A large-scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo

Xiaodong Zhu[1,2], Gregory Mahairas[3,*], Michele Illies[1], R. Andrew Cameron[4], Eric H. Davidson[4] and Charles A. Ettensohn[1,‡]

[1]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[2]Stowers Institute for Medical Research, Kansas City, MO 64110, USA
[3]Department of Molecular Biotechnology, University of Washington School of Medicine, Seattle, WA 98195, USA
[4]Division of Biology, California Institute of Technology, Pasadena, CA 91124, USA
*Present address: Institute for Systems Biology, Seattle, WA 98105, USA
‡Author for correspondence: (e-mail: ettensohn@andrew.cmu.edu)

SUMMARY

The primary mesenchyme cells (PMCs) of the sea urchin embryo have been an important model system for the analysis of cell behavior during gastrulation. To gain an improved understanding of the molecular basis of PMC behavior, a set of 8293 expressed sequenced tags (ESTs) was derived from an enriched population of mid-gastrula stage PMCs. These ESTs represented approximately 1200 distinct proteins, or about 15% of the mRNAs expressed by the gastrula stage embryo. 655 proteins were similar ($P<10^{-7}$ by BLAST comparisons) to other proteins in GenBank, for which some information is available concerning expression and/or function. Another 116 were similar to ESTs identified in other organisms, but not further characterized. We conservatively estimate that sequences encoding at least 435 additional proteins were included in the pool of ESTs that did not yield matches by BLAST analysis. The collection of newly identified proteins includes many candidate regulators of primary mesenchyme morphogenesis, including PMC-specific extracellular matrix proteins, cell surface proteins, spicule matrix proteins and transcription factors. This work provides a basis for linking specific molecular changes to specific cell behaviors during gastrulation. Our analysis has also led to the cloning of several key components of signaling pathways that play crucial roles in early sea urchin development.

Key words: Morphogenesis, Gastrulation, Cell migration, Cell fusion, Sea urchin embryo, Primary mesenchyme cells, EST analysis

INTRODUCTION

The primary mesenchyme cells (PMCs) of the sea urchin embryo have been a powerful experimental system for the analysis of morphogenesis at the cellular level. The optical transparency of the sea urchin embryo and the ease with which PMCs can be isolated and manipulated both in vivo and in vitro have led to a detailed understanding of PMC behavior during gastrulation and later embryogenesis (reviewed by Gustafson and Wolpert, 1967; Okazaki, 1975a; Solursh, 1986; Ettensohn et al., 1997; Ettensohn, 1999).

The PMCs are the sole descendants of the large micromeres, four blastomeres that form near the vegetal pole of the 32-cell stage embryo. The progeny of the large micromeres become incorporated into the epithelial wall of the blastula near the center of the vegetal plate. At the beginning of gastrulation, these cells undergo a conversion from an epithelial to a mesenchymal phenotype. They become motile and ingress into the blastocoel, migrating on the inner surface of the gastrula wall by means of numerous filopodia (Gustafson and Wolpert, 1967; Malinda et al., 1995; Miller et al., 1995). PMC filopodia

interact with a complex mixture of extracellular matrix (ECM) molecules that form a thin basal lamina lining the blastocoel cavity. The PMCs gradually accumulate in a characteristic ring-like pattern near the equator of the embryo, guided by substrate-associated cues that arise progressively during the blastula and gastrula stages. As the PMCs migrate, their filopodia fuse, forming long cables that link the cells in a syncytial network (Okazaki, 1965; Hodor and Ettensohn, 1998). Within these filopodial cables, the PMCs secrete the crystalline rods (spicules) that constitute the elaborate larval skeleton (Decker and Lennarz, 1988; Wilt, 1999). These cellular events have been described in considerable detail. Indeed, there is (arguably) a more complete understanding of the morphogenetic behavior of PMCs than that of any other population of embryonic cells.

An elucidation of molecular mechanisms that underlie PMC morphogenesis has lagged behind our understanding of PMC behavior at the cellular level. Recent studies have pointed to molecular changes that accompany ingression (Miller and McClay, 1997; Hertzler and McClay, 1999) and a PMC substrate molecule has been cloned – the proteoglycan core

protein-like molecule, ECM3 (Hodor et al., 2000). Two other ECM molecules, pamlin and ECM18, have been identified that may also play a role in PMC migration (Katow, 1995; Berg et al., 1996). Approximately 15 gene products expressed specifically by PMCs, or enriched in these cells, have been cloned. These include four spicule matrix proteins, SM50 (Benson et al., 1987), SM30 (George et al., 1991), PM27 (Harkey et al., 1995) and SM37 (Lee et al., 1999a); the cytoskeletal proteins α-spectrin (Wessel and Chen, 1993), profilin (Smith et al., 1994) and actin CyIIa (Cox et al., 1986); the cell surface protein MSP130 (Leaf et al., 1987); an ETS-family transcription factor (Kurokawa et al., 1999), several collagens (Angerer et al., 1988; Wessel et al., 1991; Suzuki et al., 1997); and lamin B (Holy et al., 1995). PMCs also express at least two β-integrins (Marsden and Burke, 1997; Marsden and Burke, 1998). In some cases, the functions of these molecules have been partly defined. For example, spicule matrix proteins are integral components of the skeletal rods and play an important role in regulating the process of biomineralization (Wilt, 1999). Secretion of collagen by the PMCs appears to provide a necessary microenvironment for skeletogenesis (Blankenship and Benson, 1984; Wessel et al., 1991), perhaps by regulating the presentation of growth factors that control the expression of specific spicule matrix protein genes (see Ettensohn et al., 1997).

To gain a more detailed understanding of the molecular basis of PMC morphogenesis, we have carried out a large-scale analysis of mRNAs expressed by these cells during gastrulation. We took advantage of the fact that PMC precursors, the micromeres of the 16-cell stage embryo, can be isolated in large quantities and cultured in vitro under conditions that allow the cells to undergo a normal program of differentiation (Okazaki, 1975b; Harkey and Whiteley, 1983). This analysis has led to the identification of candidate regulators of cell migration, cell fusion and skeletogenesis, and to the cloning of key components of signaling pathways that have been shown to function in a variety of contexts in the early embryo.

MATERIALS AND METHODS

Embryo and cell culture

Adult *Strongylocentrotus purpuratus* were purchased from Marinus (Long Beach, CA). Gametes were obtained by intracoelomic injection of 0.5 M KCl. Micromeres were isolated and cultured according to a protocol provided by Steve Benson (personal communication). Briefly, eggs were fertilized in 10 mM para-aminobenzoic acid and rinsed twice with fresh artificial seawater (SW). The fertilized eggs were cultured at 15°C. At the four-cell stage, the seawater was replaced with Ca^{2+}-free seawater (CF-SW). At the 16-cell stage, fertilization membranes were removed by passing the embryos through 53 μm Nitex mesh. The embryos were then rinsed several times with Ca^{2+}/Mg^{2+}-free seawater at 4°C (50× the packed embryo volume/rinse) and suspended in CF-SW (15× the packed embryo volume). The embryos were dissociated by pipetting using a 9 inch (23 cm) Pasteur pipette. The dissociated cells were loaded on a 3-30% sucrose gradient at 4°C and separated at 1 *g* for 40 minutes. The micromeres, which formed a clear band one-quarter to one-half inch (5-12 mm) below the top of the gradient, were drawn off and plated at a density of $2×10^4$ cells/cm^2 on 100 mm tissue culture dishes. After the cells attached to the plates they were rinsed three times with sterile SW and cultured in sterile SW supplemented with 2% horse serum

and 1× penicillin-streptomycin-glutamine (Gibco Life Technologies). The cells were cultured at 15°C, until sibling embryos reached the mid-gastrula stage, when they were collected for RNA isolation (below). About 70% of the cultured cells were positive when immunostained with monoclonal antibody 6e10, which recognizes the PMC-specific cell surface glycoprotein MSP130. The remaining cells were presumably large micromeres that did not differentiate into PMCs, small micromere derivatives and cells derived from contaminating mesomeres and macromeres.

cDNA library construction and arraying

Total RNA was extracted from cultured cells using Trizol reagent (Gibco Life Technologies). Poly(A)+ RNA was isolated using a MicroPoly(A)Pure kit (Ambion). cDNA was synthesized using an oligo(dT) primer and cloned directionally into the pSPORT plasmid vector following the manufacturer's instructions (Gibco Life Technologies). The average insert size was 1.5-2.0 kb. The library was arrayed in 384-well plates using a Genetix Q-Bot robot.

DNA sequencing and sequence analysis

Plasmid template DNA was prepared from individual clones isolated from wells of the 384-well plates. Cells were grown overnight in 400 μl of LB medium and lysed under alkaline conditions (Birnboim and Doly, 1979). Lysates were cleared using Millipore lysate clearing plates (catalog code, MANANLY50) and DNA was purified using Millipore multiscreen glass fiber filter plates (catalog code, 52EM108M8). Most steps in the plasmid isolation procedure were carried out robotically using Beckman Multimek 96 pipetting robots and automated filtration stations.

A single sequencing reaction was carried out with each template using dideoxy chain terminators and a T7 primer, which provided sequence from the 5′ ends of the directionally cloned cDNAs. Sequencing reactions were resolved using ABI 3700 DNA analyzers. The average length of readable sequence was 600-800 nucleotides. Several clones of interest were subsequently sequenced fully on both DNA strands using an ABI 377 sequencer at the University of Pittsburgh School of Medicine DNA Sequencing Facility.

DNA sequences were loaded into an Oracle database and subjected to quality control using Phred (Ewing and Green, 1998; Ewing et al., 1998). Sequences were trimmed to remove contaminating vector (pSport), *Escherichia coli* and *S. purpuratus* repetitive sequences. Sequences that included >200 bp of Q20 bases were considered of high quality and subjected to BLASTX searching against the public non-redundant protein databases in GenBank.

Whole-mount in situ hybridization

Whole-mount in situ hybridization was performed as described previously (Guss and Ettensohn, 1997), with minor modifications. Tween-20 (0.1%) was included in all wash solutions to prevent embryos from adhering to the walls of microfuge tubes. After hybridization with digoxigenin-labeled probes, embryos were washed with $0.1 × SSC$ (rather than $1 × SSC$) to reduce nonspecific staining.

RESULTS

Overall distribution of sequences

A total of 8293 high-quality sequences were generated and subjected to BLASTX analysis. The initial data for each clone consisted of a single sequencing reaction primed at the 5′-most end of the cDNA, although a number of clones were later sequenced fully (see below). All 8293 DNA sequences can be accessed through GenBank (Accession Numbers BG780044-BG789442) or through the sea urchin genome project (http://sea-urchin.caltech.edu:8000/genome/databases).

Table 1. Overall distribution of sequences

A   Number of sequences analyzed by BLASTX = 8293
B   Sequences that matched known proteins* = 1629
  →Distinct proteins matched = 771
    →Distinct, previously characterized proteins matched‡ = 655
    →Distinct ESTs matched§ = 116
C   Sequences in 'no match' category = 6664
  →Sequences of poor quality or with short inserts¶ = 600
  →rRNA sequences = 450
  →Untranslatable, interspersed repetitive sequences = 200
  →Protein coding sequences in 'no match' category** = 913
  →Distinct proteins represented‡‡ = 435
D   Total number of distinct proteins identified§§ = ~1200

*Based on BLAST analysis of GenBank at a probability value $P<10^{-7}$.

