Development 128, 2893-2904 (2001)
Printed in Great Britain © The Company of Biologists Limited 2001
DEV5914

Gene expression profiles in *Ciona intestinalis* tailbud embryos

Yutaka Satou[1], Naohito Takatori[1], Lixy Yamada[1], Yasuaki Mochizuki[1], Makoto Hamaguchi[1], Hisayoshi Ishikawa[1], Shota Chiba[1], Kaoru Imai[1], Shungo Kano[1], Seiko D. Murakami[1], Akie Nakayama[1], Atsuo Nishino[1], Yasunori Sasakura[1], Gohki Satoh[1], Taishin Shimotori[1], Tadasu Shin-i[2], Eiichi Shoguchi[1], Miho M. Suzuki[1], Norio Takada[1], Nanami Utsumi[3], Natsue Yoshida[3], Hidetoshi Saiga[3], Yuji Kohara[2] and Nori Satoh[1,*]

[1]Department of Zoology, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan
[2]Center for Genetic Resource Informations, National Institute of Genetics, Mishima 411-8540, Japan
[3]Department of Biological Sciences, Graduate School of Science, Tokyo Metropolitan University, Hachiohji, Tokyo 192-0397, Japan
*Author for correspondence (e-mail: satoh@ascidian.zool.kyoto-u.ac.jp)

Accepted 15 May 2001

SUMMARY

A set of 3423 expressed sequence tags derived from the *Ciona intestinalis* tailbud embryos was categorized into 1213 independent clusters. When compared with DNA Data Bank of Japan database, 502 clusters of them showed significant matches to reported proteins with distinct function, whereas 184 lacked sufficient information to be categorized (including reported proteins with undefined function) and 527 had no significant similarities to known proteins. Sequence similarity analyses of the 502 clusters in relation to the biosynthetic function, as well as the structure of the message population at this stage, demonstrated that 390 of them were associated with functions that many kinds of cells use, 85 with cell-cell communication and 27 with transcription factors and other gene regulatory proteins. All of the 1213 clusters were subjected to whole-mount in situ hybridization to analyze the gene expression profiles at this stage. A total of 387 clusters showed expression specific to a certain tissue or organ; 149 showed epidermis-specific expression; 34 were specific to the nervous system; 29 to endoderm; 112 to mesenchyme; 32 to notochord; and 31 to muscle. Many genes were also specifically expressed in multiple tissues. The study also highlighted characteristic gene expression profiles dependent on the tissues. In addition, several genes showed intriguing expression patterns that have not been reported previously; for example, four genes were expressed specifically in the nerve cord cells and one gene was expressed only in the posterior part of muscle cells.

This study provides molecular markers for each of the tissues and/or organs that constitutes the *Ciona* tailbud embryo. The sequence information will also be used for further genome scientific approach to explore molecular mechanisms involved in the formation of one of the most primitive chordate body plans.

Key words: *Ciona intestinalis*, Tailbud embryos, cDNA project, ESTs, Gene expression profiles

INTRODUCTION

Ascidians belong to the subphylum Urochordata, which is one of the three chordate groups. Adult ascidians are sessile and specialized for filter feeders, but in most species their fertilized eggs develop rather quickly into tadpole-type larvae. The ascidian tadpole consists of about 2600 cells that form several distinct types of tissues (Satoh, 1994). The tadpole is organized into a trunk and tail. The trunk contains a dorsal central nervous system (CNS) with two sensory organs (otolith and ocellus), endoderm, mesenchyme, trunk lateral cells (TLCs) and trunk ventral cells (TVCs). The tail contains a notochord flanked dorsally by the nerve cord (non-neuronal ependymal cells), ventrally by endodermal strand, and bilaterally by three rows of muscle cells. The entire surface of the larva is covered by an epidermis. This configuration of the ascidian tadpole is thought to represent one of the most simplified and primitive chordate body plans (reviewed by Satoh and Jeffery, 1995; Di Gregorio and Levine, 1998; Satou and Satoh, 1999; Wada and Satoh, 2001; Nishino and Satoh, 2001).

