Available online at www.sciencedirect.com
ScienceDirect
Developmental Biology 300 (2006) 74–89
DEVELOPMENTAL BIOLOGY
www.elsevier.com/locate/ydbio

Identification and characterization of homeobox transcription factor genes in *Strongylocentrotus purpuratus*, and their expression in embryonic development

Meredith Howard-Ashby, Stefan C. Materna, C. Titus Brown, Lili Chen, R. Andrew Cameron, Eric H. Davidson *

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Received for publication 18 April 2006; revised 4 August 2006; accepted 16 August 2006
Available online 22 August 2006

Abstract

A set of 96 homeobox transcription factors was identified in the *Strongylocentrotus purpuratus* genome using permissive blast searches with a large collection of authentic homeodomain sequences from mouse, human and fly. A phylogenetic tree was constructed to compare the sea urchin homeobox gene family to those of vertebrates, with the result that with the only a few exceptions, orthologs of all vertebrate homeodomain genes were uncovered by our search. QPCR time course measurements revealed that 65% of these genes are expressed within the first 48 h of development (late gastrula). For genes displaying sufficiently high levels of transcript during the first 24 h of development (late blastula), whole mount in situ hybridization was carried out up to 48 h to determine spatial patterns of expression. The results demonstrate that homeodomain transcription factors participate in multiple and diverse developmental functions, in that they are used at a range of time points and in every territory of the developing embryo.

Keywords: Transcription factor; Sea urchin; Homeobox; Development

Introduction

Transcription factors are the key players in the gene networks directing development. These networks consist essentially of genes encoding sequence-specific regulatory proteins, the targets of which encode other transcription factors, thereby initiating cascades of overlapping directives which ultimately specify the many embryonic territories. To solve the architecture of developmental gene networks requires primary knowledge of which transcription factors are active in the embryo and when and where they are expressed. The availability of the *Strongylocentrotus purpuratus* genome sequence, which has just been obtained by the Human Genome Sequencing Center at Baylor College of Medicine (http://www.hgsc.bcm.tmc.edu/projects/seaurchin/; http://www.ncbi.nlm.nih.gov/genome/guide/seaurchin/), has made it possible to identify systematically all the transcription factors encoded in the genome. Thus, we sought to find and annotate all genes encoding sequence-specific DNA binding proteins predicted by the genome sequence. We then determined whether each is expressed in the early to mid-stage embryo and, for active genes, we established the temporal and spatial modes of expression.

Transcription factors fall into several large families defined by the structures of their DNA binding domains. The largest of these families in *S. purpuratus* is the Zn Finger family, an analysis of which is described in another paper of this series (Materna et al., 2006). The next largest is our present subject, the homeodomain family. Here we consider all subclasses of homeodomain regulatory genes except for the *hox* and *parahox* genes, which are the subject of a separate report (Arnone et al., 2006). Other classes of transcription factors are dealt with in additional papers (Ets family factors (Rizzo et al., 2006);

* Corresponding author. Fax: +1 626 793 3047.
E-mail address: Davidson@caltech.edu (E.H. Davidson).

0012-1606/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2006.08.039

Table 1

Gene name	Index	Glean ID	Subfamily
Sp-alx1	–	SPU_22817,SPU_25302	Paired
Sp-alx4	184	SPU_22816	Paired
Sp-arx	297	SPU_19338	Paired
Sp-arxl	298	SPU_17249	Paired
Sp-arxl2	389	SPU_21491	Paired
Sp-atbf1	78	SPU_17348	Other
Sp-awh	122	SPU_18954	Lim
Sp-barhl	259	SPU_14164	Hox
Sp-barx	260	SPU_01519,SPU_03920	Hox
Sp-brn124	–	SPU_16443	Other
Sp-brn3	18	SPU_25632	Other
Sp-cdx2	300	SPU_24715,SPU_19656	Hox
Sp-chx10	146	SPU_00485	Paired
Sp-cutl	331	SPU_03595	Other
Sp-dbx1	261	–	Nk
Sp-dlx	309	SPU_02815	Other
Sp-emx	150	SPU_02592	Hox
Sp-en	12	SPU_20975	Other
Sp-eve	257	SPU_12253	Hox
Sp-exd	68	SPU_05435,SPU_23739	Atypical
Sp-eyg	321	SPU_19129	Paired
Sp-eygl	393	SPU_16786	Paired
Sp-gbx	610	SPU_25492	Hox
Sp-gsc	–	SPU_15982	Paired
Sp-gsh1	317	SPU_13436	Hox
Sp-hb9	258	SPU_02816	Hox
Sp-hbn	324	SPU_23177	Paired
Sp-hex	263	SPU_27215	Nk
Sp-hlx	340	SPU_14802	Nk
Sp-hnf1	56	SPU_08196	Atypical
Sp-hnf6	–	SPU_16449	Other
Sp-hox1.tlx1	85	SPU_17352	Hox
Sp-hox11.13a	97	SPU_02632	Hox
Sp-hox11.13b	256	SPU_02631	Hox
Sp-hox11.13c	294	SPU_00388	Hox
Sp-hox2	293	SPU_12252,SPU_00386	Hox
Sp-hox3	253	SPU_27568	Hox
Sp-hox4.5	50.1	SPU_05169	Hox
Sp-hox6	254	SPU_05171	Hox
Sp-hox7	255	SPU_05170,SPU_02634	Hox
Sp-hox8	50.2	SPU_02630,SPU_21309	Hox
Sp-hox9.10	45	SPU_02633	Hox
Sp-irxA	200	SPU_10351	Atypical
Sp-irxB	299	SPU_11246	Atypical
Sp-isl	32	SPU_23730	Lim
Sp-lass6	388	SPU_00948	Other
Sp-lbx	115	SPU_14177	Nk
Sp-lhx2	268	SPU_04021	Lim
Sp-lhx3.4	105	SPU_01975	Lim
Sp-lim1	44	SPU_06991	Lim
Sp-lmx1	314	SPU_14157	Lim
Sp-mbx1	270	SPU_11297	Paired
Sp-meis	345	SPU_11202	Atypical
Sp-mox	109	SPU_23868,SPU_25486	Hox
Sp-msx	74	SPU_22049	Other
Sp-msxl	395	SPU_20565	Other
Sp-not	–	SPU_02129	Hox
Sp-nk1	265	SPU_12491	Nk
Sp-nk2.1	266	SPU_00757	Nk
Sp-nk2.2	75	SPU_00756	Nk
Sp-nk2.5	14	SPU_05472	Nk
Sp-nk3.2	267	SPU_13047	Nk
Sp-nk6.1	127	SPU_12699	Nk
Sp-nk7	327	SPU_22573	Nk
Sp-oct1.2	26	SPU_09262	Other