‡Distinct proteins about which there is information other than sequence; e.g. information concerning expression pattern, biochemical function, etc.

§Distinct ESTs from other organisms about which no further information was available in GenBank.

¶Sequences with >10% unreadable residues or inserts <50 nucleotides in length. All estimates in C) were based on a random sample of 315 sequences in the 'no match' category analyzed by BLASTN.

**Calculated based on the fraction of 'no match' sequences (excluding those with short inserts, poor sequence, rRNAs, and sequences similar to untranslatable repetitive mRNA) with +1, +2 or +3 ORFs equal to or longer than 350 nucleotides.

‡‡Calculated assuming the same average frequency of matches/protein as the population of identified proteins (2.1 matches/protein).

§§The sum of lines B1 and C5.

Table 2. Most frequently identified proteins

Protein	Number of matches
Cytochrome C oxidase, subunit I	103
MSP130	44
Ribonucleotide reductase, small subunit	30
Cyclin B	25
Ubiquitin	24
ATP synthase, subunit 6	19
NADH dehydrogenase, subunit 5	18
Tubulin, alpha	18
eIF-1, α	16
Cytochrome B	15
ADP/ATP translocase	14
eIF-4A	13
Histone H1, embryonic	13
IGF-II mRNA-binding protein	13
Cyclin A	12
MSP130-related 2	11
PM27	10
SM37	10
Histone H2A variant	9
HMG1	9
NADH dehydrogenase, subunit 4	9
SM50	9

BLASTX analysis showed that of the 8293 ESTs, 1629 were strong matches ($P<10^{-7}$) to previously identified proteins (Table 1). The frequency of matches was therefore 1629/8293 (about 0.20). Further analysis of the 1629 matches showed that they represented 771 distinct proteins. Of the 771 proteins, 116 were matches to ESTs (mostly from *Caenorhabditis elegans* and *Homo sapiens*), while the remaining 655 were matches to proteins that have been characterized to varying extents. In some cases, matches were to proteins that have been characterized only with respect to their pattern of expression, while other matches were to proteins with well-defined biochemical and cellular functions. A complete list of the 655 proteins, grouped according to major cellular function, is shown in the Appendix. The great majority of proteins were identified only once in our analysis, although a few highly abundant proteins were identified many times (Table 2). The average number of hits/protein was 2.1 (1629/771).

About 80% of the ESTs fell into the 'no match' category (6664/8293 sequences). By examining a random sample of 315 sequences in this category, we estimated that a very small fraction (3%, or ~200 total sequences) could be accounted for by poor sequence data (i.e. sequences with >10% unreadable bases) that were not eliminated in the initial sequence screening. Approximately 6% of the cases in the 'no match' category (~400 total sequences) contained no insert or one shorter than ~50 nucleotides when analyzed by BLASTN and therefore represented an artifact of library construction. A similar proportion of the 'no match' cases (6.8%, or ~450 total sequences) were rRNA sequences. The most commonly identified sequence in this class was mitochondrial 16S rRNA, a super-abundant polyadenylated transcript (see also Davidson, 1986; Poustka et al., 1999). In addition, approximately 2% of the ESTs (~200 total) represented sequences similar to untranslatable, interspersed repetitive sequences that have been identified in the egg (Costantini et al., 1978). Taken together,

these findings indicate that a relatively small fraction, about 18% of the 'no match' category, can be accounted for by these classes of transcripts. The remaining 82% are therefore likely to represent other untranslated sequences and bona fide proteins that did not match entries in GenBank.

We arrived at a conservative estimate of the number of protein-coding sequences in the 'no match' category by comparing the distribution of maximum open reading frame (ORF) lengths of sequences in this category to the distribution of maximum ORF lengths in untranslated sequences (see also Lee et al., 1999b). Because the PMC library was oligo(dT)-primed, we assumed that most clones in the 'no match' category represented 3′ UTR sequences. We analyzed all *S. purpuratus* genes in GenBank, as of 7/11/2000, for which 3′-UTR sequences >600 nucleotides were available (39 genes). We divided these 3′-UTRs into 120 non-overlapping segments with an average length of 800 nucleotides, a value chosen to match the average read length of a random sampling of 'no match' sequences (described below). The longest ORF in each fragment was determined and the distribution of these values is plotted as a histogram in Fig. 1A. Most of the ORFs were quite short and in no case did we find an ORF greater than 350 nucleotides in length. For comparison, the maximum ORF lengths of a sample of 120 ESTs that yielded strong BLAST matches ($P<10^{-7}$) were also plotted (Fig. 1A). The average length was much greater and the distribution only slightly overlapped that of the 3′-UTR ORFs.

We then chose 120 ESTs at random from the 'no-match' category, excluding those with no inserts, low-quality sequences and rRNA sequences. The average read length of these sequences was 796 nucleotides. As above, we determined the length of the longest ORF in each sequence and plotted these as a histogram (Fig. 1B). Because of the directional cloning strategy used to construct the library, the great majority (>95%) of strong BLAST matches to known proteins were in one orientation (reading frames +1, +2 and +3) and the same must be true of cryptic protein-coding sequences in the 'no match' population. We therefore restricted our analysis of ORFs in the 'no match' sequences to the three positive reading

A

B
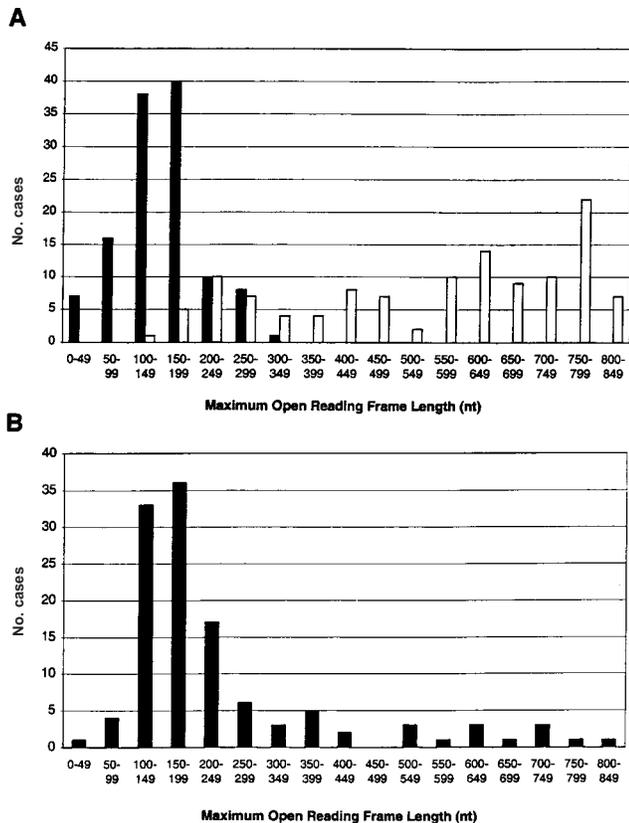
Fig. 1. Histograms of maximum ORF lengths. (A) Black bars are plots of the lengths of the longest ORFs represented in 120 segments (average length=800 nucleotides) of sea urchin 3′-UTRs. The segments were obtained from all sea urchin genes in GenBank as of 7/11/2000 for which 3′-UTR sequences >600 nucleotides were available (39 genes). Most of the ORFs are short and none was >350 nucleotides in length. White bars are of the lengths of the longest ORFs in a random sample of 120 ESTs (average length=794 nucleotides) that yielded strong BLAST matches ($P<10^{-7}$). There is little overlap between this distribution and the one derived from 3′-UTR sequences. (B) Plot of maximum ORF lengths in 120 randomly selected ESTs from the 'no match' category (average length=796 nucleotides). Because the cDNA library was directionally cloned, only the longest ORFs in the three positive reading frames (+1, +2, +3) were considered. 20/120 clones (16.7%) contained an ORF equal to or longer than 350 nucleotides.

frames. As expected, most sequences in this category contained only short ORFs (100-200 nucleotides) but longer ORFs were also apparent in the population. 20/120 clones (16.7%) contained an ORF equal to or longer than 350 nucleotides.

We used the 16.7% value as an estimate of the fraction of ESTs in the 'no match' category that represented bona fide coding sequences. This is likely to be a conservative estimate. As shown in Fig. 1A, a small but significant fraction (22.5%) of protein-coding ESTs (i.e. those with strong BLAST matches) had maximum ORF lengths of <350 nucleotides. Undoubtedly, some cryptic protein-coding sequences also had maximum ORF lengths shorter than 350 nucleotides. Nevertheless, by using the 16.7% value, and after eliminating from consideration poor sequences, clones with very short inserts, rRNA sequences, and sequences similar to

untranslatable, repetitive mRNA (18% of the sequences in the 'no match' category, as described above), we estimate that 913 sequences in the 'no match' population represent bona fide protein-coding sequences. If these cryptic proteins exhibit, on average, the same prevalence distribution as proteins identified by BLAST matches, then 913 sequences would represent 435 distinct proteins (2.1 hits/protein). By adding this value to the 771 distinct proteins already identified by BLAST analysis, we conservatively estimate that the EST database contains sequences corresponding to ~1200 different proteins.

Further sequence analysis of selected cDNA clones

The EST analysis identified a number of cDNA clones that encoded especially strong candidates for regulators of PMC morphogenesis. As a first step in the further characterization of such gene products the complete sequences of several clones were determined.

Extracellular matrix molecules

Fibronectin

Clone 03-0233 contained a large (>4 kb) insert that encoded 1079 amino acids of a molecule with significant similarity to vertebrate fibronectin. The closest BLAST match was to bovine fibronectin ($P=3\times10^{-41}$). The ORF encoded a large C-terminal portion of the sea urchin protein which consisted of nine tandem fibronectin Type III repeats followed by a 200-300 amino acid region at the extreme C terminus that was not similar to known proteins at a significant level. An RGDT sequence was identified within the sixth Type III repeat. This tetrapeptide has a cell-binding function in vertebrates (Pierschbacher and Ruoslahti, 1984).

Fibrinogen-related protein

Clone 0016_B2_H02 contained the entire ORF of a protein related to fibrinogen. This protein was 308 amino acids in length and consisted of an N-terminal signal sequence, a short (~70 amino acid) segment without significant similarity to known proteins, and a C-terminal fibrinogen-related domain (FRD). This globular domain is found at the C terminus of a variety of extracellular proteins in both vertebrates and invertebrates, including fibrinogens β and γ, angiopoietins, tenascin, ficolins, the product of the *Drosophila scabrous* gene, and some lectins (see Xu and Doolittle, 1990; Conklin et al., 1999; Gokudan et al., 1999).

Nidogen

Clone 0026_B2_F03 contained a single long ORF encoding more than 1200 amino acids with significant similarity to vertebrate and invertebrate nidogen/entactin. The closest match by BLAST analysis was to human nidogen ($P=3\times10^{-63}$). Based on comparison with the human sequence, the sea urchin clone appears to lack only the N-terminal-most 10-20 amino acids of this protein.