The ascidian embryogenesis is well documented. The cleavage pattern is invariant, and cleavage is bilaterally symmetrical. Gastrulation is initiated around the 118-cell stage, and it involves epibolic movements of ectodermal cells and migration of endodermal and mesodermal cells inside the embryo. Neurulation is accomplished by folding of the presumptive neural plate, as in vertebrate embryos. Then, the tailbud embryo is formed, which eventually develops into a tadpole larva. In addition, the lineage of embryonic cells is characterized by detailed descriptions of the epidermis, CNS, endoderm, mesenchyme, TLCs, muscle and notochord (Conklin, 1905; Ortolani, 1955; Nishida and Satoh, 1983; Nishida, 1987; Nicol and Meinertzhagen, 1988). These research findings show that the ascidian embryo is an appropriate experimental system with which to explore cellular and molecular mechanisms that underlie the embryonic cell

specification and pattern formation of the embryo (reviewed by Satoh, 1994; Satoh, 1999; Nishida, 1997; Jeffery, 2001). Regulatory mechanisms of specific developmental gene expressions as well as cell-cell interactions are able to be investigated at single cell level (e.g. Yasuo and Satoh, 1993; Nakatani and Nishida, 1994; Corbo et al., 1997).

The genome size of *Ciona intestinalis* is estimated to be about 160 Mb, and the number of genes approximately 15,500 (Simmen et al., 1998). This genome size and gene number are comparable with those of *Drosophila*. *C. intestinalis* is a cosmopolitan species used by researchers worldwide, and they spawn all year round. Their eggs are self-fertile and the generation time is about 2-3 months. These conditions have allowed us to screen mutants that affect developmental processes (Nakatani et al., 1999 in *Ciona savignyi*; Sordino et al., 2000 in *Ciona intestinalis*).

Given these research results, we decided to attempt cDNA projects of *C. intestinalis*, in collaboration with *Ciona* genome project consortium members. We have focused on the gene expression profiles of embryos at the tailbud stage because it is expected that, at this stage, most of tissues and organs that constitute the future larva begin to be differentiated; therefore, the genes responsible for the formation of these tissues are expressed. We had several specific aims. One was to obtain molecular markers for each of the tissues and/or organs. Although molecular markers for differentiation of embryonic cells have been well characterized in another species (*Halocynthia roretzi*), to date, in *Ciona* embryos, molecular markers have been reported for only epidermis, CNS, endoderm, muscle and notochord (Chiba et al., 1998; Takamura, 1998; Takahashi et al., 1999; Hotta et al., 2000; Imai et al., 2000). If our study could characterize such markers, they might be used in future studies to understand molecular developmental mechanisms of *Ciona* embryos. Second, our study should provide information about the overall gene expression profiles in *Ciona* tailbud embryos. Lee et al. have reported EST analysis of gene expression in early cleavage-stage sea urchin embryos, in which, for example, 29 of 153 proteins were identified as being associated with cell-cell communication (Lee et al., 1999). Similar data should be obtained in our study. In addition, our study has investigated the gene expression profiles of more than 1000 independent genes by whole-mount in situ hybridization. We discuss how many and what kind of genes are expressed in each tissue. Third, determination of the entire genome sequence is now under way. Our data could be used for annotating these genes to genome sequencing data in the future.

MATERIALS AND METHODS

Biological materials

Ciona intestinalis were cultivated at the Maizuru Fisheries Research Station of the Kyoto University, Maizuru, and the Education and Research Center of Marine Bio-Resources of Tohoku University, Miyagi, Japan. Adults were maintained under constant light to induce oocyte maturation. Eggs and sperm were obtained surgically from the gonoduct. After insemination, eggs were dechorionated by immersing them in seawater that contain 1.3% sodium thioglycolate (Wako Pure Chemical Industries, Osaka, Japan) and 0.065% actinase E (Kaken Pharmaceutical Company, Tokyo, Japan). After washing, they were maintained in agar-coated dishes with Millipore-filtered seawater (MFSW) containing 50 μg/ml streptomycin sulfate at room temperature (18-20°C). They developed to tailbud embryos by 10-13 hours of development. These were collected for RNA isolation and whole-mount in situ hybridization.

