

Table S3. Newly identified components of the PMC GRN

Protein	Biochemical function	Motifs (NCBI CDD)	Notes
α -Actinin	Actin crosslinking protein	a) Calponin homology (2) b) EF hand c) Spectrin repeats	α -actinin functions to bundle actin filaments.
Alx4	Transcription factor	a) Homeodomain (1) b) OAR domain (1)	The <i>alx4</i> gene is closely linked to <i>alx1</i> . The gene is expressed by PMCs during gastrulation, but later by NSM cells in the coelomic pouches.
Arp1	Binds ATP and several proteins	Actin-related protein	Arp1 is a component of the dynactin complex, which regulates cytoplasmic dynein activity (1).
Arp3	Binds ATP and several proteins	Actin-related protein	Arp3 is a component of the ARP (Arp2/3) complex, which regulates the nucleation of actin filaments (2).
Calumenin	Calcium binding protein	a) ER signal sequence b) EF hand (one or more depending on stringency) c) HEDF ER retention signal	The ER retention signal of mammalian CALU is only partially functional and some protein is secreted. CALU interacts with thrombospondin (3). There are other CALU-like proteins in <i>S. purpuratus</i> but this is the closest BLAST match to human CALU.
Casc1	Unknown	None	Casc1 (cancer susceptibility candidate 1) is highly conserved in diverse metazoans and has some similarity to a flagellar dynein subunit (4). There is evidence of early embryonic expression in several species. There are no closely related genes in <i>S. purpuratus</i> .
Can1	Carbonic anhydrase	a) ER signal sequence b) CA superfamily c) GPI anchor	Carbonic anhydrases catalyze the reversible hydration of carbon dioxide. There are at least 19 carbonic anhydrase genes in sea urchins.
Cdc42	GTPase	Ras-like GTPase	Cdc42 acts as a molecular switch that transduces signals to the cytoskeleton; among its functions is the activation of Wasp. Cdc42 activity is associated with filopodium formation in several cell types (5, 6).
Cdi	Unknown	None	Cdi is a basic, arg/ser-rich protein with very weak similarity to cyclin-dependent kinase inhibitors, There are no closely related genes in <i>S. purpuratus</i> .
Cofilin	Actin binding protein	ADF superfamily	Cofilin regulates the dynamic assembly/disassembly of actin filaments.
Ctd	Phosphatase	HAD superfamily	This gene model lacks the N-terminus of the protein. There are several related hydrolase domain (1) proteins in <i>S. purpuratus</i> . Ctd closely resembles human SCPs; small ser/thr phosphatases that play a role in regulating the activity of specific transcription factors. These enzymes are related to CTD (C-terminal domain of RNAPIII) phosphatase.
EGF-rich/P41	Matrix protein	a) ER signal seq b) VWA matrilin (2) c) EGF (1)	Other motif prediction programs (Pfam and ExPASy) find 9-16 EGF repeats in this protein. There are many EGF domain-containing proteins in <i>S. purpuratus</i> . The most closely related proteins in vertebrates are fibrillins/matrilins.
Elav	RNA binding protein	RRM (4)	Elav proteins have been implicated in gene regulation and cell differentiation via post-transcriptional mechanisms in diverse organisms. There are other RRM-containing RNA-binding protein in <i>S. purpuratus</i> but this is the closest BLAST match to human Elav.
Fos	Transcription factor	bZIP (1)	Fos can dimerize with Jun to form the AP-1 transcription factor complex. Jun is also enriched in PMCs. Fos has been shown to play a role in many types of cells that, like PMCs, are invasive and motile (7, 8).
Lasp1	Actin binding phosphoprotein	a) LIM (1) b) nebulin (1) c) SH3 (1)	A second, closely related <i>lasp</i> gene is located within a large intron of this gene. Lasp proteins are enriched in cell protrusions and have been implicated in cell migration (9).
Pks2	Type I polyketide synthase	a) ketoacylsynthase (1) b) acyltransferase (1) c) dehydrogenase (1) d) ketoacyl reductase (1) e) thioesterase (1)	PKSs are multifunctional enzymes that carry out the biosynthesis of a wide variety of polyketides. Phylogenetic analysis of sea urchin <i>pks</i> genes has been carried out by Castoe et al. (10).
P11	Unknown	ER signal sequence	P11 is a small (101 aa), basic, cysteine-rich, secreted protein. There are no strong BLAST matches to any other proteins in any organism.
P16rel1	Biomineralization protein	a) ER signal sequence b) Transmembrane (1)	P16rel1 is an acidic, serine-rich protein, similar to P16 (11) in general features. The <i>p16rel1</i> gene is located ~100 kb from <i>p16</i> . There are no apparent orthologs outside the echinoderms.
P16rel2	Biomineralization protein	a) ER signal sequence b) Transmembrane (1)	P16rel2 is a basic protein with glycine- and proline-rich repeats that resemble those found in spicule matrix

