Chapter IV

Summary

Approaches:

I have taken a series of approaches to study endomesodermal transcriptional regulators in the purple sea urchin. At first, we did not know many of the players now included in the EM-GRN, and few had their spatial-temporal patterns of expression studied at the time. Subtractive hybridization screens using macroarrays identified several new genes overexpressed in the endomesoderm (for an example see Ransick et al., 2002). A large number of gene perturbations, including knock-downs using MASO and expression of dominant negative forms by mRNA injection, allowed for the exploration of the function of several genes expressed within the vegetal half of the embryo (Davidson et al., 2002a; 2002b).

Initially, I selected several transcription factors from different classes, all expressed in partially overlapping spatial domains during similar stages of development of the endomesoderm (Figures 1A and 1B), and I isolated lambda clones containing genomic sequences surrounding their exons. The idea was to compare the sequences between them and derive common regulators as well as unique features of the regulatory regions that explain the differences in their expression patterns. It very quickly became clear that this was too much to be taken by a single researcher. At the same time, new approaches were developed both in terms of computational tools as well as molecular resources for sea urchin genomics that greatly facilitated my research (Rast et al., 2000; Cameron et al., 2000; Brown et al., 2002).

I proceeded to focus on a zinc finger transcription factor, *Spblimp1/krox*, a gene first expressed in the endomesoderm and subsequently in the endoderm of the purple sea urchin. *Spblimp1/krox* is an alternatively transcribed gene with two isoforms described to date. *blimp1/krox1a* (Figure 2) is the isoform expressed starting at gastrulation in the mid and hindgut. *blimp1/krox1b* (Figure 3) is the isoform expressed during early development in the endomesoderm. (Livi and Davidson, 2006a, 2006b).

Initial studies utilizing an engrailed fusion with the DNA binding domain of *Spblimp1/krox,* B1/K-eng, indicated that it acted as a transcriptional regulator of other endomesodermal regulatory genes. Further analysis, including anti-sense morpholino knock-down of Blimp1/Krox, refined our understanding of its connections within the GRN by allowing us to establish whether it had a positive or negative input into its downstream targets. Among the network properties of Blimp1/Krox is a positive feedback loop involving *otx*, where Otx is a positive regulator of blimp1/krox and Blimp1/Krox is a positive regulator of *otx* (Davidson et al., 2002a; 2002b; Livi and Davidson, 2006a; Yuh et al., 2004). Another interesting linkage involves a negative autoregulatory loop in mesoderm cells where Blimp1/Krox represses its own transcription (Livi and Davidson, 2006a).

During this period, I also worked on the regulation of *Spblimp1/krox*. Initially, I performed phylogenetic footprinting analysis comparing BAC sequences from *S. purpuratus* and *L. variegatus* surrounding the *blimp1/krox* locus (Figure 4). I started by subcloning the most conserved patches of sequences into EpGFPII, an expression vector

containing the basal promoter of Endo16 driving the expression of GFP (Cameron et al., 2004). I noticed that only fragments containing sequences immediately next to the transcriptional start site for each isoform contained any transcriptional activity. In any event, the level of expression of all constructs was very low using linearized vector-based constructs for injection.

It became clear that a different strategy would be required. I proceeded to utilize the endogenous basal promoter of each alternatively transcribed isoform, *Spblimp1/krox1a* and *Spblimp1/krox1b*, including the 5' untranslated sequence driving the marker genes (either green fluorescent protein or chroramphenicol acetyl transferase) to test each conserved patch of sequences. I created my constructs using a fusion PCR strategy (Hobert, 2002) so that there were no bacterial vector sequences in the linear PCR product I injected. This piece of the work quickly led to a fragment just 5' of exon 1a that was able to recapitulate the late expression pattern of the *Spblimp1/krox1a* isoform (Livi and Davison, 2006b).

The regulation of the early isoform is much more complex. An important feature to note is that there is a proximal element with strong ubiquitous transcriptional activation function (Figure 5) that is then spatial-temporally restricted by a series of other more distal modules (Figure 6). Conserved sequence patches in the large *1a-1b* intron, 3' of position 50kb on the *Spblimp1/krox* BAC, control expression of the early form, as will be described elsewhere (J. Smith, C. Livi, and E. Davidson, unpublished data).

Why Build GRN Models

To understand the causal relationship between molecular players and developmental processes, it is necessary to understand their regulation. A thorough understanding of how a single gene is regulated is still insufficient to elucidate the series of necessary events leading up to territory specification. Dozens of transcription factors are required for the specification of a single territory in early embryogenesis (Levine and Davidson, 2005). Remembering all of their upstream regulators and downstream targets to generate the big picture view is a challenging task. In an age of developmental biology rich in experimental information, visualization tools become essential for a progression towards completeness (Longabaugh et al., 2005).

Thus, having all of the information in tabular format of the molecular consequences of perturbing the expression of any single gene or set of genes is not sufficient. Instead, a realistic model displaying as much of the biological detail as is possible needs to be constructed to aid in the understanding of the data. Through this organizational and logic system building step, many properties not yet noticed often "pop out". At the same time, holes in the model quickly become apparent, and experiments to fill them in can be designed and carried out as a reiterative process.

This kind of model is also very useful when trying to coordinate data acquired by different labs. A consensus of definitions and normalizations of the data can be carried out prior to their display and allow the data be directly comparable. Hopefully, being able to

compile multiple layers of data from many experimentalists into one place will lead to more informed decisions on how to proceed minimizing the need for duplicating efforts. Finally, models are invaluable for introducing the often complex data already acquired to new members of the scientific community who would often need to read and re-read dozens or even hundreds of scientific papers to accumulate all of the facts that can often be summarized in a few pages.

References:

Arnone, M.I., Davidson, E.H., 1997. The hardwiring of development: organization and function of genomic regulatory systems. Development. 124, 1851-1864.