cDNA library

Total RNA was isolated from *C. intestinalis* tailbud embryos by the acid guanidinium thiocyanate-phenol-chloroform method (Chomczynski and Sacchi, 1987). Poly $(A)^+$ RNA was purified twice using Oligotex beads (Roche Japan, Tokyo). Poly $(A)^+$ RNA was converted to double-stranded cDNA, which contains *Eco*RI site at the 5′-end and *Xho*I site at the 3′-end, using a cDNA synthesis kit (Stratagene). The cDNAs were ligated to pBSII-SK(–) digested with *Eco*RI and *Xho*I. The cDNAs were electroporated into XL-1 Blue MRF′ bacteria (Stratagene). The library was arrayed in 384-well plates in a Genetix Q-Pix robot. The first stage of the arrayed library preparation was done in the Division of Biology, California Institute of Technology with kind help of Dr Eric H. Davidson.

EST sequencing

Clones were picked up from the 384-well plates and cDNA inserts were amplified using PCR. Successful amplifications were confirmed using agarose electrophoresis. After purification of the PCR products, their sequences were determined by the conventional procedures using the big-dye terminators on ABI3700 autosequencers at the Academia DNA Sequencing Center, National Institute of Genetics Japan. The primer for 3′-sequencing was the anchored oligo dT primers (5′-(T)17-V-3′) and that for 5′-sequencing was BS740 (5′-CCGC-TCTAGAACTAGTG-3′).

Clustering and homology search using ESTs

Using 3′-most sequence tags, clones were grouped into clusters, each of which contained cDNA clones encoding the same gene. The DDBJ DNA database and protein database (DAD) were searched with 5′-most sequence tags using BLAST algorithm (BLASTN and BLASTX). We categorized clusters into several groups according to their functions that were predicted from BLASTX results (see Results).

Whole-mount in situ hybridization

For convenience, we applied a method of whole-mount in situ hybridization using DNA probes to ascidian embryos. DNA probes were synthesized basically as described by Tabara et al. (Tabara et al., 1996) with several modifications. First, cDNA inserts were PCR-amplified with SK and T7 primers. Each 10 μl reaction mixture contained 1.4 ng template DNA, 0.1 μM primers, 0.2 mM 4dNTP, 1.5 mM $MgCl_2$, 1× buffer, and 0.25 U of Taq DNA polymerase (TOYOBO). Reactions proceeded through 35 cycles (30 seconds at 94°C, 30 seconds at 55°C, 3 minutes at 72°C). Digoxigenin-labeled DNA probes were also synthesized by PCR reactions. Each 10 μl reaction mixture contained 1 μl first PCR product, 0.0525 mM DIG-dUTP (Boehringer Mannheim), 0.0975 mM dTTP, 0.15 mM dATP, 0.15 mM dCTP, 0.15 mM dGTP, 1.5 mM $MgCl_2$, 0.53 μM (dT)17dG, 0.53 μM (dT)17dC, 2.13 μM (dT)17dA, 1× buffer and 0.5 U of Taq DNA polymerase (TOYOBO). Reactions proceeded through 30 cycles (30 seconds at 95°C, 1 minute at 45°C, 1 minute at 72°C). After synthesis, free nucleotides and primers that were not incorporated were removed by ethanol precipitation.

Embryos for whole-mount in situ hybridization were fixed in 4% paraformaldehyde in 0.1 M MOPS (pH 7.5), 0.5 M NaCl at 4°C for 16 hours, before storage in 80% ethanol at −30°C. After thorough washes with PBST (phosphate-buffered saline containing 0.1% Tween 20), the fixed specimens were partially digested with 2 μg/ml proteinase K in PBST for 20 minutes at 37°C and boiled in a water bath for 5 minutes. After being washed with PBST, the specimens were post-fixed with 4% paraformaldehyde in PBST for 1 hour at room temperature, then washed again with PBST. After

prehybridization at 50°C for 1 hour, the specimens were hybridized with digoxigenin-labeled probes at 50°C for 16 hours. The hybridization buffer contained 50% formamide, 5× SSC, 5× Denhardt's solution, 100 μg/ml salmon sperm DNA, 0.1% Tween 20, and DIG-labeled DNA probe.