Osteonectin

Clone 0016_B2_D04 encoded a full-length protein with significant similarity to vertebrate and invertebrate osteonectin (SPARC, BM-40). The closest BLAST match was to osteonectin/SPARC from *Caenorhabditis elegans* ($P=4\times10^{-36}$). Sea urchin osteonectin is 270 amino acids in length, a size similar to that of osteonectins from other organisms (250-310

amino acids). Like other osteonectins, the sea urchin protein has a putative signal sequence, is acidic (calculated pI=4.35) and relatively rich in cysteines.

Potential regulators of PMC migration, fusion, and skeletogenesis

Rac

Clone PM990802-08-0472 encoded a full-length homolog of the small GTPase Rac. The closest BLAST match was to human Rac1 (4×10^{-94}). Sea urchin Rac is 194 amino acids in length, two amino acids shorter than human Rac1. The N-terminal two-thirds of sea urchin Rac and human Rac1 are identical at 118/121 positions, and the two proteins show ~90% amino acid identity overall.

Tetraspanin NET-5

Clone PM990802-06-0460 encoded a full-length protein highly similar to human tetraspanin NET-5 ($P=3\times10^{-39}$). The sea urchin protein exhibited the characteristic organization of tetraspanins; three putative transmembrane domains near the N terminus and a fourth near the C terminus, with a large extracellular loop between the third and fourth transmembrane domains. Both the human tetraspanin NET-5 and its sea urchin counterpart are 239 amino acids in length.

Tetraspanin NET-7

Clone 0016_A2_H04 encoded a full-length protein most similar to human tetraspanin NET-7 ($P=5\times10^{-35}$). The sea urchin protein is 243 amino acids in length, five amino acids longer than human tetraspanin NET-7. It contained the distinctive spacing and number of transmembrane domains described above.

DOCK180/Myoblast city

Clone PM990802-03-0379 encoded a C-terminal fragment, 520 amino acids in length, of a protein highly similar to DOCK180/Myoblast city. The closest BLAST match was to human DOCK180 ($P=4\times10^{-54}$).

Discoidin-domain receptor tyrosine kinase

Clone PM990802-08-0413 encoded the N-terminal two-thirds (~650 amino acids) of a sea urchin homolog of a discoidin-domain receptor tyrosine kinase. The closest BLAST match was to human DDR1 (TrkE) ($P=6\times10^{-75}$). The sea urchin protein has the characteristic domain organization of this class of receptor tyrosine kinase: an N-terminal signal sequence and discoidin domain, a central transmembrane domain, and a cytoplasmic protein tyrosine kinase domain.

Putative spicule matrix proteins and a protein related to MSP130

SM50-related

Clone 0022_B2_H02 encoded the C-terminal 226 amino acids of a previously unidentified spicule matrix protein. The C-terminal 90 amino acids of the protein were organized in 28 tandem copies of a distinctive repeat element of the form P-X-Y, where X is N, F or T (usually N), and Y is Q, N, T, A or R (usually Q). The presence of tandem copies of a proline- and/or glycine-rich repeat is a common feature of many spicule matrix proteins, although the primary sequence of the repeat and the

copy number varies between these proteins (Katoh-Fukui et al., 1991; Livingston et al., 1991; Harkey et al., 1995; Lee et al., 1999a). The remainder of the SM50-related sequence showed a high degree of similarity to the N-terminal region of SM50 ($P=1\times10^{-41}$), which contains a C-lectin-like domain (see Harkey et al., 1995). The two proteins were 60% identical over this 140-amino acid region. The SM50-related protein therefore exhibited the distinctive two-domain structure characteristic of other spicule matrix proteins. The alignment of the SM50-related protein with the N-terminal region of SM50 suggested that the first ~60 amino acids are missing from the SM50-related clone.
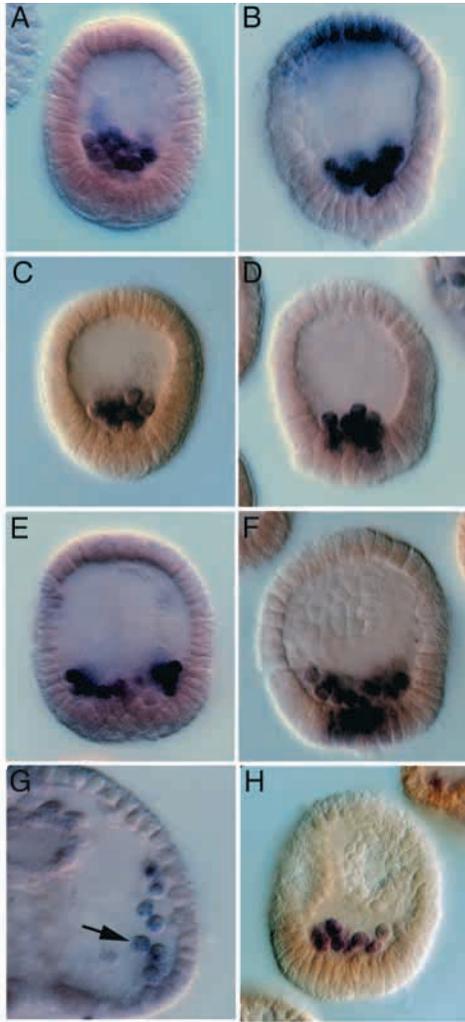
C-lectin

Clone 0014_A1_A09 contained the complete ORF of a small (186 amino acid) protein similar to C-type lectins from several organisms. The closest BLAST match (3×10^{-20}) was to echinoidin, a C-type lectin identified in *Anthocidaris crassispina* (Giga et al., 1987), but the degree of amino acid identity was sufficiently low (34%) to make it doubtful these proteins are homologues. Moreover, another C-lectin similar to echinoidin has been identified in *S. purpuratus* (Smith et al., 1996) and is clearly distinct from the protein identified here. While the C-lectin we identified lacks the obvious repeat elements of other spicule matrix proteins, it exhibits several features that suggest it may belong to this class of proteins: (1) it includes an N-terminal signal sequence and is presumably secreted; (2) it includes a C-lectin domain, as do the previously identified spicule matrix proteins (Harkey et al., 1995; Killian and Wilt, 1996); and (3) it is expressed at high levels specifically by PMCs, as shown by whole-mount in situ hybridization (Fig. 2).

MSP130-related 1

Clone 0025_B2_A08 encoded a large C-terminal region (609 amino acids) of a protein closely related to, but distinct from, the PMC-specific cell surface glycoprotein MSP130 ($P=4\times10^{-48}$). The MSP130-related 1 sequence aligned over its entire length with *S. purpuratus* MSP130 at an overall amino acid identity level of 35-40%. The alignment included all regions of the MSP130 protein, except the N-terminal-most ~90 amino acids (the corresponding amino acids are missing from MSP130-related 1, which is a partial clone) and amino acids 226-378, which correspond to the second glycine-rich domain of MSP130 (Parr et al., 1990). This domain is absent from the MSP130-related 1 protein. Like MSP130, MSP130-related 1 includes 14-16 hydrophobic amino acids at the extreme C terminus that may function as a GPI-anchor domain (Parr et al., 1990). We also identified several cDNAs encoded by a second gene, MSP130-related 2, that is clearly distinct from both MSP130 and MSP130-related 1. MSP130 and the MSP130-related proteins therefore represent a small gene family consisting of at least three members.

Whole-mount in situ hybridization studies

The expression patterns of 20 mRNAs were examined by in situ hybridization. Of these, we found that eight were expressed exclusively or predominantly by PMCs (Fig. 2). Included in this group were two transcription factors (ERG and aristaless), two extracellular matrix proteins (fibronectin and fibrinogen-related protein), two cell surface proteins (MSP130-

Fig. 2. Whole-mount in situ hybridization analysis. (A) Aristaless. (B) Fibrinogen-related protein (FRP). (C) MSP130-like. (D) C-lectin. (E) NET-7. (F) ERG. (G) Fibronectin. (H) SM50-related. Expression of aristaless, MSP130-like, SM50-related and C-lectin is restricted to PMCs at the mesenchyme blastula stage. FRP is expressed by PMCs and by ectoderm cells at the animal pole. ERG mRNA is expressed by PMCs and non-skeletogenic mesoderm cells in the vegetal plate. At the late gastrula stage, fibronectin mRNA is enriched in PMCs (arrow, G).

related 1 and NET-7), and two new putative spicule matrix proteins (SM50-related and C-lectin). Probes against the other 12 mRNAs showed more general labeling patterns consistent with expression in PMCs, as well as other cell types in the embryo.

DISCUSSION

The results of two other sea urchin cDNA sequencing projects, each considerably smaller in scale than the present study, have recently been reported (Lee et al., 1999b; Poustka et al., 1999). Following a strategy essentially similar to that described here, Lee et al. (Lee et al., 1999b) examined 956 ESTs from an arrayed cDNA library generated using *S. purpuratus* cleavage-stage poly(A)+ RNA. Using criteria similar to ours, they

identified 232 ESTs with significant matches to known protein-coding sequences in GenBank. These 232 ESTs were found to represent 153 different proteins. The average number of hits/protein was therefore about 1.5, less than the value we obtained (2.1 hits/protein). This difference is undoubtedly partly due to the larger sample size of our study, as the probability that a given EST will match a previously identified protein increases with the sample size. Another likely contributing factor is the relatively lower diversity of the pool of mRNAs expressed by a specific cell type (in this case, PMCs) from a later developmental stage (Davidson, 1986).

One significant difference between the study by Lee and co-workers and the present work was the method used to prime cDNA synthesis (oligo(dT) versus random priming). Oligo(dT) priming undoubtedly led to a relatively greater representation of 3′UTR sequences in our analysis. Nevertheless, the overall frequency of matches (fraction of total sequences that were significantly similar to previously identified protein coding sequences) was only slightly lower in our study (0.20 versus 0.24). One reason the difference may not have been greater is that the level of rRNA contamination was lower in the PMC library (6-7% versus 14%), probably due at least in part to the method of priming. We may also have had relatively less contribution from untranslatable, interspersed-repeat-containing poly(A)+ RNA. These sequences are on average at least five times as long as mRNAs and would tend to be more highly represented in cDNA libraries generated by random priming than those produced by oligo(dT) priming (Davidson, 1986; Lee et al., 1999b). A minor, but useful, feature of the oligo(dT) priming strategy and 5′ orientation of sequencing was that when N termini of proteins were identified by BLAST analysis; such clones nearly always contained the complete coding sequences of the corresponding proteins.

Poustka et al. (Poustka et al., 1999) used oligonucleotide fingerprinting to generate a normalized cDNA collection representing about one-third of all genes expressed in the fertilized egg of *S. purpuratus*. Starting with an oligo(dT)-primed cDNA library generated from poly(A)+ RNA, 21,925 clones were fingerprinted by hybridization with 217 different 8-mer oligonucleotide probes and grouped into 6291 clusters (corresponding to different transcripts) ranging in size from 1 to 265 clones. In a pilot analysis, the 5′ ends of representative clones from 711 clusters were sequenced and the sequences of 90 clones (12.7%) were found to show significant similarity to 80 distinct proteins in the databases ($P<10^{-5}$). The potential advantage of a fingerprinting approach is that by grouping cDNAs into clusters before selecting clones for sequencing, the probability of resequencing prevalent mRNA species repeatedly is greatly reduced.