Brown, C.T., Rust, A.G., Clarke, P.J., Pan, Z., Schilstra, M.J., De Buysscher, T., et al., 2002. New computational approaches for analysis of *cis*-regulatory networks. Dev. Biol. 246, 86-102.

Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., et al., 2000. A sea urchin genome project: sequence scan, virtual map, and additional resources. Proc. Natl. Acad. Sci. U S A. 97, 9514-9518.

Cameron, R.A., Oliveri, P., Wyllie, J., Davidson, E.H., 2004. cis-Regulatory activity of randomly chosen genomic fragments from the sea urchin. Gene Expression Patterns 4, 205–213.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Schilstra, M.J., Clarke, P.J., Rust, A.G., Pan, Z., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev Biol. 246, 162-190.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z., Schilstra, M.J., Clarke, P.J., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002. A genomic regulatory network for development. Science. 295, 1669-1678.

Hobert, O., 2002. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. Biotechniques. 32, 728-730.

Livi, C.B., Davidson, E.H., 2006. Expression and function of *blimp1/krox*, an alternatively transcribed regulatory gene of the sea urchin endomesoderm network. Dev. Biol. In press.

Livi, C.B., Davidson, E.H. Regulation of *Spblimp1/krox1a*: an alternatively transcribed isoform expressed in the mid and hindgut of the sea urchin gastrula. Gene Expr. Patterns. In press.

Longabaugh, W.J., Davidson, E.H., Bolouri, H., 2005. Computational representation of developmental genetic regulatory networks. Dev. Biol. 283, 1–16.

Ransick, A., Rast, J.P., Minokawa, T., Calestani, C., Davidson, E.H., 2002. New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. Dev Biol. 246, 132-147.

Rast, J.P., Amore, G., Calestani, C., Livi, C.B., Ransick, A., Davidson, E.H., 2000. Recovery of developmentally defined gene sets from high-density cDNA macroarrays. Dev Biol. 228, 270-286.

Revilla-i-Domingo, R., Davidson, E.H., 2003. Developmental gene network analysis. Int J Dev Biol. 47, 695-703.

Yuh, C.H., Dorman, E.R., Howard, M.L., Davidson, E.H., 2004. An otx *cis*-regulatory module: a key node in the sea urchin endomesoderm gene regulatory network. Dev. Biol. 269, 536-551.

Figure 1: WMISH of 25 h embryos using anti-sense probes targeting different genes expressed in the vegetal plate. (A) side view; (B) vegetal view. This figure shows a comparison of control and LiCl-treated embryos (15 mM and 30 mM) showing the effect on vegetal plate specific gene expression of *Spendo16* (Endo16), *Spgatae* (Gata e), *Spblimp1/krox* (Krox), and *Spfkhd1* (FoxB).

Figure 1A

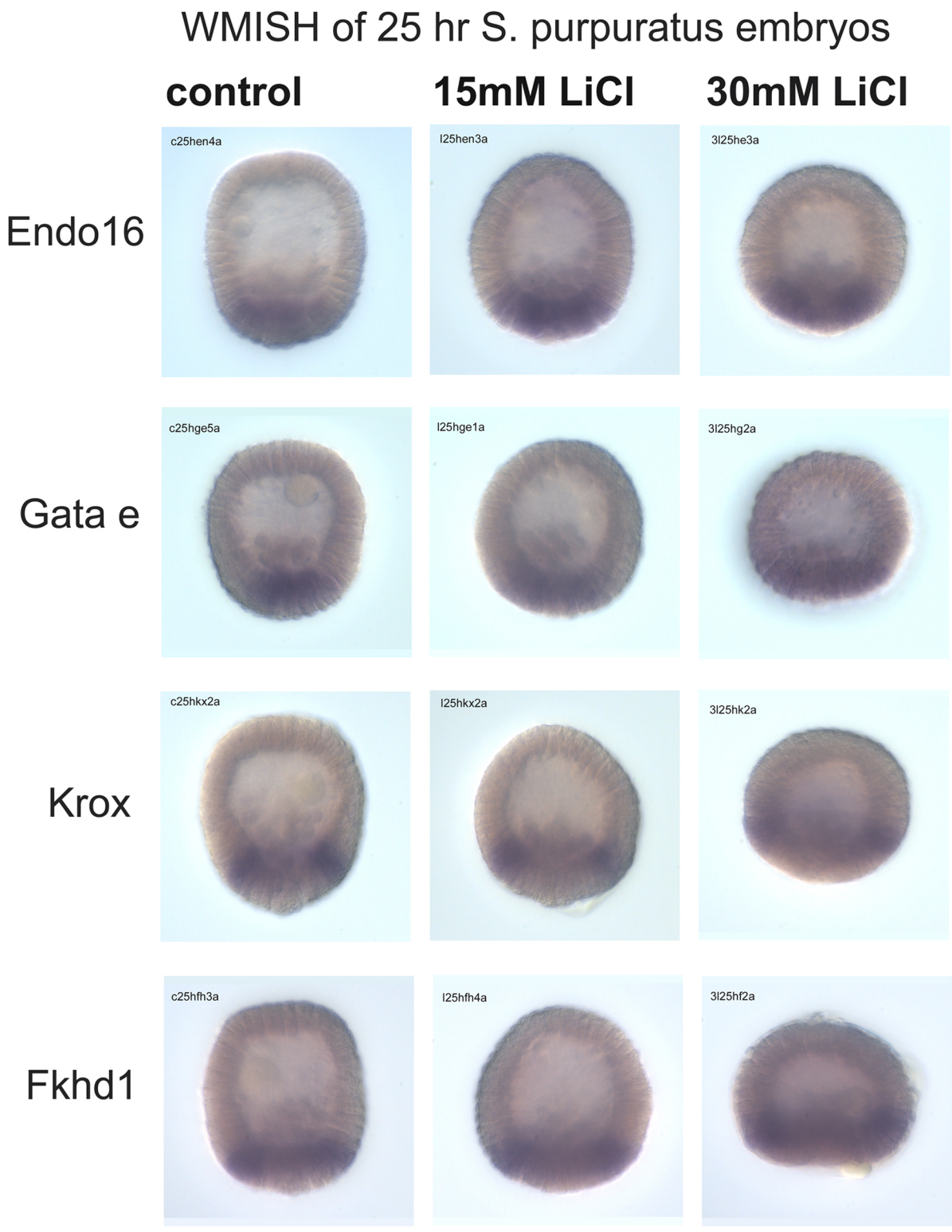
WMISH of 25 hr S. purpuratus embryos
control
15mM LiCl
30mM LiCl
Endo16
Gata e
Krox
Fkhd1
c25hen4a
l25hen3a
3l25he3a
c25hge5a
l25hge1a
3l25hg2a
c25hkx2a
l25hkx2a
3l25hk2a
c25hfh3a
l25hfh4a
3l25hf2a