After hybridization, the specimens were washed twice for 20 minutes in the wash solution (50% formamide, 2× SSC, 0.1% Tween 20) at 50°C, and then twice in the wash solution/PBST (1:1). They were washed in PBST at 50°C four additional times. After being cooled to room temperature, the specimens were blocked with 0.5% blocking reagent (Boehringer Mannheim) in PBST for 30 minutes before 1 hour incubation with 1:2000 alkaline-phosphatase-conjugated anti-digoxigenin antibody (Boehringer Mannheim). The specimens were washed in PBST for 10 minutes four times and then in alkaline phosphatase buffer (100 mM NaCl, 50 mM $MgCl_2$, 100 mM Tris-HCl, pH 9.5) twice. For signal detection, the embryos were incubated with NBT/BCIP/alkaline phosphatase buffer following the supplier's instructions (Boehringer Mannheim). The reaction was stopped in PBST. The embryos were dehydrated in a graded series of ethanol, then cleared in a 1:2 mixture of benzyl alcohol:benzyl benzoate (BABB).

RESULTS AND DISCUSSION

Overall distribution of sequences

In the present study, we determined sequences of both the 5′-most and 3′-most ends of a total of 3423 cDNA clones derived from *Ciona intestinalis* tailbud embryos. The average length of the sequences was 400~500 nucleotides. The sequences of the 3′-most end were used to examine the overlap of clones, and this analysis categorized the 3423 clones into 1213 independent clusters (Table 1). Our consortium is now proceeding with further EST analysis of *C. intestinalis* tailbud embryos, and sequences of more than 10,000 clones are accessible with DDBJ (GenBank/EMBL) database. To obtain the gene expression profiles we first selected *C. intestinalis* tailbud embryos, as we wish to re-examine more carefully the sequence data when we eventually obtain nearly all the saturated ESTs. The present analysis is therefore partial and will be described rather briefly.

Sequences of the 5′-most end, as well as the 3′-most end, of 1213 clusters that were subjected to BLASTX analysis (Table 1). BLASTX analysis showed that, of the 1213 clusters, 502 were strong matches (P<E-15) to previously identified proteins with distinct functions. The frequency of matches was therefore 502/1213, or about 0.41. As summarized in Table 1, these 502 proteins were categorized into three major classes, basically according to previous classification (Lee et al., 1999); Class A, which contains 390 proteins associated with functions that many kinds of cell use; Class B, which involves 85 proteins associated with cell-cell communication; and Class C, which includes 27 proteins that function as transcription factors and other gene regulatory proteins. Furthermore, the 390 proteins of Class A were categorized into nine subclasses (AI~AIX), the 85 proteins of Class B into three subclasses (BI~BIII) and the 27 proteins of Class C into three subclasses (CI~CIII) (Table 1). The number of genes belonging to each subclass is also shown in Table 1. The great majority of proteins were identified only once in our analysis, although a few highly abundant proteins were identified many times (data not shown). The average number of hits/protein was 3.01 (1511/502; Table 1).

Table 1. The number of different genes expressed by class

Class			Number of clusters	Number of clones
(A) Functions that many kinds of cells use				
AI	Transportation and binding proteins for ions and other small molecules		17	47
AII	RNA processing, polymerizing, splicing and binding proteins, and enzymes		44	86
AIII	Cell replication, histones, cyclins and allied kinases, DNA polymerases, topoisomerases, DNA modification		38	97
AIV	Cytoskeleton and membrane proteins		48	296
AV	Protein synthesis co-factors, tRNA synthetases, ribosomal proteins		59	305
AVI	Intermediary synthesis and catabolism enzymes		117	301
AVII	Stress response, detoxification and cell defense proteins		11	37
AVIII	Protein degradation and processing, proteases		29	59
AIX	Transportation and binding proteins for proteins and other macromolecules		27	68
		Total	390	1296
(B) Cell-cell communication				
BI	Signaling receptors, including cytokine and hormone receptors, and signaling ligands		11	15
BII	Intracellular signal transduction pathway molecules including kinases and signal intermediates		61	99
BIII	Extracellular matrix proteins and cell adhesion		13	52
		Total	85	166
(C) Transcription factors and other gene regulatory proteins				
CI	Sequence-specific DNA-binding proteins		21	38
CII	Non-DNA binding proteins that perform positive or negative roles		4	5
CIII	Chromatin proteins other than AIII with regulatory function		2	6
		Total	27	49
(D) Miscellaneous				
DI	Not enough information to classify		184	1121
DII	Not significant similarities to known proteins		527	791
		Total	711	1912
		Total	1213	3423