			proteins. The <i>p16rel2</i> gene is directly adjacent to <i>p16</i> . There are no apparent orthologs outside the echinoderms.
P21-arc	ARC complex protein	None	P21-arc (ARPC3) is a subunit of the ARP (Arp2/3) complex, which regulates the nucleation of actin filaments (2).
P58A/B	Unknown	a) ER signal sequence b) Transmembrane (1)	The <i>p58a</i> and <i>p58b</i> genes are immediately adjacent to one another. The corresponding proteins are 63% similar at the amino acid level. P58A and P58B are more similar to each other than to any other proteins in <i>S. purpuratus</i> . Both proteins are required for biomineralization (12).
P133	Unknown	a) LamG (at least four) b) EGF (at least three) c) Transmembrane (1)	This gene model lacks the N-terminus of the protein. Combinations of EGF and LamG domains are found in a number of cell adhesion molecules and ECM receptors, including neurexin and some cadherins.
Smad1/5/8	Transcription factor	a) MH1 (1) b) MH2 (2)	
SM20	Spicule matrix protein	C-type lectin (1)	This gene is clustered with sm29 and sm21. The SM20 protein is found in purified spicules (13).
SM49	Spicule matrix protein	C-type lectin (1)	SM49 is found in purified spicules (13).
β -Spectrin	Actin crosslinking protein	a) Calponin homology (2) b) Spectrin repeats	α -Spectrin is also enriched in PMC (14).
Stomatin	Membrane/lipid raft protein	a) Band 7/stomatin (1) b) Transmembrane (1)	Stomatin is a conserved, widely-expressed, integral membrane protein that may regulate membrane transporters or ion channels (15). <i>S. purpuratus</i> has a cluster of eight tandem, highly similar <i>stomatin</i> genes, several of which are expressed by PMCs.
Talin	β -integrin/actin binding protein	a) FERM (1) b) I/L-WEQ (2)	One of the functions of talin is to bind to β -integrin. Three β -integrins (β C, β D, and β G) are represented in the PMC EST collection (supplementary material Table S1) and all have talin-binding motifs (16).
Tsp	Unknown	a) ER signal sequence b) Thrombospondin type 1 (1)	Tsp is a small (123 aa), secreted protein. An adjacent gene (NCBI LOC593393, GLEAN3_09954) encodes a very similar (but distinct) protein with an ER signal sequence and thrombospondin repeat.
Vigilin	RNA binding protein	KH type I (14)	Vigilin is conserved from yeast to mammals and has been implicated in heterochromatin formation and mRNA stability/translation (17).
Wasp	Binds multiple proteins	CRIB (1)	WASP regulates the organization of the actin cytoskeleton by binding to actin, the Arp2/3 complex, and other proteins; it is activated by binding to Cdc42.

References

- (1) Schroer, 2004
- (2) Pollard; 2007.
- (3) Honoré; 2009.
- (4) Wirschell et al., 2009.
- (5) Ahmed et al., 2010.
- (6) Mellor; 2010.
- (7) Ozanne et al., 2007.
- (8) Sherwood, 2006.
- (9) Raman et al., 2010.
- (10) Castoe et al., 2007.
- (11) Cheers and Etensohn, 2005.
- (12) Adomako-Ankomah and Etensohn, 2011.
- (13) Mann et al., 2010.
- (14) Wessel and Chen; 1993.
- (15) Kadurin et al., 2009.
- (16) Whittaker et al., 2006).
- (17) Zhou et al., 2008.