfe9 and rendezvin (Fig. 5B). While it is clear that selective pressure has retained tandem, repetitive sequences in single exons, the initial appearance and the consequence of each exon’s length have not been determined for these genes.

Does motif organization within a protein influence its behavior? Tandem expansion of one MUFFLER, such as LDLrA repeats, might define a new protein architecture. Alternatively, tandem heterogeneous MUFFLERs, such as TSP1-CUB pairs, could establish new functional arrays. Local arrangements within multi-domain proteins affect the stability of a particular motif within a gene (Tordai et al., 2005), implying that selective forces can positively enrich a gene’s mosaicism based on how these domains ultimately function together in the cognate protein. Within the fertilization envelope, for example, CUB domains interact specifically with proteins containing LDLrA repeats (submitted for publication). Almost one quarter of the predicted S. purpuratus genes with CUB domains also contain LDLrA repeats (63/257), implying that proteins can possess both motifs without impairing their function. Given the latter case, why keep the CUB domains and LDLrA repeats separate among the fertilization envelope proteins (Fig. 5B)? If CUB-LDLrA associations are high affinity, then use of multidomain proteins with both CUB and LDLrA motifs in the starting material of the fertilization envelope could be detrimental to the assembly process since these proteins would preferentially self-associate—implying polymerize with another protein—a useless endpoint when trying to establish a heterogeneous network from cortical granule- and vitelline layer-derived proteins. Thus, one parsimonious rationalization for keeping CUB and LDLrA domains on different fertilization envelope proteins is to separate them until the last minute because their interaction is rapid and irreversible—two important traits for a structure that is required to provide both a block to polyspermy and protection during early development.

Summary

Throughout its lifecycle (Fig. 1), the oocyte constantly undergoes changes unlike any other cell in the adult. These events require dramatic shifts in global gene expression to initiate meiosis, to harness and store nutrients, and to synthesize specialized organelles and an extracellular matrix that are all essential for fertilization and subsequent development. In
surveys the *S. purpuratus* genome, we have begun to understand how the genomic loci and transcription factors may be coordinated to orchestrate the processes necessary to make a viable egg. This preliminary survey now enables us to better understand the influence of genome dynamics and evolution on reproduction and development in this basal deuterostome.

**Acknowledgments**

We thank Maryanna Aldrich and Holly Careeskey (Brown University, RI) for their work on meiosis genes; and Dr. Annette Coleman (Brown University, RI) for use of PAUP. We thank Dr. Richard Hynes (Massachusetts Institute of Technology, MA) for annotating the vitellogenin and YP30 genes, and Meredith Howard-Ashby, Stefan Materna, Qiang Tu, and other members of Eric Davidson’s lab (California Institute of Technology, CA) for providing us with transcription factor primer sets and data prior to publication. This work was supported by grants from NSF and NIH awarded to G.M.W. J.L.S is supported by an NRSA.

**Appendix A. Supplementary data**

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ydbio.2006.07.041.

**References**