Table 2. Expressed sequence tag similarities, gene description and probability of occurrence by chance

Class	ClusterID	Accession number	Database entry name	Organism	Probability
AI	00121	AF221690	Voltage-dependent anion channel	*Squalus acanthias*	3E-59
	00385	CAA37609	Signal sequence receptor β subunit	*Canis familiaris*	1E-42
	00794	AAF35832	Zinc transporter hZIP2	*Homo sapiens*	7E-25
	00885	CAA73906	Calmodulin	*Ciona intestinalis*	1E-80
	01239	CAB40132	Calmodulin 2	*Branchiostoma floridae*	1E-16
	01624	AAD26138	Cl^- channel protein p64H1	*Bos taurus*	3E-27
	01661	BAA01643	H^+-transporting ATPase	*Rattus norvegicus*	2E-25
	02049	AAA40991	Ca^{2+} transporting ATPase	*Rattus norvegicus*	1E-88
AII	00006	AAF66160	RRN3	*Homo sapiens*	5E-26
	00237	CAA74583	Heterogeneous nuclear ribonucleoprotein H	*Mus musculus*	1E-17
	00554	AAD38877	p68 RNA helicase	*Molgula oculata*	E-109
	00714	AAA49949	Ribonucleoprotein	*Xenopus laevis*	4E-45
	00795	BAA87912	TBP-binding protein ABT1	*Mus musculus*	9E-34
	00843	AAF19376	RNA polymerase	*Xenopus laevis*	2E-42
	00865	BAB03404	PEM-3	*Ciona savignyi*	7E-76
	00888	AAF72188	snRNP-associated protein; SmB	*Danio rerio*	8E-32
	00908	AAB58717	TFIIA small subunit	*Rattus norvegicus*	4E-36
	01310	AAA16347	Splicing factor	*Homo sapiens*	6E-33
	01538	CAA45354	snRNP C	*Xenopus laevis*	5E-25
	01808	AAF37578	Serine-arginine-rich splicing regulatory protein SRRP86	*Rattus norvegicus*	6E-31
	01994	AAC52056	RNA polymerase II subunit hsRPB4	*Homo sapiens*	6E-53
	01997	AAC39540	Heterogeneous nuclear ribonucleoprotein R	*Homo sapiens*	4E-52
AIII	00029	CAA61665	Cyclin D2	*Xenopus laevis*	3E-55
	00342	AAD11940	Mitotic checkpoint protein kinase BUB1B	*Mus musculus*	5E-35
	00383	CAA32968	Histone H2A.X	*Homo sapiens*	1E-40
	00973	AAF43778	Cyclin-dependent protein kinase H	*Homo sapiens*	1E-59
	01456	CAA43807	Cell division kinase. CDC2 homolog	*Homo sapiens*	2E-84
	01505	AAB60379	DNA topoisomerase I	*Homo sapiens*	1E-33
	01629	AAC52080	Cdc7	*Homo sapiens*	6E-30
	01816	AAF05755	Anaphase-promoting complex subunit 8	*Homo sapiens*	1E-67
	01843	AAB18946	Cyclin C	*Gallus gallus*	7E-83
	02009	AAB09784	Replication factor C, 36 kDa subunit	*Homo sapiens*	4E-81
	02045	BAA25400	CsCDC42	*Ciona savignyi*	E-106
AIV	00025	BAA23596	CsMA-1	*Ciona savignyi*	E-127
	00031	AAA59890	Myosin regulatory light chain	*Homo sapiens*	2E-52
	00046	AAC28357	Cytoskeletal actin 1	*Molgula occulta*	E-110