In our study, we were able to identify a large number of different proteins without any normalization of the cDNA library, simply by sequencing large numbers of clones. Nucleic acid hybridization studies indicate that there are some 8500 diverse mRNA species at the gastrula stage, assuming a mean length of 2 kb (Galau et al., 1974; Davidson, 1986). Not all these mRNAs are expressed by PMCs, as many transcripts expressed at the gastrula stage have tissue-specific distributions (Kingsley et al., 1993). If we accept, for the sake of argument, that 5000 different mRNA species are expressed by PMCs at the midgastrula stage, then our EST analysis identified approximately one quarter of those gene products. The average

number of hits/protein was still quite low (2.1). If the sample size were increased further, it would become progressively more difficult to identify new mRNAs, and some method of selectively enriching for rare sequences would probably be required to obtain a complete catalog of expressed genes. Nevertheless, a more comprehensive catalog of genes expressed by PMCs could certainly be obtained simply by additional high-throughput sequencing. Such an approach would be facilitated to a modest extent by first performing filter hybridization using gene-specific probes to identify those clones in the arrayed library that correspond to highly abundant sequences (rRNAs, cytochrome C oxidase subunit I, MSP130, etc.), which together represent 10-15% of the clones in the library, and then eliminating those from further analysis.

The gene products that emerged from the EST analysis appear to mirror closely the cellular composition of the cDNA library. Of the 21 gene products identified more than eight times in our analysis (Table 2), four are known to be expressed specifically by PMCs (MSP130, PM27, SM37 and SM50) and the others are proteins with housekeeping functions that are likely to be expressed by many cell types, including PMCs. Moreover, every gene product currently known to be expressed exclusively or primarily by PMCs at the gastrula stage (including SM30, profilin, spectrin, collagens, lamin B, etc.) was identified at least once in our analysis. We have also shown by in situ hybridization that many of the proteins identified for the first time through our sequencing analysis are expressed primarily or exclusively by PMCs (Fig. 2). Based on these observations and our determination of the purity of the cell population used to generate the library, we expect that the great majority of proteins identified in our analysis are expressed by PMCs. Nevertheless, a small number of mRNAs were also identified that are unlikely to be expressed by these cells. Clones were identified that encoded various members of the Spec gene family (10 cases), arylsulfatase (four cases), and hatching enzyme (three cases). All are abundant mRNAs expressed specifically by presumptive or definitive ectoderm cells. Therefore, independent methods will be required to confirm that any specific protein identified in our analysis is expressed by PMCs.

We chose to study PMC gene expression at the equivalent of the mid-gastrula stage. Analysis of proteins synthesized by cultured micromeres by two dimensional gel electrophoresis indicates that the major transition in the molecular program of differentiation of the cells occurs prior to that stage, approximately concomitant with ingression (Harkey and Whiteley, 1983). Most proteins that are upregulated at ingression continue to be synthesized throughout later development. This pattern of protein expression is consistent with studies demonstrating that most major morphogenetic activities of the PMCs are activated by the early to mid-gastrula stage but persist much later in development. For example, the ability of the cells to migrate directionally in response to guidance information in the blastocoel is clearly established by the mid-gastrula stage, when the subequatorial ring forms, and persists at least until the late gastrula stage (Ettensohn, 1990). PMCs first become fusogenic at the early gastrula stage but remain capable of fusing with other PMCs throughout embryogenesis (Hodor and Ettensohn, 1998). Thus, by focussing on the population of mRNAs expressed by PMCs at the mid-gastrula stage, we are very likely to include most of the gene products that regulate the major morphogenetic activities of these cells.

Because our library was not normalized, the frequencies with which we identified specific mRNAs should reflect their relative abundance within the sequence population (Lee et al., 1999b). A potential limitation is that transcripts with unusually long or short 3′-UTRs could be under- or over-represented, respectively, in the pool of ESTs that yielded matches to known protein-coding sequences. Nevertheless, of the 21 proteins identified more than eight times, four are terminal differentiation gene products of PMCs (MSP130, PM27, SM37 and SM50) and the remainder have general housekeeping functions. All these proteins might therefore be expected to be expressed at high levels by PMCs. It has been estimated that there are >200 *SM50* mRNA molecules per PMC and about 16 *PM27* mRNA molecules per PMC at peak expression levels during gastrulation (Killian and Wilt, 1989; Harkey et al., 1995). The single most frequently identified sequence in the EST analysis, mitochondrial 16S rRNA (169 hits) has been shown by independent methods to be the most prevalent poly(A)+ RNA species in the embryo (Davidson, 1986; Poustka et al., 1999). Finally, four of the other proteins in the collection of 21 (cyclins A and B, α-tubulin, and the small subunit of ribonucleotide reductase) were also among a subset of sequences identified multiple times in a random-primed, cleavage stage cDNA library (Lee et al., 1999b). Based on all these considerations, it seems likely that the frequency with which a specific sequence was identified in our EST analysis provides a good indication of the prevalence of the corresponding mRNA in PMCs, at least in most cases. The fact that the great majority of proteins were identified only once or twice in the EST pool indicates that most mRNAs expressed by PMCs are in the moderate-to-low prevalence class.

The EST analysis identified a large number of new, potential regulators of PMC morphogenesis that will be attractive candidates for further study. For example, we identified several proteins that have been shown to play a role in mediating cell-cell fusion in other developing systems. These include the small GTPase Rac, three members of the tetraspanin family, and DOCK180/myoblast city. Tetraspanins are a recently identified family of four-pass transmembrane proteins that function in multiple cellular processes. These proteins regulate integrin function and play a role in the fusion of myoblasts and gametes (Hemler, 1998; Tachibana and Hemler, 1999; Kaji et al., 2000). Genetic and biochemical studies have shown that Rac and DOCK180/myoblast city are important regulators of myoblast fusion and interact directly with one another (Luo et al., 1994; Erickson et al., 1997; Kiyokawa et al., 1998; Nolan et al., 1998; Frasch and Leptin, 2000). We found that NET-7 is expressed at high levels specifically by PMCs, supporting the view that this protein has a special function in these cells.

Several proteins were identified that have been implicated in the regulation of filopodial motility. These include three proteins that regulate actin polymerization and the formation of filopodia and other cell protrusions: Arp3 ($P=2\times10^{-82}$) (cloned previously from *Hemicentrotus pulcherrimus*; GenBank Accession Number, AB016822), N-WASP ($P=4\times10^{-10}$) and cdc-42 ($P=7\times10^{-16}$) (Miki et al., 1998; Carlier et al., 1999; Rohatgi et al., 1999; Borisy and Svitkina, 2000). In addition, we identified putative sea urchin homologs of the cell surface proteoglycan syndecan ($P=1\times10^{-7}$) and

syntenin ($P=3\times10^{-54}$), a cytoplasmic protein that interacts with the C-terminal region of syndecan (Grootjans et al., 1997). Syndecans have been implicated in a variety of processes related to cell adhesion, signaling and motility, including the formation of filopodia (Woods and Couchman, 1998; Granes et al., 1999).

The major biosynthetic activity of the PMCs is the secretion of the calcareous skeleton. We identified two new candidate spicule matrix proteins, SM50-like and a C-lectin, and showed that both were expressed specifically by PMCs. In addition, we identified a discoidin domain receptor (DDR) tyrosine kinase that might function in skeletogenesis. DDRs are an ancient class of receptor tyrosine kinase that have recently been found to act as collagen receptors, undergoing a slow autophosphorylation and activation in response to that ligand (Shrivastava et al., 1997; Vogel et al., 1997; Vogel, 1999; Vogel et al., 2000). A variety of evidence (reviewed by Ettensohn et al., 1997) indicates that PMCs must interact with a self-produced collagenous substrate in order to synthesize spicules, probably in part through the activation of the *SM30* gene, and sea urchin DDR is a candidate for mediating such an interaction. Finally, we identified two proteins closely related to MSP130. MSP130 is a novel, GPI-linked protein that appears to function in facilitating Ca^{2+} import (Farach-Carson et al., 1989). We have found that MSP130 and MSP-related proteins form a small gene family consisting of at least three members. cDNAs encoding MSP130-related 2 were identified 11 times in our analysis, indicating that this mRNA is expressed at high levels by PMCs. MSP130-related 1 was identified only once, but in situ hybridization analysis suggests that this mRNA is also abundant (Fig. 2).

We cloned at least seven new ECM molecules from the sea urchin: perlecan ($P=9\times10^{-67}$), fibronectin ($P=3\times10^{-41}$), fibrinogen-related protein ($P=2\times10^{-44}$), fibrillin ($P=9\times10^{-43}$), F-spondin ($P=3\times10^{-19}$), nidogen ($P=1\times10^{-135}$) and osteonectin ($P=4\times10^{-36}$). One of these, fibronectin, has been implicated in PMC migration in several previous studies (Fink and McClay, 1985; Katow and Hayashi, 1985; Katow et al., 1990). These studies relied on probes against vertebrate fibronectin, however, and sea urchin fibronectin had proven refractory to cloning for many years. Our findings resolve the long-standing issue of whether sea urchins have fibronectin and will allow further analysis of the function of the endogenous protein. We also identified several ECM-degrading enzymes that might function in PMC ingression, migration, or skeletogenesis, including membrane-type matrix metalloprotease 15 ($P=8\times10^{-46}$), matrix metalloprotease 1 (collagenase) ($P=6\times10^{-30}$), heparanase ($P=5\times10^{-18}$) and a metalloelastase ($P=6\times10^{-30}$).

A large number of transcription factors emerged from the EST analysis, most of which were isolated as full-length clones. These included aristaless ($P=8\times10^{-30}$), MTA1 ($P=0$), interleukin enhancer binding factor 2/NF45 ($P=1\times10^{-125}$), Sox11 ($P=5\times10^{-27}$), Sox21 ($P=6\times10^{-38}$), MCG4 ($P=1\times10^{-58}$), MED7 ($P=2\times10^{-66}$), AP-1/c-jun ($P=3\times10^{-32}$), HEX ($P=3\times10^{-19}$) and ERG ($P=1\times10^{-110}$; a partial ERG sequence from sea urchin was previously reported by Qi et al., 1992). Several of these factors are expressed selectively by mesodermal cells in other systems (e.g. aristaless, NF45 and ERG) and two, MTA1 and ERG, have been implicated in the regulation of cell movements in metastatic and embryonic cells

(Nicolson and Moustafa, 1998; Herman et al., 1999; Vlaeminck-Guillem et al., 2000). In situ hybridization analysis showed that at least two of these transcription factors, ERG and aristaless, are expressed predominantly or exclusively by PMCs.