Figure 1B

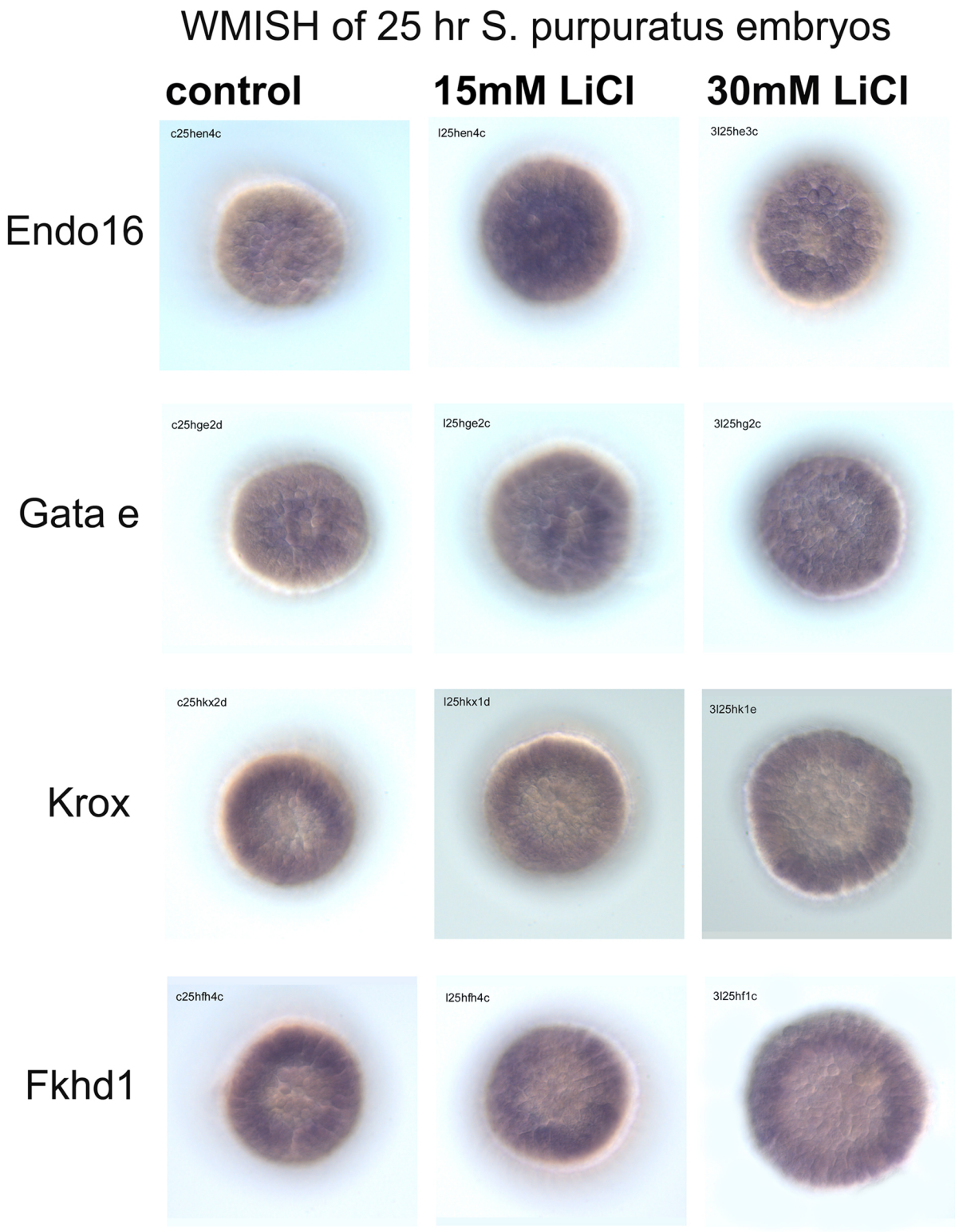
WMISH of 25 hr S. purpuratus embryos
control
15mM LiCl
30mM LiCl
Endo16
c25hen4c
l25hen4c
3l25he3c
Gata e
c25hge2d
l25hge2c
3l25hg2c
Krox
c25hkx2d
l25hkx1d
3l25hk1e
Fkhd1
c25hfh4c
l25hfh4c
3l25hf1c

Figure 2: *Spblimp1/krox1a* genbank entry from NCBI. Gene description, amino acid and nucleotide sequence of the late transcriptional isoform.