	00086	BAA22381	β-tubulin	*Halocynthia roretzi*	E-129
	00173	AAD09271	Troponin I	*Ciona intestinalis*	5E-92
	00189	AAA86910	Fast-twitch myosin light chain 1	*Bos taurus*	1E-37
	00259	CAC03999	Intermediate filament protein C	*Styela clava*	E-58
	00327	BAA08111	Embryonic muscle myosin heavy chain	*Halocynthia roretzi*	3E-49
	00372	AAA48656	Capping protein α2 isoform	*Gallus gallus*	6E-16
	00608	CAA45469	Tropomyosin	*Ciona intestinalis*	3E-48
	00982	CAA62350	αII spectrin	*Rattus norvegicus*	E-101
	00988	AAA60580	β-spectrin	*Homo sapiens*	2E-40
	02046	AAC00533	Capping protein α subunit isoform 1	*Homo sapiens*	2E-50
AV	00010	AAB26418	Ribosomal protein 49	*Drosophila persimilis*	4E-40
	00045	CAA40592	Ribosomal protein S1a	*Xenopus laevis*	5E-95
	00075	CAA63732	Ribosomal protein L10a	*Rattus norvegicus*	8E-68
	00080	AAA50025	Translational elongation factor 1α	*Danio rerio*	6E-88
	00150	S69004	Translation initiation factor eIF-4E, long splice form	*Xenopus laevis*	2E-51
	00227	AAF64459	Ribosomal protein L18	*Tilapia mossambica*	3E-71
	00514	P17008	40S ribosomal protein S16	*Rattus norvegicus*	5E-55
	00693	A47151	Ethionine adenosyltransferase (EC 2.5.1.6)	*Mus musculus*	2E-44
	01228	CAA73167	Translation initiation factor eIF4A I	*Xenopus laevis*	3E-61
	01237	A53221	Acidic ribosomal protein P1	*Polyorchis penicillatus*	8E-22
	01820	AAC38014	Ribosomal protein S6	*Xenopus laevis*	6E-56
	01852	BAA06623	eIF-4E protein	*Xenopus laevis*	3E-42
AVI	00037	AAB50594	Cytochrome oxidase 1	*Psolus chitonoides*	4E-16
	00058	AAA83428	ALDH7	*Homo sapiens*	6E-43
	00066	P91924	ADP-ribosylation factor	*Dugesia japonica*	1E-72
	00114	BAA05020	HrEpiB	*Halocynthia roretzi*	9E-77
	00138	BAA32086	Natural killer cell enhancing factor	*Cyprinus carpio*	1E-85
	00140	AAA70333	Adenylosuccinate synthetase	*Schizosaccharomyces pombe*	2E-44
	00155	AAA30359	*S*-adenosylmethionine decarboxylase	*Bos taurus*	2E-52
	00156	AAC39661	Pyruvate dehydrogenase complex protein X subunit precursor	*Homo sapiens*	4E-27
	00190	CAB52415	Carnitine palmitoyltransferase I	*Drosophila melanogaster*	62-34
	00203	AAB02271	Vacuolar ATPase subunit A	*Drosophila melanogaster*	2E-52
	00249	AAC64398	ATP-specific succinyl-CoA synthetase beta subunit	*Mus musculus*	6E-53
	00283	BAA88254	Cytochrome oxidase subunit III	*Halocynthia roretzi*	4E-19
	00302	AAC72372	Succinate dehydrogenase Ip subunit	*Gallus gallus*	3E-61