The EST analysis also identified many components of conserved signaling pathways that play important roles in early sea urchin development. With respect to the Wnt signaling pathway (reviewed by Kikuchi, 2000; Peifer and Polakis, 2000), we identified (1) a frizzled-related protein most similar to mouse frizzled-1 ($P=4\times10^{-27}$); (2) axin, a key scaffolding protein that regulates the phosphorylation and degradation of β-catenin ($P=5\times10^{-47}$); (3) a Wnt protein most similar to vertebrate Wnt-8 ($P=6\times10^{-84}$); (4) protein kinase B/Akt, a kinase that phosphorylates and inactivates GSK3 ($P=1\times10^{-115}$); and (5) regulatory subunit B of protein phosphatase 2A (PP2A), another component of the multi-protein complex that regulates β-catenin phosphorylation ($P=3\times10^{-61}$). Differential nuclearization of β-catenin along the animal-vegetal axis plays an important role in patterning the early sea urchin embryo (reviewed by Davidson et al., 1998; Angerer and Angerer, 2000; Ettensohn and Sweet, 2000). The identification of these components of the β-catenin pathway will facilitate further analysis of the regulation of β-catenin nuclearization.

With respect to the Notch pathway (reviewed by Artavanis-Tsakonas et al., 1999), we identified for the first time in the sea urchin a putative Notch ligand, a homolog of the protein Delta ($P=1\times10^{-135}$). Notch signaling is required for specification of non-skeletogenic mesoderm (NSM), and activation of Notch depends on inductive signals from micromere progeny (Sherwood and McClay, 1999; Sweet et al., 1999). It has therefore been speculated that micromere descendants might express a ligand for the Notch receptor (Sweet et al., 1999). The identification of sea urchin Delta in our PMC EST analysis supports this hypothesis. The EST analysis also led to the identification of sea urchin homologs of two proteins that regulate the post-translational processing of Notch, TNFα-converting enzyme ($P=3\times10^{-18}$) and presenilin I ($P=3\times10^{-31}$; Chan and Jan, 1999; Brou et al., 2000).

Considerable progress is currently being made in elucidating molecular pathways that pattern the early sea urchin embryo (see Davidson et al., 1998; Angerer and Angerer, 2000; Ettensohn and Sweet, 2000). The specification of PMC fate normally requires the presence of β-catenin in micromeres (Emily-Fenouil et al., 1998; Wikramanayake et al., 1998; Logan et al., 1999) and may also require the zygotic activation of an ETS transcription factor in the micromere lineage at the late blastula stage (Kurokawa et al., 1999). PMC specification is also linked to a change in the properties of the vegetal cortex at the eight-cell stage and/or the unequal cell division that produces the micromeres (reviewed by Ettensohn and Sweet, 2000). An apparently complete program of PMC specification can also be elicited in other cells of the early embryo, in some lineages even as late as the mid-late gastrula stage, by experimentally perturbing cellular interactions (Ettensohn, 1992; McClay and Logan, 1996). Ultimately, both normal and regulative pathways of PMC fate specification must be linked to the specific activation of the downstream effector molecules that execute the remarkable morphogenetic program of PMCs.

APPENDIX

Proteins identified by BLAST analysis

Cell cycle, cell growth and cell death
Alix (ALG-2-interacting protein)
Anaphase-promoting complex, subunit 10
Bcl-X (apoptosis regulator)
BTG1 (B-cell translocation gene 1 protein)
BUB-3
Cdc5
Cdc6
Cdc45
Cdc47
Chromodomain helicase DNA binding protein 3
Cyclin A
Cyclin B
Cyclin C
Cyclin D-interacting protein
Cyclin K
Cyclin 1
Cyclin-dependent kinase 2
Cyclin-dependent kinase 3
Cyclin-dependent kinase 8
Cyclin-dependent kinase inhibitor 3
DAD ('defender against cell death') 1
DNA helicase
DNA ligase III
DNA polymerase, α subunit
DNA polymerase, β subunit
DNA polymerase, epsilon subunit
ERCC-6 (excision-repair protein)
Fizzyl
GADD45
Histone acetyltransferase
Histone H1 (cleavage stage)
Histone H1 (embryonic)
Histone H2A (cleavage stage)
Histone H2A variant
Histone H2B (cleavage stage)
Histone H2B (embryonic)
Histone H3 (embryonic)
Inner centromere protein
Mad2 (spindle assembly checkpoint protein)
MOB1 (mitosis and ploidy protein)
Nim ('never-in-mitosis')-related kinase
PCD6 (programmed cell death protein)
Prohibitin
PRB1 (pRB-associated protein)
RAD1
RAD21
RAD23
RAD51
RCC1 (regulator of chromatin condensation 1)
RecQ (DNA helicase)
Replication origin recognition complex, subunit 4
RBP2 (pRB binding protein 2)
SUDD protein
Topoisomerase
UV-damaged DNA binding factor
XRCC ('x-ray repair cross-complementing' protein) 3

Cell signaling, growth factors, kinases and phosphatases
ACK protein kinase
Adenylyl cyclase
Axin
BMP2/4 (univin)
Calmodulin
CAM kinase I
Casein kinase I, α
Casein kinase I, γ
Cysteine-rich FGF receptor
Δ
DVR-1 (Vg1-like)

Frizzled-1
G protein, β1 subunit
GTPase-activating protein
HP28 (PDGF-associated protein)
Inositol 1,4,5-triphosphate-binding protein
IRE1 (ER kinase)
JNK protein kinase
MAPKAPK-4
MAPK phosphatase
Myotubularin (dual specificity phosphatase)
Notch- like protein
Nucleoside diphosphate kinase B
PERK (ER ser/thr kinase)
Phosphatidyl inositol-4-phosphate-5-kinase
Phosphorylase B kinase, α subunit
Phosphorylase B kinase, γ subunit
Phosphotyrosyl phosphatase activator
Pim-3/KID-1 kinase
Pleiotrophin (heparin-binding growth factor)
Pre-B-cell colony-enhancing factor
Presenilin I
Protein kinase, 5'-AMP-activated
Protein kinase B/Akt
Protein kinase C inhibitor, 14-3-3 protein
Protein phosphatase 1, γ subunit
Protein phosphatase 2A, 74 kDa regulatory subunit
Protein phophatase 2A inhibitor (SET protein)
Protein phosphatase 2C
Protein phosphatase 4, regulatory subunit 1
Protein phosphatase with EF-hands
Protein ser/thr kinase 11 (PAR-4)
Protein ser/thr kinase, RING3
Protein tyrosine kinase 9
Protein tyrosine phosphatase
Protein tyrosine phosphatase receptor interacting protein (liprin)
PTEN tumor suppressor
Rac
Raf
Rap1b
Regulator of protein phosphatase 4
Rho1
Scavenger receptor, cysteine-rich
SMAD5
Src-type protein kinase
SpAN protease
TNFα converting enzyme
TRAF (tumor necrosis factor receptor associated factor)
TRAP170 (thyroid hormone receptor associated factor)
Wnt-8
Vav-2 oncogene

Channels/transporters and their regulators
ABC transporter
ADP/ATP translocase
Ammonium transporter
Anion exchange protein, AE-2
Annexin VI
Cationic amino acid transporter
Cystinosin
Glutamate receptor (AMPA-type)
Glycine transporter
L-amino acid transporter, LAT-1
Lysosomal proton pump ATPase
Lysosomal proton pump ATPase, δ subunit
MTRP (Golgi 4-transmembrane spanning transporter)
N-type Ca^{2+} channel, α1 subunit
Na^+/Ca^{2+} exchange protein
Na^+/H^+ exchange regulatory factor 2
Na^+/K^+ ATPase, α chain
Na^+/K^+ ATPase, β chain
Na^+/phosphate cotransporter
Organic anion transporter
Organic cation transporter
Porin

Proline transporter
SAP97 (discs-large homolog, PDZ protein)
SLOB protein (K^+ channel-interacting protein)
Sodium bicarbonate transporter
Sodium-dependent phosphate transporter, type II
Stomatin
Sulfonylurea receptor 2B
Tetracycline transporter-like protein

Cytoskeleton, cell adhesion and cell motility
Abp1 (SH3P7)
Actin, muscle-specific
Actin, cytoskeletal
Actin-binding LIM protein
Actin-like protein, 13E
Actin-like protein, BAF53
A-actinin
Anillin
Ankyrin
p21-Arc
p41-Arc
Arp1/centractin
Arp3
Arp6p
Attractin (CUB family of adhesion/guidance molecules)
Bicaudal-D
Bystin
Canoe
Cdc10 (septin 1)
Cdc42
Cdc42-interacting protein 4
Centrin
Cofilin
Crumbs
Cut1
Del-1 (integrin-binding protein)
Discoidin-domain receptor tyrosine kinase
DOCK 1/myoblast city
Dynactin, subunit p25
Dynein heavy chain
Dynein heavy chain, isotype 3A
Dynein light chain
Dyskerin/CBF5 (centromere/MT-binding protein)
EWAM (actin binding protein)
Filamin
Flamingo (protocadherin)
Flightless I
Interaptin
Kelch-motif protein
Kinesin-like protein 1
Kinesin heavy chain
LASP-1
MAP (77 kDa)
Moesin
Myosin heavy chain, nonmuscle
Myosin heavy chain, β
Myosin light chain kinase
N-WASP
Outer dense fiber protein 2
PAK-interacting exchange factor
Profilin
Prominin
p55-related MAGUK protein (multiple PDZ-domain protein)
Roadblock
Semaphorin VIa
Slit-2
Spectrin, α chain
Spectrin, β chain
Syndecan-2
Syntenin (syndecan-binding PDZ protein)
Talin
Tektin B1
Tensin
Tetraspan CD-53
Tetraspan NET-4
Tetraspan NET-5

2624   X. Zhu and others

T-plastin
Tropomyosin
Tubulin, α

Extracellular matrix
Choriogenin
Coagulation factor V
Collagen, α2 (IV)
Collagen, α3 (IV)
ECM18
Fibrillin
Fibrinogen-related
Fibronectin
Fibropellin Ia
Fibropellin Ib
Fibropellin II
F-spondin
Glypican (HSPG)
Heparanase
HLC-32
Hyalin
Laminin, α chain
Laminin-like protein
Matrix metalloprotease 1 (MMP-1, collagenase)
Matrix metalloprotease 15 (MT2MMP-15,
   membrane-type)
Metalloelastase
Nidogen/entactin
Osteonectin
Perlecan