Figure 2 – Page 1

```
LOCUS       DQ225099     3974 bp    mRNA    linear   INV 01-DEC-2005
DEFINITION  Strongylocentrotus purpuratus Blimp1/Krox1a late isoform
mRNA, complete cds, alternatively spliced.
ACCESSION   DQ225099
VERSION     DQ225099.1  GI:81174740
KEYWORDS    .
SOURCE      Strongylocentrotus purpuratus
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 3974)
  AUTHORS   Livi,C.B. and Davidson,E.H.
  TITLE     Expression and function of blimp1/krox, an alternatively
            transcribed regulatory gene of the sea urchin endomesoderm
network
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3974)
  AUTHORS   Livi,C.B. and Davidson,E.H.
  TITLE     Regulation of Spblimp1/krox1a, an alternatively transcribed
isoform expressed in the mid and hindgut of the sea urchin gastrula
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3974)
  AUTHORS   Livi,C.B. and Davidson,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2005) Biology, California Institute of
            Technology, MSC: 156-29, 1200 E. California Blvd.,
Pasadena, CA 91125, USA
FEATURES             Location/Qualifiers
     source          1..3974
                     /organism="Strongylocentrotus purpuratus"
                     /mol_type="mRNA"
                     /db_xref="taxon:7668"
     CDS             159..2270
                     /note="SET domain-containing zinc finger
transcription
                     factor isoform 1a; SpBlimp1/Krox1a; alternatively
spliced"
                     /codon_start=1
                     /product="Blimp1/Krox1a late isoform"
                     /protein_id="ABB58913.1"
                     /db_xref="GI:81174741"


/translation="MGCNDNAVPSDMTEIDFQKKCIYEVKDPERDPASTEPRSVTSLP

GNLRFKRRNEQVVGVLADEYIPKGTRFGPLAGKIYREEEVPEDANKKYFWRIYNGQKF

VHYIDGFNKKRSNWMQYVSPAGASAEQNIVACQYKTDIFFYTTKAVEPGQELFVWYCT

EFAKRLEKPVVASRASPVMKREPAVPVSAPSEDVLDFSRRPHQLSSTVNDVKSSFPMS

SSLARGSSDHLAAEAGVSKTLYPSLPPTFAFSHRVGCSPPGTDIKQEPPSPTSTSPMG

SSTLLPTQPAIPMSRMYPTSGAPVVVPLPMYGLSPNYISTAGGLTLLASQPIYSMPSS
```

Figure 2 – Page 2

RMMDSRSLEMTKNRHIQPKPLLNQPLADSNGYLSEHEASMKQDVSRRVRRSSNPNYGH

RALGYELPRRNGKIVYECNVCKREFGQLSNLKVHLRVHTGEKPFSCDLCGKGFTQFAH

LQKHHLVHTGEKPHCCHVCEKRFSSTSNLKTHMRLHSGDKPYNCKECPAKFNQQVHLR

LHKKAHLEGNVELFMDFSDDADTPSTDGSVIAGLQDHLASTEVTDYFGNTDVTSPSSR

DRLDDDEDDEGKLIIQESGRSLETSPMGYMFRSDAMDETSQELCDSPSTGITASSCNG

VEDDAMEVTKTDQDQSAAFESKDNVALPLQSELHEASSHATKRSHPELEFGSRSKLSK

SNPALKLENVVQKILNRTNKS"

ORIGIN

```
1    tcaagttgga aagcacttgc ttgctgttac caccacagaa aagagaaaaa gcgaaaaaaa
61   gaagtctaaa gaaattcccc gattgaaacc gcaagctatt ttttctacgc ttcaacgagg
121  tcttcgatca ggtcttcgac agccgccacc gcctggccat ggggtgcaac gacaacgccg
181  tgccgtccga catgaccgag atcgacttcc agaagaaatg catttacgag gtcaaggacc
241  ccgaacggga ccccgcgagc acggagccga ggtcggtgac gtcattacca ggcaacctcc
301  gattcaagcg ccggaatgaa caggtggtcg gtgtgttagc agatgagtac ataccaaaag
361  gtacgcggtt cggaccccta gctgggaaga tttacaggga agaggaggtg ccggaagatg
421  ccaacaagaa atacttttgg aggatttaca atggccagaa gtttgtccac tacattgacg
481  gtttcaacaa gaagcgaagc aactggatgc agtacgtcag ccccgccgga gcttcggcgg
541  aacagaacat tgttgcctgc cagtacaaga ccgatatttt cttctacact acgaaagctg
601  tcgagccggg ccaggagctg ttcgtgtggt attgtaccga gttcgctaaa cgcctagaga
661  aaccagttgt cgcgtccaga gctagtccag ttatgaagag agaacccgcc gttcctgtct
721  ctgctccttc agaggatgtt ctcgacttca gtcgtcgtcc acaccaactg tcttccacag
781  tcaatgatgt caagtcttcc tttccgatgt cttcaagcct tgccagaggt tcttccgatc
841  accttgctgc tgaggccgga gtcagtaaga ccttgtaccc cagtcttcca cccaccttcg
901  ccttcagcca ccgggttgga tgcagcccac cgggtacgga catcaaacag gagccaccaa
961  gtccaacgtc cacttcccct atgggcagtt caaccctgtt gccgacccaa ccagccattc
1021 ctatgtcaag gatgtacccg accagtggtg ctcctgttgt tgtaccatta cctatgtatg
1081 gcctgtcacc aaactacatc agtactgcag gtggtctcac gttgcttgcc tcacagccca
1141 tctactctat gccgtccagc agaatgatgg attctcgctc tttggaaatg accaagaatc
1201 gccacataca gccaaagcct cttttaaatc aaccactcgc agattccaat ggctatcttt
1261 cagaacacga agccagcatg aagcaagatg tttccaggcg ggtcaggcgc agcagtaacc
1321 cgaactacgg ccatcgtgct cttggctacg agcttccacg ccgcaacggc aagattgtct
1381 acgaatgcaa tgtctgcaag agggaattcg ggcagctgtc gaacctcaag gtccacctcc
1441 gcgtccacac cggcgagaaa cccttctcct gcgatctttg cgggaagggc ttcacacagt
1501 tcgctcatct tcagaagcat catctcgttc acacagggga gaagccgcat tgctgtcacg
1561 tctgcgagaa gcggtttagc agcaccagca acctcaagac ccacatgcga ctccacagtg
1621 gggacaagcc ctacaactgc aaagagtgcc cagccaagtt caatcaacaa gtccatctca
1681 ggcttcacaa gaaggcacat cttgaaggca acgtcgaact cttcatggac ttcagcgacg
1741 atgctgacac cccaagcacc gacgggtccg tcatcgctgg tcttcaagat cacctggcca
1801 gcactgaggt caccgactat ttcggaaata ccgatgtgac gtcccctagc agtcgagatc
1861 gtctcgacga tgatgaggac gacgaaggca agttgatcat ccaagagagt gggagatcgc
1921 tggaaacatc acccatgggc tacatgtttc gatctgacgc catggacgaa acgtctcaag
1981 aactctgcga ctccccctcg actggaatca ccgcgtccag ctgcaacggc gtagaagacg
2041 acgcgatgga agtcactaaa accgaccaag accaatcagc agcctttgag agcaaagaca
2101 acgttgccct tccattgcaa tctgaactcc atgaagcatc atctcatgca accaaaagat
2161 cccatcctga actggagttc ggatcaagaa gcaagttatc caagtccaat cctgccctca
2221 aactcgagaa cgtcgtccag aagatactca atcgtacgaa caaatcatag aacttcttct
2281 ttagtcgaac aattttgatt ttatcagtgt gctcgctttg gctagtaaat gctaaattac
```