Table 1 (*continued*)

Gene name	Index	Glean ID	Subfamily
Sp-otp	272	SPU_19290	Paired
Sp-otx	–	SPU_10424	Paired
Sp-pax1.9	16	SPU_06683	Paired
Sp-pax258	47	SPU_14539	Paired
Sp-pax4l	394	SPU_17635,SPU_17636	Paired
Sp-pax6	296	SPU_06786	Paired
Sp-paxA	273	SPU_27334	Paired
Sp-paxB	274	SPU_18351	Paired
Sp-paxC	108	SPU_00276	Paired
Sp-phb1	392	SPU_08112	Paired
Sp-phb2	396	SPU_24093	Paired
Sp-pbx	–	SPU_23739	Paired
Sp-phox2	269	SPU_13464	Paired
Sp-pitx1	163	SPU_14461,SPU_24163	Paired
Sp-pitx2	275	SPU_04599	Paired
Sp-pitx3	84	SPU_06159,SPU_04598	Paired
Sp-pknox	330	SPU_12122	Atypical
Sp-pmar1	–	SPU_14721	Paired
Sp-pou6	618	SPU_10438	Other
Sp-prox1	343	SPU_15984	Atypical
Sp-prx	311	SPU_18951	Paired
Sp-rough	606	SPU_07242	Other
Sp-rx	151	SPU_16786	Paired
Sp-shox	310	SPU_19268	Paired
Sp-sip	81	SPU_22242	Other
Sp-six1.2	15	SPU_17379	Atypical
Sp-six3	2	SPU_18908	Atypical
Sp-six4	21	SPU_17380	Atypical
Sp-tgif	43	SPU_18126	Atypical
Sp-unc4.1	334	SPU_01739,SPU_13704	Paired
Sp-xlox	40	SPU_20637,SPU_26099	Hox

Forkhead family factors (Tu et al., 2006); and all other families (Howard-Ashby et al., 2006)).

Materials and methods

Identification of transcription factor sequences

Most of the transcription factors considered here were initially identified from the unassembled sea urchin genome traces and the November 2004 Baylor University draft genome assembly using a reference database of known transcription factors (excluding zinc fingers). This "rake" was assembled from two sources: nr human, mouse and fly sequences tagged as "transcription factor" and the GO seqdblite databases GO:0003700, GO:0000130, GO:0030528, GO:0003705, GO:0003702 and GO:0003677. Entries were removed if they contained the descriptors "general transcription factor II," "TFII," "TFIII," "protease," "histone," "reverse transcriptase," "nucleosome," "RNA polymerase," DNA replications," "chromatin," "helicase," "DNase" or "exonuclease." Any nonhomeodomain/non-GATA zinc finger proteins were also removed from the rake database. The final rake contained approximately 4900 protein sequences.

Tblastn (Altschul et al., 1990) of the protein sequences in the rake against the individual traces, as well as the translated Baylor draft assembly (cutoff=e^{-10}) was used to coarsely identify all traces or contigs potentially encoding transcription factors. Blastx of this subset of sequences vs. the rake protein database (cutoff=e^{-12}) was then used to highlight the locations of exons encoding transcription factor-specific conserved domains (e.g. bHLH, homeodomain, sox). Finally, the isolated conserved domains were blasted (tblastn) against NCBI's nr database to establish the closest known homologues. To avoid redundancy, efforts were made to group multiple exons from the same protein. Complementary exons from the same large contig as well as complementary exons from smaller contigs with the same closest homologues were assigned one unique number/gene name. PCR of sea urchin cDNA was used to confirm that different