General metabolism and other enzymes
Acetyl-CoA-acyltransferase A
Acetyl-serotonin N-methyltransferase
Aconitate hydratase
Acyl-CoA dehydrogenase, long chain
Acyl-CoA oxidase, subunit II
Acyl-CoA-synthetase
ADE-2 (multifunctional protein)
Adenosylhomocysteinase
Adenylosuccinate lyase
ADP ribosyltransferase
Adenylosuccinate synthetase
Alanine aminotransferase
Aldehyde dehydrogenase 4
Aldose-1-epimerase
Aminocyclopropane carboxylate deaminase
Aminocyclopropane carboxylate synthase
AMP deaminase
Arginine methyltransferase
Arylsulfatase
Aspartate transaminase
ATPase N2B
ATP synthase, αsubunit
ATP synthase, β subunit
ATP synthase, γ subunit
ATP synthase F0 subunit 6
ATP synthase coupling factor 6
cGMP-specific phosphodiesterase
Chondroitin-6-sulfotransferase
CoA-thioester hydrolase
CTP synthase
Cytochrome B
Cytochrome B5
Cytochrome C
Cytochrome C-1 (heme protein precursor)
Cytochrome C oxidase, subunit I
Cytochrome C oxidase, subunit II
Cytochrome C oxidase, subunit III
Cytochrome C oxidase, subunit VIa
Cytochrome C reductase, Complex III, subunit 2
Cytochrome C reductase, Complex III, subunit 6
Cytochrome C reductase, iron-sulfur subunit
Cytochrome P450 monooxygenase
Deoxyribonuclease I
Diacylglycerol acyltransferase
2,4-Dienoyl-CoA-reductase
Dihydroflavonol-4-reductase

Dihydrofolate reductase
Electron transfer flavoprotein, β subunit
Endo-β1,4-glucanase
Enhancer of *Rudimentary* protein
Esterase
Fructose biphosphate aldolase
Galactosylceramide sulfotransferase
GalNAc-α2,6-sialyltransferase
GalNAc transferase I
GalNAc transferase II
GlcNAc sulfotransferase
GlcNAc transferase I
GlcNAc transferase II
Glutamate-cysteine ligase, regulatory subunit
Glutamine synthetase
Glutathione-S-transferase
Glycerol kinase
Glycine hydroxymethyltransferase
Glycogen debranching enzyme
GMP synthase
GPI anchor biosynthesis protein
HemeA:farnesyltransferase
Hexosaminidase
Hexosamindase B, β subunit
Holocytochrome C synthetase
Hydrolase, α/β
8-Hydroxyguanine glyosylase
Hydroxyisobutyrate dehydrogenase
Hydroxymethylglutaryl-CoA-synthase
Isocitrate dehydrogenase
Lipoic acid synthetase
Lipoyl transferase
Malate dehydrogenase
Maltase
Methyl sterol oxidase
NADH-dependent glutamate synthase
NADH deydrogenase, α/β subcomplex, 8 kDa
   subunit
NADH dehydrogenase, Complex I, B14 subunit
NADH dehydrogenase, Complex I, 15 kDa subunit
NADH dehydrogenase, Complex I, 19 kDa subunit
NADH dehydrogenase, Complex I, 20 kDa subunit
NADH dehydrogenase, Complex I, 39 kDa subunit
NADH dehydrogenase, Complex I, 75 kDa subunit
NADH dehydrogenase, subunit 1
NADH dehydrogenase, subunit 2
NADH dehydrogenase, subunit 4
NADH dehydrogenase, subunit 5
NAD(P) transhydrogenase
NAD(P)H steroid dehydrogenase
N-arginine dibasic convertase
Nitrogen fixation protein
Ornithine decarboxylase
Ornithine decarboxylase antizyme
Peptidyl-prolyl *cis-trans* isomerase
Peroxide reductase
Phosphate transfer protein B
Phosphatidylinositol transfer protein
Phosphoadenosine phosphosulfate synthase
Phospholipase B
Phospholipid scramblase, TRA1
Pyruvate dehydrogenase phosphatase
Retinoic acid hydrolase
Ribonucleotide reductase, small subunit
Selenophosphate synthase
Serine hydroxymethyltransferase
Sterol 14-α demethylase
Succinate dehydrogenase
Succinyl-CoA-synthetase, α subunit
Sucrose isomaltase
Tricarboxylate carrier
tRNA pseudouridine synthase A
Uronyl 2-sulfotransferase
Uroporphyrin decarboxylase

Membrane/protein trafficking
ADP ribosylation factor 1
ADP ribosylation factor 4

ADP ribosylation factor-directed GTPase activating
   protein
Amphiphysin
AP17 (clathrin coat assembly protein)
BET3
Clathrin, heavy chain
Coatomer, β subunit
Coatomer, γ subunit
Copine I
Copine III
Cytohesin
ER lumen protein retaining receptor
EXO70
Exportin(tRNA)
Importin α4
Importin β3 (RanBP5)
Importin 7 (RanBP7)
NSF-1
Nucleoporin p58
Prenylated Rab acceptor 1
Rab1
Rab2
Rab6
Rab7
Rab10
Rab11b
Rab21
Rab23
Rab26
RAMP4
RanBP1
Sar-1a
Sec13
Sec23
Sec24
Sec61
SNAP-25-interacting protein
Sorting nexin 4
Sorting nexin 12
Synaptophysin
Syntaxin 7
Transitional endoplasmic reticulum ATPase

Protein folding and degradation
Aminopeptidase A
Aminopeptidase N
Aqualysin
Capn7
Carboxypeptidase A
Cathepsin C
Cathepsin D
Cathepsin L
Chymotrypsin inhibitor 2
ClpX
Cullin 1
DnaJ
HSP70
HSP90
HSP90-binding protein, p23
HSP, 97 kD
HSPCO14
HSPCO17
Legumaine
Lysosomal carboxypeptidase
Prefoldin 1
Proteasome, δ chain
Proteasome, ε chain
Proteasome, 26S, subunit S3
Proteasome, 26S, subunit 6
Proteasome, 26S, subunit 7
Proteasome, 26S, subunit 9
Proteasome, 26S, subunit 10b
PRT1
Skp1
Smt3A (ubiquitin-like protein)
SUMO-1 activating enzyme, subunit 1
TCP-1, δ subunit
TCP-1, η subunit

Tubulin-specific chaperone
Ubiquitin
Ubiquitin carboxyterminal hydrolase
Ubiquitin-conjugating enzyme E2
Ubiquitin-conjugating enzyme
Ubiquitin-specific protease 3
Ubiquitin-specific protease 14

Protein synthesis (including translational regulators)
eIF-1α
eIF-1β
eEF-2
eIF-2C
eIF-2G
eIF-3, subunit 8
eIF-4A
eIF-5
Nanos/Xcat2
NAT1 translational repressor
Ribosomal protein, 60S subunit, L3
Ribosomal protein, 60S subunit, L5
Ribosomal protein, 60S subunit, L7a
Ribosomal protein, 60S subunit, L10
Ribosomal protein, 60S subunit, L11
Ribosomal protein, 60S subunit, L15
Ribosomal protein, 60S subunit, L30
Ribosomal protein, 60S subunit, L44
Ribosomal protein, 40S subunit, S3
Ribosomal protein, 40S subunit, S4
Ribosomal protein, 40S subunit, S15a
Ribosomal protein P0
Signal peptidase
Signal recognition particle, 14 kDa protein
Signal recognition particle, 54 kDa protein
Signal sequence receptor, β subunit
Signal sequence receptor, δ subunit
Signal sequence receptor, γ subunit
SUI1
tRNA synthetase, arginyl
tRNA synthetase, asparaginyl
tRNA synthetase, glycyl
tRNA synthetase, phenylalanyl
tRNA synthetase, valyl
WHO translational regulator

RNA metabolism
Abstrakt (DEAD box protein)
AU-rich RNA-binding protein
BAT1 (nuclear ATP-dependent RNA helicase)
CIRP (cold-inducible RNA-binding protein)
Cleavage stimulation factor, 50 kDa subunit
Crooked-neck protein
GRY-RBP (RNA-binding protein)
hnRNP protein F
hnRNP protein K
hnRNP protein L
hnRNP protein R
IGF-II mRNA-binding protein
Lark (RNA-binding protein)
MVP-100 (major vault protein)
NOP4
NSAP1 (RNA-binding protein)
Poly(A)-binding protein
Rae1
Ribonuclease
Ribonuclease P
RNA helicase, p68
RNase L inhibitor
RRP5
SAP ('spliceosome-associated protein')-130
SAP-145
SAP-155
SF3a (splicing factor)
Sm D3 (small nuclear ribonucleoprotein)
SnRNP assembly defective 1
SnRNP protein B
Splicing factor, arginine/serine-rich 7

Splicing factor, CC1.3
Splicing factor, KH-type
Splicing factor, polyU-binding
Splicing factor, SC 35
Splicing factor, proline/glutamine-rich
SR protein
SS-A, 60 kD ribonucleoprotein
U3 snRNP protein, 55 kDa
U5 snRNP protein, 116 kDa
U5 snRNP protein, dim1
U5 snRNP protein, Prp8
Zinc-finger RNA-binding protein

Spicule matrix proteins/MSP130
C-lectin
MSP130
MSP130-related 1
MSP130-related 2
PM27
SM30
SM37
SM37-related
SM50
SM50-related

Transcriptional regulators
AMY-1
AND-1
AP1/c-Jun
Aristaless
CHD 1 (chromodomain-helicase-DNA-binding protein)
DR1
ERG
Eyes absent
Glucose-regulated repressor
HEX
Hexamer-binding protein
HMED7
HMG1
HMG2
Interleukin enhancer binding factor 2
MAX-like bHLHZIP protein
MCG4
MTA-1
NFkB
NF-X1/*shuttle craft*
Nuclear receptor co-repressor 1
p100 transcriptional coactivator
p300 transcriptional cofactor JMY
RNA polymerase II, subunit RPB4
Scaffold attachment factor B
SNF2α
SNF2-related CBP activator protein
SoxB1
Sox11
Sox21
SpShr2
Stage-specific activator protein (SSAP)
SWI/SNF complex, 170 kDa subunit
TATA-binding protein-related factor 2
TFIIH
TFIIS
Zinc-finger protein 2
Zinc-finger protein 84
Zinc-finger protein 184 (Kruppel-related)
Zinc finger protein, KRAB
Zinc-finger protein, MEX-1
Zinc-finger protein, OZF

Other
Acinus L protein
Adrenal gland protein (lozenge-like)
Amyloid-β(A4) precursor
Androgen-induced protein
Aut1p
BCRP1 (breast cancer resistance protein)
BING4 (WD40 protein)

Brain protein I3
Brain protein 44
Butyrate response factor 1
Calcyclin-binding protein
Calcyphosine
Cell surface antigen 4F2
Coiled-coil protein
Degenerative spermatocyte protein (transmembrane)
EDRK-rich factor 2
EF-hand protein
EGF-repeat-containing transmembrane protein
Egg receptor for sperm
F protein
Fatvg
GARP ('glutamic acid-rich protein')
Glucose-regulated protein, 170 kDa
GOB4 (cement gland protein)
Growth arrest-specific gene 11
HAN11 (WD-repeat protein)
Hatching enzyme
HDL-binding protein/vigilin
HOOK1
Human surface glycoprotein
Huntingtin-interacting protein
Hydroxyproline-rich glycoprotein
IgGFc-binding protein
JTV1 protein
Lamin B
Lamin B receptor
LDL receptor-related protein 2
Leukemia virus receptor
LYAR (growth-regulating nucleolar Zn-finger protein)
MAMA
Meiosis-specific nuclear structural protein 1
Melastatin (down-regulated in metastatic melanoma)
Metallothionein
Nasopharyngeal epithelium-specific protein 1
Neuralized
Neuronal protein 15.6
Neuropathy target esterase (NPE)
Nuclear phosphoprotein p150TSP
Nucleolar protein P120
NMP200 (nuclear matrix protein)
Nuclear protein np95
Out-at-first
Oxysterol binding protein
Parkin
Peroxisomal biogenesis factor
p22, calcium-binding (EF hand) protein
Phosphoprotein α 4
PIWI
Polyposis locus protein 1
Polyprotein
Pregnancy-induced growth inhibitor
Prostate cancer overexpressed gene 1
Protein B
Rcd ('required for cell differentiation') 1
Reduced expression in cancer
Reticulocalbin
Reverse transcriptase-like protein
SART-1 tumor antigen
Selenoprotein W
SMCY
Spastin
Spec1a
Spec2c
Spec2d
Stromal interaction molecule 1
Testis-specific Zn-finger protein
TRABID protein
Transposase
Traube
Trichohyalin
Tubby
Tumor-suppressing subtransferable candidate 1
Wolf-Hirschhorn syndrome candidate 1

REFERENCES

Angerer, L. M., Chambers, S. A., Yang, Q., Venkatesan, M., Angerer, R. C. and Simpson, R. T. (1998). Expression of a collagen gene in mesenchyme lineages of the Strongylocentrotus purpuratus embryo. *Genes Dev.* **2**, 239-246.