Figure 2 – Page 3

```
2341 caacataaga gaaatgctca agcttgaatc ttggttttcc gacttgttta atgtattatg
2401 gatttgcgat cagacttgaa taagcaattt ggatgctacg aacaatgtca gcacagaaac
2461 aagcataata atattagtca acagactcat aatattttgt tcgaaagaac gcatttgtat
2521 atatattgaa tattgccaac tgtacgatag gtgtatgaat aaatattaat catactgaaa
2581 tcattacgcc aaaataagcc gattatgact atcaatgagt tagtgccata ttctgccgaa
2641 tcatctttat tttgtataaa ttattctcta ttgatcatag ctaaaataag aataatacta
2701 tttttgtcaa attaatgatt gaaatttgga catgctctga atgttactat tttacctgtg
2761 ctttgtgatt tcaaatcagt tcacattagt taattaccga tttgaatacc tctataagac
2821 tttccggatc gaacggattc tctgtattca actttcatct tgaaaagtta aattatatat
2881 gtgtctttca aaaatgatct gtgttaacat tcaagtaatt atgtaaagag caggagtttt
2941 ctgcttcttt tgctttggag aaaattgatt taattatgat gtgattgcag aatgttttct
3001 gaagttttaa acgtttatta tagtctaccg aggtctaccg atgtaatatt attgtctttc
3061 aaaccaggaa gtcacaagca attatgaaat tgtctttcgc tgtataactg tctgatctgt
3121 agaagttgta ggattactca atacaatgga tgtaaaacgt gatcagaaag aatttcaaga
3181 aatcaaattt aaacaatatc aatatccttg ctcgtaaccc tttgaaacgt gcgaatatcg
3241 caaggacaaa tatcattatg tgaatacgat tttcaaataa aatgcaaaag ggtatttaaa
3301 tgacccttga gtcgatttct tgaaggcctt tgcgatcgtg atgtaagtta gagtatataa
3361 ttttacccaa gagtgccaaa acatgcagta gcggggtagc gatacattgc ggcccgaatt
3421 cacaaaggat gttgtcgaca aacctagttt tatgtgcgtc atatgacgcg tatcgacgcg
3481 tttttacaaa ccaaggggtc tatgtgcccc ccccccctcat tccagtatca tattataatc
3541 atgtgcacaa atgtaatgct cgaccgtcag cttgccgtgg tgaatgaaat attaatcaat
3601 taactaatta gtattgatat caagatattt tctaaactag acactagcga aacttttatg
3661 attcgaaccc tagtattcag ttatttatta atgatagtca caggcaaggt aattcaaaat
3721 aggatgtgta ttctcacgtg gtcttactat ttttaagata aatgttattg ttagaagaat
3781 agtttatttt ctatgtatat atatgtgatg ttctttaata caactcatgg gataatagca
3841 gtgtgcctag cagtgcgagg cgtgtatgta cattagatct tccaacagga ttgtgcctta
3901 tattgatgac atttttttgt ataaaattga aagtgacatt tttgccaata aaaattatgt
3961 ttatgtcaac agtg
//
```

Figure 3: *Spblimp1/krox1b* genbank entry from NCBI. Gene description, amino acid and nucleotide sequence of the early transcriptional isoform.

Figure 3 – Page 1

```
LOCUS       DQ177152  3974 bp  mRNA  linear INV 28-SEP-2005
DEFINITION  Spblimp1/krox1a, SET domain containing zinc
finger transcription factor, late isoform.
ACCESSION   DQ177152;
KEYWORDS    .
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata;
Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 3974)
  AUTHORS   Livi,C.B. and Davidson,E.H.
  TITLE     Regulation of Spblimp1/krox1a: an alternatively
transcribed isoform expressed in the mid and hindgut of the
sea urchin gastrula
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3974)
  AUTHORS   Livi,C.B. and Davidson,E.H.
  TITLE     EXPRESSION AND FUNCTION OF BLIMP1/KROX, AN
ALTERNATIVELY TRANSCRIBED REGULATORY GENE OF THE SEA URCHIN
ENDOMESODERM NETWORK
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3974)
  AUTHORS   Livi,C.B. and Davidson,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2005) Biology, California
Institute of Technology, MSC: 156-29, 1200 E. California
Blvd, Pasadena, CA 91125, USA
FEATURES             Location/Qualifiers
     source          1..3974
                     /organism="Strongylocentrotus
purpuratus"
                     /db_xref="taxon:7668"
     CDS             159..2270
                     /note="SET domain containing zinc
finger transcription factor, isoform 1a"
                     /codon_start=1
                     /product="SpBlimp1/Krox1a"

/translation="MGCNDNAVPSDMTEIDFQKKCIYEVKDPERDPASTEPRSVTSLP
GNLRFKRRNEQVVGVLADEYIPKGTRFGPLAGKIYREEEVPEDANKKYFWRIYNGQKF
VHYIDGFNKKRSNWMQYVSPAGASAEQNIVACQYKTDIFFYTTKAVEPGQELFVWYCT
EFAKRLEKPVVASRASPVMKREPAVPVSAPSEDVLDFSRRPHQLSSTVNDVKSSFPMS
SSLARGSSDHLAAEAGVSKTLYPSLPPTFAFSHRVGCSPPGTDIKQEPPSPTSTSPMG
```