Table 2. Continued

Class	ClusterID	Accession number	Database entry name	Organism	Probability
AVII	00496	CAA06233	Heat shock cognate 70	*Gallus gallus*	1E-87
	00668	AAC48718	Heat shock protein 90A	*Sus scrofa*	5E-90
	01282	P81926	Superoxide dismutase [CU-ZN]	*Halocynthia roretzi*	1E-46
	01295	AAB21614	Prohibitin	*Homo sapiens*	8E-58
	01307	AAA18335	Heat shock protein HSP27	*Mus musculus*	2E-44
	01586	AAD17992	PRx III	*Rattus norvegicus*	2E-49
AVIII	00197	CAA80851	Ubiquitin	*Phanerochaete chrysosporium*	E-114
	00221	AAC09297	N-myristoyltransferase 2 protein	*Mus musculus*	4E-45
	00349	AAC50477	Amyloid precursor protein-binding protein 1	*Homo sapiens*	8E-62
	00509	AAC59636	Carboxypeptidase H	*Lophius americanus*	8E-44
	00640	AAB06237	Cyclin-specific ubiquitin carrier protein E2-C	*Spisula solidissima*	2E-58
	00680	CAA96580	Herpesvirus-associated ubiquitin-specific protease (HAUSP)	*Homo sapiens*	4E-33
	00689	BAA11338	Proteasome subunit p42 protein	*Homo sapiens*	2E-91
	01687	AAC26141	Ubiquitin-conjugating enzyme 12	*Homo sapiens*	7E-39
	01988	BAA89276	α4 subunit of 20S proteasome	*Carassius auratus*	4E-73
AIX	00022	BAA05019	HRSec61 protein	*Halocynthia roretzi*	8E-86
	00771	CAA65774	Sec23 protein	*Homo sapiens*	3E-69
	00798	CAA34386	SRP 54K subunit (AA 1-504)	*Mus musculus*	6E-82
	00854	AAC52154	Transitional endoplasmic reticulum ATPase	*Rattus norvegicus*	6E-57
	00887	AAA37707	Facilitated glucose transporter	*Mus musculus*	2E-36
	00971	AAC53372	Importin α Q2	*Mus musculus*	8E-46
	01281	AAC79495	EH domain binding protein epsin 2	*Rattus norvegicus*	4E-81
	01366	AAD40007	Small zinc finger-like protein	*Mus musculus*	3E-22
	01494	AAD13577	VAMP-associated protein B	*Homo sapiens*	1E-51
BI	00334	AAC42665	Nuclear receptor I	*Ciona intestinalis*	3E-94
	00700	BAA76876	Nicotinic acetylcholine recepter α-subunit	*Canis familiaris*	2E-57
	00817	AAC77361	Frizzled-8	*Xenopus laevis*	1E-25
	00914	AAC38017	x-Delta-1	*Xenopus laevis*	5E-26
	00991	AAA42186	Serotonin transporter	*Rattus norvegicus*	6E-28
	01003	AAC50948	snRNA activating protein complex 50kD subunit	*Homo sapiens*	3E-30
BII	00060	AAB81617	Receptor for activated protein kinase C	*Danio rerio*	1E-66
	00321	BAA92185	β-catenin protein	*Ciona intestinalis*	E-102
	00329	BAA96292	GTP-binding protein tc10	*Rattus norvegicus*	4E-62
	00340	BAA92186	Glycogen synthase kinase protein	*Ciona intestinalis*	6E-74
	00830	CAB81555	B-Raf protein	*Mus musculus*	1E-20
	00863	BAA05878	TGF-β type I receptor	*Mus musculus*	6E-53
	01342	AAC14343	14-3-3 protein beta	*Mus musculus*	4E-24
	02050	AAA48773	Protein-tyrosine kinase	*Gallus gallus*	2E-29
BIII	00177	AAA50293	α-2 type IV collagen	*Mus musculus*	3E-38
	00311	CAA45920	Restrictin-precursor	*Gallus gallus*	2E-40
	01986	AAA59486	Laminin B1	*Homo sapiens*	3E-81
CI	00074	AAD11962	NK homeobox protein	*Homo sapiens*	7E-35
	00220	AAC60126	xGCNF	*Xenopus laevis*	7E-19
	00250	BAA08722	As-MEF2	*Halocynthia roretzi*	8E-66
	00107	AAC50893	FUSE binding protein 3	*Homo sapiens*	5E-17
	00304	AAA70033	TAR DNA-binding protein-43	*Homo sapiens*	8E-49
	00306	BAA89664	Cas-associated zinc finger protein	*Rattus norvegicus*	5E-21
	00319	AAF60348	Cdx	*Herdmania curvata*	8E-17
	00350	AAD24209	CCCH zinc finger protein C3H-3	*Xenopus laevis*	9E-34
	00511	AAB91435	CAGF9	*Homo sapiens*	6E-30
	00517	AAD00562	Transcriptional regulator Sox-11B	*Danio rerio*	1E-20
	00681	BAA74520	Similar to *Drosophila ash2* gene	*Homo sapiens*	9E-62
	00801	BAA89208	Bromodomain PHD finger transcription factor	*Homo sapiens*	4E-20
	00869	AAC53022	skm-BOP2	*Mus musculus*	4E-37
	00905	BAA96136	PBX1B	*Gallus gallus*	5E-39
	00968	AAA19853	COUP-TFI	*Mus musculus*	4E-33
	00989	A35913	Regulatory factor X	*Homo sapiens*	7E-32
	01414	BAA01482	Zinc-finger protein	*Mus musculus*	9E-45
	01659	CAA69928	Nuclear orphan receptor ROR-β	*Gallus gallus*	3E-49
	01826	CAA62353	Polybromo 1 protein	*Gallus gallus*	3E-18
	01880	CAB64386	SoxNeuro	*Drosophila melanogaster*	3E-35
	02043	AAA20993	NF45 protein	*Homo sapiens*	9E-77
CII	00077	AAB18236	TAT interactive protein	*Homo sapiens*	7E-22
	00684	AAA61194	Transducin-like enhancer protein	*Homo sapiens*	9E-38
	00711	AAC15912	Nuclear protein Skip	*Homo sapiens*	1E-42
	00841	AAF35860	PPAR interacting protein PRIP	*Mus musculus*	8E-17
CIII	00488	Q08945	Structure-specific recognition protein 1 (SSRP1)	*Homo sapiens*	3E-61
	00531	JC6179	Dorsal switch protein 1	*Drosophila melanogaster*	3E-47