Angerer, L. M. and Angerer, R. C. (2000). Animal-vegetal axis patterning mechanisms in the early sea urchin embryo. *Dev. Biol.* **218**, 1-12.

Artavanis-Tsakonas, S., Rand, M. D. and Lake, R. J. (1999). Notch signaling: cell fate control and signal integration in development. *Science* **284**, 770-776.

Benson, S., Sucov, H., Stephens, L., Davidson, E. and Wilt, F. (1987). A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. I.Authentication of the cloned gene and its developmental expression. *Dev. Biol.* **120**, 499-506.

Berg, L. K., Chen, S. W. and Wessel, G. M. (1996) An extracellular matrix molecule that is selectively expressed during development is important for gastrulation in the sea urchin embryo. *Development* **122**, 703-713.

Birnboim, H. C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **24**, 1513-1523.

Blankenship, J. and Benson, S. (1984). Collagen metabolism and spicule formation in sea urchin micromeres. *Exp. Cell Res.* **152**, 98-104.

Borisy, G. G. and Svitkina, T. M. (2000). Actin machinery: pushing the envelope. *Curr. Opin. Cell Biol.* **12**, 104-112.

Brou, C., Logeat, F., Gupta, N., Bessia, C., LeBail, O., Doedens, J. R., Cumano, A., Roux, P., Black, R. A. and Israel, A. (2000). A novel proteolytic cleavage involved in Notch signaling: the role of the disintegrin-metalloprotease TACE. *Mol. Cell* **5**, 207-216.

Carlier, M. F., Ducruix, A. and Pantaloni, D. (1999). Signalling to actin: the Cdc42-N-WASP-Arp2/3 connection. *Chem. Biol.* **6**, R235-R240.

Chan, Y. M. and Jan, Y. N. (1999). Presenilins, processing of beta-amyloid precursor protein, and Notch signaling. *Neuron* **23**, 201-204.

Conklin, D., Gilbertson, D., Taft, D. W., Maurer, M. F., Whitmore, T. E., Smith, D. L., Walker, K. M., Chen, L. H., Wattler, S., Nehls, M. and Lewis, K. B. (1999). Identification of a mammalian angiopoietin-related protein expressed specifically in liver. *Genomics* **62**, 477-482.

Costantini, F. D., Scheller, R. H., Britten, R. J. and Davidson, E. H. (1978). Repetitive sequence transcripts in the mature sea urchin oocyte. *Cell* **15**, 173-187.

Cox, K. H., Angerer, L. M., Lee, J.J., Davidson, E. H. and Angerer, R. C. (1986). Cell lineage-specific programs of expression of multiple actin genes during sea urchin embryogenesis. *J. Mol. Biol.* **188**, 159-172.

Davidson, E. H. (1986). *Gene Activity in Early Development*. 3rd edn. New York: Academic Press.

Davidson, E. H., Cameron, R. A. and Ransick, A. (1998). Specification of cell fate in the sea urchin embryo: summary and some proposed mechanisms. *Development* **125**, 3269-3290.

Decker, G. L. and Lennarz, W. J. (1988). Skeletogenesis in the sea urchin embryo. *Development* **103**, 231-247.

Emily-Fenouil, F., Ghiglione, C., Lhomond, G., Lepage, T. and Gache C. (1998). GSK3beta/shaggy mediates patterning along the animal-vegetal axis of the sea urchin embryo. *Development* **125**, 2489-2498.

Erickson, M. R., Galletta, B. J. and Abmayr, S. M. (1997). Drosophila myoblast city encodes a conserved protein that is essential for myoblast fusion, dorsal closure, and cytoskeletal organization. *J. Cell Biol.* **138**, 589-603.

Ettensohn, C. A. (1990). The regulation of primary mesenchyme cell patterning. *Dev. Biol.* **140**, 261-271.

Ettensohn, C. A. (1992). Cell interactions and mesodermal cell fates in the sea urchin embryo. *Development* **115 Suppl.**, 43-51.

Ettensohn, C. A. (1999). Cell movements in the sea urchin embryo. *Curr. Opin. Genet. Dev.* **9**, 461-465.

Ettensohn, C. A., Guss, K. A., Hodor, P. G. and Malinda, K. M. (1997). The morphogenesis of the skeletal system of the sea urchin embryo, In *Reproductive Biology of Invertebrates* (ed. J. R. Collier), pp. 225-265. New Delhi: Oxford and IBI Publishing.

Ettensohn, C.. A. and Sweet, H. C. (2000). Patterning the early sea urchin embryo. *Curr. Top. Dev. Biol.* **50**, 1-44.

Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186-194.

Ewing, B., Hillier, L., Wendl, M. C. and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175-185.

Farach-Carson, M. C., Carson, D. D., Collier, J. L., Lennarz, W. J., Park, H. R. and Wright, G. C. (1989). A calcium-binding, asparagine-linked oligosaccharide is involved in skeleton formation in the sea urchin embryo. *J. Cell Biol.* **109**, 1289-1299.

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.

Frasch, M. and Leptin, M. (2000). Mergers and acquisition: unequal partnerships in Drosophila myoblast fusion. *Cell* **102**, 127-129.

Galau, G. A., Britten, R. J. and Davidson, E. H. (1974). A measurement of the sequence complexity of polysomal messenger RNA in sea urchin embryos. *Cell* **2**, 9-20.

George, N. C., Killian, C. E. and Wilt, F. H. (1991). Characterization and expression of a gene encoding a 30.6-kDa Strongylocentrotus purpuratus spicule matrix protein. *Dev. Biol.* **147**, 334-342.

Giga, Y., Ikai, A. and Takahashi, K. (1987). The complete amino acid sequence of echinoidin, a lectin from the coelomic fluid of the sea urchinAnthocidaris crassispina. Homologies with mammalian and insect lectins. *J. Biol. Chem.* **262**, 6197-6203.

Gokudan, S., Muta, T., Tsuda, R., Koori, K., Kawahara, T., Seki, N., Mizunoe, Y., Wai, S. N., Iwanaga, S. and Kawabata, S. (1999). Horseshoe crab acetyl group-recognizing lectins involved in innate immunity are structurally related to fibrinogen. *Proc. Natl. Acad. Sci. USA* **96**, 10086-10091.

Granes, F., Garcia, R., Casaroli-Marano, R. P., Castel, S., Rocamora, N., Reina, M., Urena, J. M. and Vilaro, S. (1999). Syndecan-2 induces filopodia by active cdc42Hs. *Exp. Cell Res.* **248**, 439-456.

Grootjans, J. J., Zimmermann, P., Reekmans, G., Smets, A., Degeest, G., Durr, J and David, G. (1997). Syntenin, a PDZ protein that binds syndecan cytoplasmic domains. *Proc. Natl. Acad. Sci. USA* **94**, 13683-13688.

Guss, K. A. and Ettensohn, C. A. (1997) Skeletal morphogenesis in the sea urchin embryo: regulation of primary mesenchyme gene expression and skeletal rod growth by ectoderm-derived cues. *Development* **124**, 1899-1908.

Gustafson, T. and Wolpert, L. (1967). Cell movement and contact in sea urchin morphogenesis. *Biol. Rev.* **42**, 442-498.

Harkey, M. A. and Whiteley, A. H. (1983). The program of protein synthesis during the development of the micromere-primary mesenchyme cell line in the sea urchin embryo. *Dev. Biol.* **100**, 12-28.

Harkey, M. A., Klueg K., Sheppard, P. and Raff, R. A. (1995). Structure, expression, and extracellular targeting of PM27, a skeletal protein associated specifically with growth of the sea urchin larval spicule. *Dev. Biol.* **168**, 549-566.

Hemler, M. E. (1998). Integrin associated proteins. *Curr. Opin. Cell Biol.* **10**, 578-585.

Herman, M. A., Ch'ng, Q., Hettenbach, S. M., Ratliff, T. M., Kenyon, C. and Herman, R. K. (1999). EGL-27 is similar to a metastasis-associated factor and controls cell polarity and cell migration in *C. elegans*. *Development* **126**, 1055-1064.

Hertzler, P. L. and McClay, D. R. (1999). αSU2, an epithelial integrin that binds laminin in the sea urchin embryo. *Dev. Biol.* **207**, 1-13.

Hodor, P. G. and Ettensohn, C. A. (1998). The dynamics and regulation of mesenchymal cell fusion in the sea urchin embryo. *Dev. Biol.* **199**, 111-124.

Hodor, P. G., Illies, M. R., Broadley, S. and Ettensohn, C. A. (2000). Cell-substrate interactions during sea urchin gastrulation: migrating primary mesenchyme cells interact with and align extracellular matrix fibers that contain ECM3, a molecule with NG2-like and multiple calcium-binding domains. *Dev. Biol.* **222**, 181-194.

Holy, J., Wessel, G., Berg, L., Gregg, R. G. and Schatten G. (1995). Molecular characterization and expression patterns of a B-type nuclear lamin during sea urchin embryogenesis. *Dev. Biol.* **168**, 464-478.

Kaji, K., Oda, S., Shikano, T., Ohnuki, T., Uematsu, Y., Sakagami, J., Tada, N., Miyazaki, S. and Kudo, A. (2000). The gamete fusion process is defective in eggs of CD9-deficient mice. *Nat. Genet.* **24**, 279-282.

Katoh-Fukui, Y., Noce, T., Ueda, T., Fujiwara, Y., Hashimoto, N., Higashinakagawa, T., Killian, C. E., Livingston, B. T., Wilt, F. H., Benson, S. C., Sucov, H. M. and Davidson, E. H. (1991). The corrected structure of the SM50 spicule matrix protein of Strongylocentrotus purpuratus. *Dev. Biol.* **145**, 201-202.

Katow, H. (1995). Pamlin, a primary mesenchyme cell adhesion protein, in the basal lamina of the sea urchin embryo. *Exp. Cell Res.* **218**, 469-478.