Figure 3 – Page 2

SSTLLPTQPAIPMSRMYPTSGAPVVVPLPMYGLSPNYISTAGGLTLLASQPIYSMPSS
RMMDSRSLEMTKNRHIQPKPLLNQPLADSNGYLSEHEASMKQDVSRRVRRSSNPNYGH
RALGYELPRRNGKIVYECNVCKREFGQLSNLKVHLRVHTGEKPFSCDLCGKGFTQFAH
LQKHHLVHTGEKPHCCHVCEKRFSSTSNLKTHMRLHSGDKPYNCKECPAKFNQQVHLR
LHKKAHLEGNVELFMDFSDDADTPSTDGSVIAGLQDHLASTEVTDYFGNTDVTSPSSR
DRLDDDEDDEGKLIIQESGRSLETSPMGYMFRSDAMDETSQELCDSPSTGITASSCNG
VEDDAMEVTKTDQDQSAAFESKDNVALPLQSELHEASSHATKRSHPELEFGSRSKLSK
SNPALKLENVVQKILNRTNKS"

BASE COUNT 1158 a 918 c 834 g 1064 t
ORIGIN

```
1    tcaagttgga aagcacttgc ttgctgttac caccacagaa aagagaaaaa gcgaaaaaaa
61   gaagtctaaa gaaattcccc gattgaaacc gcaagctatt ttttctacgc ttcaacgagg
121  tcttcgatca ggtcttcgac agccgccacc gcctggccat ggggtgcaac gacaacgccg
181  tgccgtccga catgaccgag atcgacttcc agaagaaatg catttacgag gtcaaggacc
241  ccgaacggga ccccgcgagc acggagccga ggtcggtgac gtcattacca ggcaacctcc
301  gattcaagcg ccggaatgaa caggtggtcg gtgtgttagc agatgagtac ataccaaaag
361  gtacgcggtt cggaccccta gctgggaaga tttacaggga agaggaggtg ccggaagatg
421  ccaacaagaa atacttttgg aggatttaca atggccagaa gtttgtccac tacattgacg
481  gtttcaacaa gaagcgaagc aactggatgc agtacgtcag ccccgccgga gcttcggcgg
541  aacagaacat tgttgcctgc cagtacaaga ccgatatttt cttctacact acgaaagctg
601  tcgagccggg ccaggagctg ttcgtgtggt attgtaccga gttcgctaaa cgcctagaga
661  aaccagttgt cgcgtccaga gctagtccag ttatgaagag agaacccgcc gttcctgtct
721  ctgctccttc agaggatgtt ctcgacttca gtcgtcgtcc acaccaactg tcttccacag
781  tcaatgatgt caagtcttcc tttccgatgt cttcaagcct tgccagaggt tcttccgatc
841  accttgctgc tgaggccgga gtcagtaaga ccttgtaccc cagtcttcca cccaccttcg
901  ccttcagcca ccgggttgga tgcagcccac cgggtacgga catcaaacag gagccaccaa
961  gtccaacgtc cacttcccct atgggcagtt caaccctgtt gccgacccaa ccagccattc
1021 ctatgtcaag gatgtacccg accagtggtg ctcctgttgt tgtaccatta cctatgtatg
1081 gcctgtcacc aaactacatc agtactgcag gtggtctcac gttgcttgcc tcacagccca
1141 tctactctat gccgtccagc agaatgatgg attctcgctc tttggaaatg accaagaatc
1201 gccacataca gccaaagcct cttttaaatc aaccactcgc agattccaat ggctatcttt
1261 cagaacacga agccagcatg aagcaagatg tttccaggcg ggtcaggcgc agcagtaacc
1321 cgaactacgg ccatcgtgct cttggctacg agcttccacg ccgcaacggc aagattgtct
1381 acgaatgcaa tgtctgcaag agggaattcg ggcagctgtc gaacctcaag gtccacctcc
1441 gcgtccacac cggcgagaaa cccttctcct gcgatctttg cgggaagggc ttcacacagt
1501 tcgctcatct tcagaagcat catctcgttc acacagggga gaagccgcat tgctgtcacg
1561 tctgcgagaa gcggtttagc agcaccagca acctcaagac ccacatgcga ctccacagtg
1621 gggacaagcc ctacaactgc aaagagtgcc cagccaagtt caatcaacaa gtccatctca
1681 ggcttcacaa gaaggcacat cttgaaggca acgtcgaact cttcatggac ttcagcgacg
1741 atgctgacac cccaagcacc gacgggtccg tcatcgctgg tcttcaagat cacctggcca
1801 gcactgaggt caccgactat ttcggaaata ccgatgtgac gtcccctagc agtcgagatc
1861 gtctcgacga tgatgaggac gacgaaggca agttgatcat ccaagagagt gggagatcgc
1921 tggaaacatc acccatgggc tacatgtttc gatctgacgc catggacgaa acgtctcaag
1981 aactctgcga ctccccctcg actggaatca ccgcgtccag ctgcaacggc gtagaagacg
2041 acgcgatgga agtcactaaa accgaccaag accaatcagc agcctttgag agcaaagaca
2101 acgttgccct tccattgcaa tctgaactcc atgaagcatc atctcatgca accaaaagat
2161 cccatcctga actggagttc ggatcaagaa gcaagttatc caagtccaat cctgccctca
2221 aactcgagaa cgtcgtccag aagatactca atcgtacgaa caaatcatag aacttcttct
2281 ttagtcgaac aattttgatt ttatcagtgt gctcgctttg gctagtaaat gctaaattac
```

Figure 3 – Page 3

```
2341 caacataaga gaaatgctca agcttgaatc ttggttttcc gacttgttta atgtattatg
2401 gatttgcgat cagacttgaa taagcaattt ggatgctacg aacaatgtca gcacagaaac
2461 aagcataata atattagtca acagactcat aatattttgt tcgaaagaac gcatttgtat
2521 atatattgaa tattgccaac tgtacgatag gtgtatgaat aaatattaat catactgaaa
2581 tcattacgcc aaaataagcc gattatgact atcaatgagt tagtgccata ttctgccgaa
2641 tcatctttat tttgtataaa ttattctcta ttgatcatag ctaaaataag aataatacta
2701 tttttgtcaa attaatgatt gaaatttgga catgctctga atgttactat tttacctgtg
2761 ctttgtgatt tcaaatcagt tcacattagt taattaccga tttgaatacc tctataagac
2821 tttccggatc gaacggattc tctgtattca actttcatct tgaaaagtta aattatatat
2881 gtgtctttca aaaatgatct gtgttaacat tcaagtaatt atgtaaagag caggagtttt
2941 ctgcttcttt tgctttggag aaaattgatt taattatgat gtgattgcag aatgttttct
3001 gaagttttaa acgtttatta tagtctaccg aggtctaccg atgtaatatt attgtctttc
3061 aaaccaggaa gtcacaagca attatgaaat tgtctttcgc tgtataactg tctgatctgt
3121 agaagttgta ggattactca atacaatgga tgtaaaacgt gatcagaaag aatttcaaga
3181 aatcaaattt aaacaatatc aatatccttg ctcgtaaccc tttgaaacgt gcgaatatcg
3241 caaggacaaa tatcattatg tgaatacgat tttcaaataa aatgcaaaag ggtatttaaa
3301 tgacccttga gtcgatttct tgaaggcctt tgcgatcgtg atgtaagtta gagtatataa
3361 ttttacccaa gagtgccaaa acatgcagta gcggggtagc gatacattgc ggcccgaatt
3421 cacaaaggat gttgtcgaca aacctagttt tatgtgcgtc atatgacgcg tatcgacgcg
3481 tttttacaaa ccaaggggtc tatgtgcccc ccccccctcat tccagtatca tattataatc
3541 atgtgcacaa atgtaatgct cgaccgtcag cttgccgtgg tgaatgaaat attaatcaat
3601 taactaatta gtattgatat caagatattt tctaaactag acactagcga aacttttatg
3661 attcgaaccc tagtattcag ttatttatta atgatagtca caggcaaggt aattcaaaat
3721 aggatgtgta ttctcacgtg gtcttactat ttttaagata aatgttattg ttagaagaat
3781 agtttatttt ctatgtatat atatgtgatg ttctttaata caactcatgg gataatagca
3841 gtgtgcctag cagtgcgagg cgtgtatgta cattagatct tccaacagga ttgtgcctta
3901 tattgatgac atttttttgt ataaaattga aagtgacatt tttgccaata aaaattatgt
3961 ttatgtcaac agtg
//
```

Figure 4: Comparative Family Relations II (FRII) analysis of genomic BAC sequences surrounding *Spblimp1/krox* and *Lvblimp1/krox*. (A) BAC maps of the *blimp1/krox* locus and its two transcriptional units, modified from Livi and Davidson (2006a). Boxes indicate the location of exons, which are numbered. The *Spblimp1/krox* BAC, 163O19, is 94 kb long. The *Lvblimp1/krox* BAC, 60B16, is 54 kb long. (B) FRII Analysis of BAC sequences in region of overlap between the *Sp* and *Lv* BACs. A pairwise view is reproduced, using a 50 bp window at 70% similarity. Top line represents *Spblimp1/krox* BAC, while the bottom line represents *Lvblimp1/krox*. Cyan-colored boxes indicate the position of *blimp1/krox* exons. Blue-colored lines between sequences indicate patches where more than 70% of the nucleotides within the 50 base pair window are conserved. The structure of the spliced *Spblimp1/krox1a* cDNA is shown above, while the structure of the predicted *Lvblimp1/krox1a* cDNA is shown below the FRII analysis (modified from Livi and Davidson, 2006b).