Katow, H. and Hayashi, M. (1985). Role of fibronectin in primary mesenchyme cell migration in the sea urchin. *J. Cell Biol.* **101**, 1487-1491.

Katow, H., Yazawa, S. and Sofuku, S. (1990). A fibronectin-related synthetic peptide, Pro-Ala-Ser-Ser, inhibits fibronectin binding to the cell surface, fibronectin-promoted cell migration in vitro, and cell migration in vivo. *Exp. Cell Res.* **190**, 17-24.

Kikuchi A. (2000). Regulation of beta-catenin signaling in the Wnt pathway. *Biochem. Biophys. Res. Commun.* **16**, 243-248.

Killian, C. E. and Wilt, F. H. (1989). The accumulation and translation of a spicule matrix protein mRNA during sea urchin embryo development. *Dev. Biol.* **133**, 148-156.

Killian, C. E. and Wilt, F. H. (1996). Characterization of the proteins comprising the integral matrix *of Strongylocentrotus purpuratus* embryonic spicules. *J. Biol. Chem.* **271**, 9150-9159.

Kingsley, P. D, Angerer, L. M. and Angerer, R. C. (1993). Major temporal and spatial patterns of gene expression during differentiation of the sea urchin embryo. *Dev. Biol.* **155**, 216-234.

Kiyokawa, E, Hashimoto, Y., Kobayashi, S., Sugimura, H., Kurata, T. and Matsuda, M. (1998). Activation of Rac1 by a Crk SH3-binding protein, DOCK180. *Genes Dev.* **12**, 3331-3336.

Kurokawa, D., Kitajima, T., Mitsunaga-Nakatsubo, K., Amemiya, S., Shimada, H. and Akasaka, K. (1999). HpEts, an ets-related transcription factor implicated in primary mesenchyme cell differentiation in the sea urchin embryo. *Mech. Dev.* **80**, 41-52.

Leaf, D. S., Anstrom, J. A., Chin, J. E., Harkey, M. A, Showman, R. M. and Raff, R. A. (1987). Antibodies to a fusion protein identify a cDNA clone encoding msp130, a primary mesenchyme-specific cell surface protein of the sea urchin embryo. *Dev. Biol.* **121**, 29-40.

Lee, Y.-H., Britten, R. J. and Davidson, E. H. (1999a). *SM37*, a skeletogenic gene of the sea urchin embryo linked to the *SM50* gene. *Dev. Growth Differ.* **41**, 303-312.

Lee, Y.-H., Huang, G. M., Cameron, R. A., Graham, G., Davidson, E. H., Hood, L. and Britten, R. J. (1999b). EST analysis of gene expression in early cleavage-stage sea urchin embryos. *Development* **126**, 3857-3867.

Livingston, B. T., Shaw, R., Bailey, A. and Wilt, F. (1991). Characterization of a cDNA encoding a protein involved in formation of the skeleton during development of the sea urchin Lytechinus pictus. *Dev. Biol.* **148**, 473-480.

Logan, C. Y., Miller, J. R., Ferkowicz,M. J. and McClay, D. R. (1999). Nuclear beta-catenin is required to specify vegetal cell fates in the sea urchin embryo. *Development* **126**, 345-357.

Luo, L., Liao, Y. J., Jan, L. Y. and Jan, Y. N. (1994). Distinct morphogenetic functions of similar small GTPases: Drosophila Drac1 is involved in axonal outgrowth and myoblast fusion. *Genes Dev.* **8**, 1787-1802.

Malinda, K. M., Fisher, G. W. and Ettensohn, C. A. (1995). Four-dimensional microscopic analysis of the filopodial behavior of primary mesenchyme cells during gastrulation in the sea urchin embryo. *Dev. Biol.* **172**, 552-566.

Marsden, M. and Burke, R. D. (1997). Cloning and characterization of novel beta integrin subunits from a sea urchin. *Dev. Biol.* **181**, 234-245.

Marsden, M. and Burke, R. D. (1998). The βL integrin subunit is necessary for gastrulation in sea urchin embryos. *Dev. Biol.* **203**, 134-148.

McClay, D. R. and Logan, C. Y. (1996). Regulative capacity of the archenteron during gastrulation in the sea urchin. *Development* **122**, 607-616.

Miki, H., Sasaki, T., Takai, Y. and Takenawa, T. (1998). Induction of filopodium formation by a WASP-related actin-depolymerizing protein N-WASP. *Nature* **391**, 93-96.

Miller, J., Fraser, S. E. and McClay, D. R. (1995). Dynamics of thin filopodia during sea urchin gastrulation. *Development* **121**, 2501-2511.

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.

Nicolson, G. L. and Moustafa, A. S. (1998). Metastasis-associated genes and metastatic tumor progression. *In Vivo* **12**, 579-588.

Nolan, K. M., Barrett, K., Lu, Y., Hu, K. Q., Vincent, S. and Settleman, J. (1998). Myoblast city, the Drosophila homolog of DOCK180/CED-5, is required in a Rac signaling pathway utilized for multiple developmental processes. *Genes Dev.* **12**, 3337-3342.

Okazaki, K. (1965). Skeleton formation of sea urchin larvae. V. Continuous observation of the process of matrix formation. *Exp. Cell Res.* **40**, 585-596.

Okazaki, K. (1975a). Normal development of metamorphosis. In: *The Sea Urchin Embryo: Biochemistry and Morphogenesis* (ed. G. Czihak). New York: Springer-Verlag.

Okazaki, K. (1975b). Spicule formation by isolated micromeres of the sea urchin embryo. *Am. Zool.* **15**, 567-581.

Parr, B. A., Parks, A. L. and Raff, R. A. (1990). Promoter structure and protein sequence of msp130, a lipid-anchored sea urchin glycoprotein. *J. Biol. Chem.* **265**, 1408-1413.

Peifer, M. and Polakis, P. (2000). Wnt signaling in oncogenesis and embryogenesis–a look outside the nucleus. *Science* **287**, 1606-1609.

Pierschbacher, M. D. and Ruoslahti, E. (1984). Variants of the cell recognition site of fibronectin that retain attachment-promoting activity. *Proc.Natl. Acad. Sci. USA* **81**, 5985-5988.

Poustka, A. J., Herwig, R., Krause, A., Hennig, S., Meier-Ewert, S. and Lehrach, H. (1999). Toward the gene catalogue of sea urchin development: the construction and analysis of an unfertilized egg cDNA library highly normalized by oligonucleotide fingerprinting. *Genomics* **59**, 122-133.

Qi, S., Chen, Z. Q., Papas, T. S. and Lautenberger, J. A. (1992). The sea urchin erg homolog defines a highly conserved erg-specific domain. *DNA Seq.* **3**, 127-129.

Rohatgi, R., Ma, L., Miki, H., Lopez, M., Kirchhausen, T., Takenawa, T. and Kirschner, M. W. (1999). The interaction between N-WASP and the Arp2/3 complex links Cdc42-dependent signals to actin assembly. *Cell* **97**, 221-231.

Sherwood, D. R. and McClay, D. R. (1999). LvNotch signaling mediates secondary mesenchyme specification in the sea urchin embryo. *Development* **126**, 1703-1713.

Shrivastava, A., Radziejewski, C., Campbell, E., Kovac, L., McGlynn, M., Ryan, T. E., Davis, S., Goldfarb, M. P., Glass, D. J., Lemke, G. and Yancopoulos, G. D. (1997). An orphan receptor tyrosine kinase family whose members serve as nonintegrin collagen receptors. *Mol. Cell* **1**, 25-34.

Smith, L. C., Harrington, M. G., Britten, R. J. and Davidson, E. H. (1994). The sea urchin profilin gene is specifically expressed in mesenchyme cells during gastrulation. *Dev. Biol.* **164**, 463-474.

Smith, L. C., Chang, L., Britten, R. J. and Davidson, E. H. (1996). Sea urchin genes expressed in activated coelomocytes are identified by expressed sequence tags. Complement homologues and other putative immune response genes suggest immune system homology within the deuterostomes. *J. Immunol.* **156**, 593-602.

Solursh, M. (1986). Migration of primary mesenchyme cells. In *Developmental Biology: A Comprehensive Synthesis*. Vol. 2 (ed. L. Browder), pp. 391-431. New York: Plenum Press.

Suzuki, H. R, Reiter, R. S., D'Alessio, M., Di Liberto, M., Ramirez, F., Exposito, J. Y., Gambino, R. and Solursh, M. (1997). Comparative analysis of fibrillar and basement membrane collagen expression in embryos of the sea urchin, Strongylocentrotus purpuratus. *Zool. Sci.* **14**, 449-454.

Sweet, H. C., Hodor, P. G. and Ettensohn, C. A. (1999). The role of micromere signaling in Notch activation and mesoderm specification during sea urchin embryogenesis. *Development* **126**, 5255-5265.

Tachibana, I. and Hemler, M. E. (1999). Role of transmembrane 4 superfamily (TM4SF) proteins CD9 and CD81 in muscle cell fusion and myotube maintenance. *J. Cell Biol.* **146**, 893-904.

Vlaeminck-Guillem, V., Carrere, S., Dewitte, F., Stehelin, D., Desbiens, X. and Duterque-Coquillaud, M. (2000). The Ets family member Erg gene is expressed in mesodermal tissues and neural crests at fundamental steps during mouse embryogenesis. *Mech. Dev.* **91**, 331-335.

Vogel W. (1999). Discoidin domain receptors: structural relations and functional implications. *FASEB J.* **13 Suppl.**, S77-S82.

Vogel, W., Gish, G. D., Alves, F. and Pawson, T. (1997). The discoidin domain receptor tyrosine kinases are activated by collagen. *Mol. Cell* **1**, 13-23.

Vogel, W., Brakebusch, C., Fassler, R., Alves, F., Ruggiero, F. and Pawson, T. (2000). Discoidin domain receptor 1 is activated independently of beta(1) integrin. *J. Biol. Chem.* **275**, 5779-5784.

Wessel, G. M., Etkin, M.and Benson, S. (1991). Primary mesenchyme cells of the sea urchin embryo require an autonomously produced, nonfibrillar collagen for spiculogenesis. *Dev. Biol.* **148**, 261-272.

Wessel, G. M. and Chen, S. W. (1993). Transient, localized accumulation of alpha-spectrin during sea urchin morphogenesis. *Dev. Biol.* **155**, 161-171.

Wikramanayake, A. H., Huang, L. and Klein, W. H. (1998). Beta-Catenin is essential for patterning the maternally specified animal-vegetal axis in the sea urchin embryo. *Proc. Natl. Acad. Sci. USA* **95**, 9343-9348.

Wilt, F. H. (1999). Matrix and mineral in the sea urchin larval skeleton. *J. Struct. Biol.* **126**, 216-226.

Woods, A. and Couchman, J. R. (1998). Syndecans: synergistic activators of cell adhesion. *Trends Cell Biol.* **8**, 189-192.

Xu, X. and Doolittle, R. F. (1990). Presence of a vertebrate fibrinogen-like sequence in an echinoderm. *Proc. Natl. Acad. Sci. USA* **87**, 2097-2101.