Figure 4

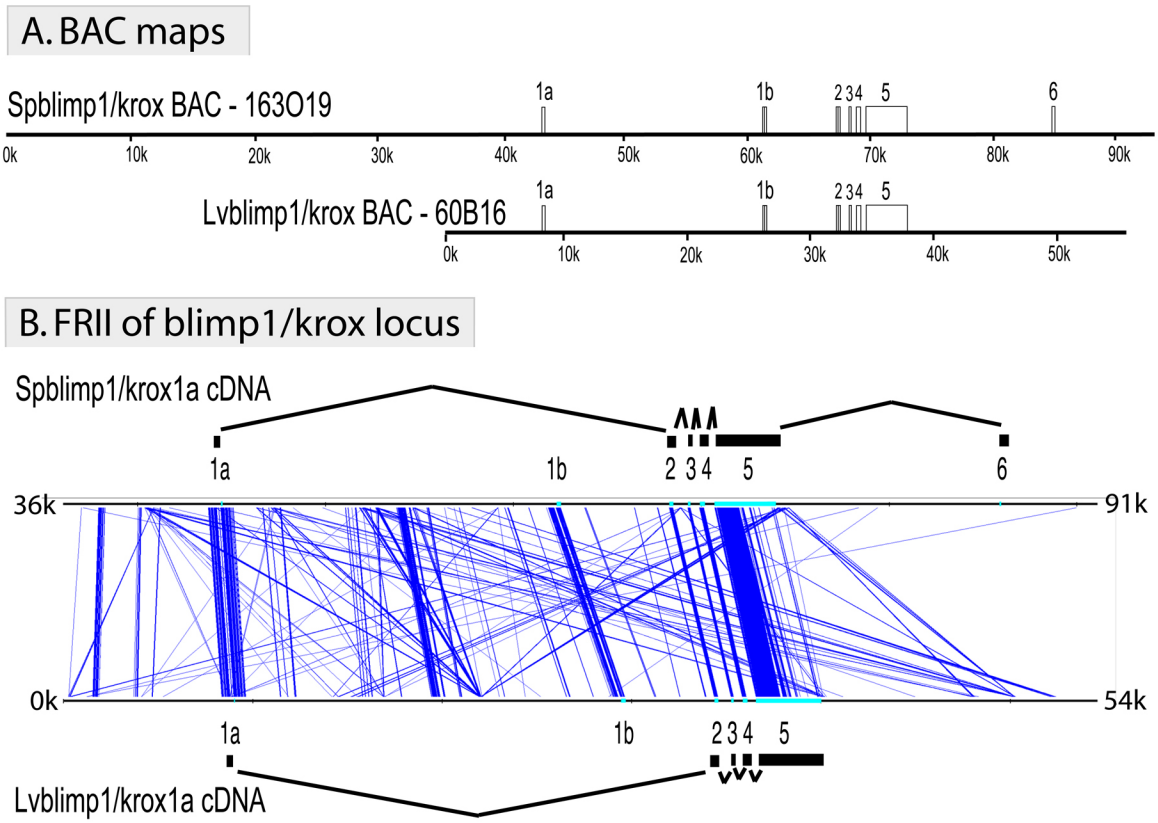
A. BAC maps
Spblimp1/krox BAC - 163O19
1a
1b
2 3 4
5
6
0k
10k
20k
30k
40k
50k
60k
70k
80k
90k
Lvblimp1/krox BAC - 60B16
1a
1b
2 3 4
5
0k
10k
20k
30k
40k
50k
B. FRII of blimp1/krox locus
Spblimp1/krox1a cDNA
1a
1b
2 3 4
5
6
36k
91k
0k
54k
1a
1b
2 3 4
5
Lvblimp1/krox1a cDNA

Figure 5: 190bp proximal positive ubiquitous module immediately 5' of exon 1b. (Top) Transgene expression table. (Bottom) Blastula stage embryos were imaged in glycerol under Nomarsky optics. WMISH of embryos injected with construct BP1b-GFP using antisense *gfp* RNA probe showing transgenic pattern of expression.

Figure 5

stained embryos	148	total stained cells/emb.	ectoderm stained cell/emb.	ectoderm stained embryos percentage	endoderm stained cell/emb.	endoderm stained embryos percentage	SMC stained cell/emb.	SMC stained embryos percentage	PMC cells/emb.	PMC stained embryos percentage
non-stained embryos	7	7.22	2.07	54.7%	2.56	61.5%	2.01	46.6%	0.58	19.6%
stained percentage	95.5%									

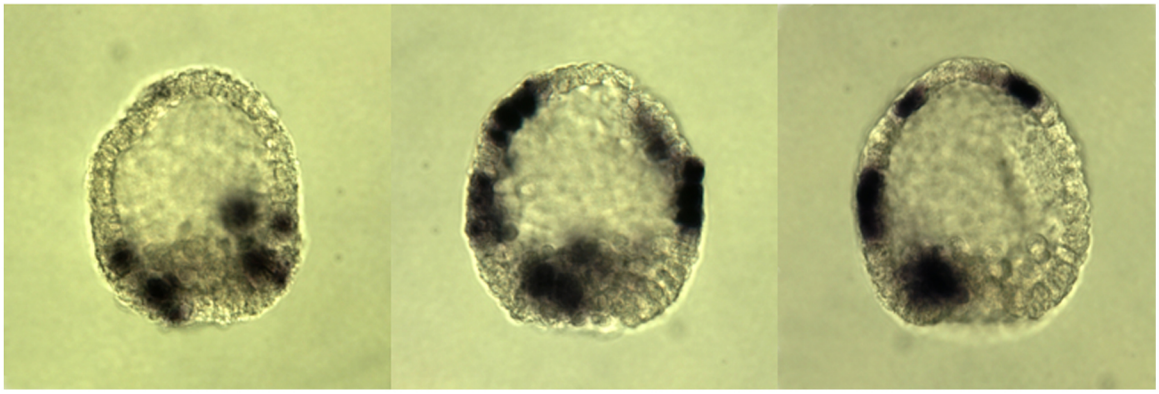

Figure 6: 3000bp region 5' of exon 1b containing multiple conserved sequence patches. (Top) Transgene expression table. (Bottom) Blastula stage embryos were imaged in sea water under Nomarsky optics and under fluorescent illumination. Overlayed images of embryos injected with construct 3k-GFP showing transgenic pattern of expression.

Figure 6

stained %	51.2%	total stained cells/ embryo	ectoderm stained cell/ embryo	ectoderm stained embryos %	vegetal plate endoderm stained cell/ embryo	vegetal plate endoderm stained embryos %	vegetal plate mesoderm stained cell/ embryo	vegetal plate mesoderm stained embryos %	PMC cells/ embryo	PMC stained embryos %
stained embryos	63	5.48	3.00	14.3%	5.08	84.1%	3.11	14.3%	2.33	14.3%
non-stained embryos	60