Besides the 502 clusters with strong matches, 184 clusters are classified into Class DI in which sequences were matches to ESTs (mostly from *Caenorhabditis elegans*, *Mus musculus* and *Homo sapiens*) or reported proteins, functions of which have not enough information to classify (for example, *posterior end mark* gene of *Ciona*; Table 1). The remaining 527 clusters were categorized into Class DII with no significant sequence similarities to other known proteins (Table 1). A simple estimation of these data suggests that nearly 43% (527/1213) of genes that are expressed in *Ciona* tailbud embryos should be characterized in future studies.

Sequence analysis of selected cDNA clones

Table 2 shows some examples of sequence analysis to identify cDNAs that encode especially strong candidates for proteins with a defined function. For example, as members of the subclass BI of signaling receptors including cytokine and hormone receptors and signal ligands, cDNAs for nuclear receptor I (cluster ID, 00334), frizzled-8 (00817) and x-Delta-1 (00914) were identified. The subclass BII for intracellular signal transduction pathway molecules including kinases and signal intermediates includes cDNAs for the receptor for activated protein kinase C (00060), β-catenin (00321), TGF-β type I receptor (00863), and protein-tyrosine kinase (02050). In addition, as members of the subclass CI of sequence-specific DNA-binding proteins, cDNAs for NK homeobox protein (00074), Cdx (00319), Sox-11B (00517) and SoxNeuro (01880) were identified. The cDNAs for transduction-like enhancer protein (00684) and nuclear protein Skip (00711) were identified as members of the subclass CII for non-DNA binding proteins that perform positive and negative roles. (Further information can be found at http://ghost.zool.kyoto-u.ac.jp)

Spatial expression profiles of genes

Overview

All of the 1213 clusters were subjected to analysis by whole-mount in situ hybridization to explore the overall gene expression profiles in *Ciona* tailbud embryos. As summarized in Table 3, 148 clusters showed no significant hybridization signals while 329 clusters showed ambiguous signals. These categories that did not show distinct gene expression patterns are presumably due to the presence of small amount of mRNAs in the embryo, because most of them constitute a category of one clone/cluster.

However, 387 of the 1213 clusters showed spatial expression patterns that were specific to a single tissue or organ (Table 3; see Fig. 2). Within these clusters, 149 genes were expressed specifically in epidermal cells, 34 genes were specific to the nervous system, including neuronal cells in the palps (an adhesive organ located at the anterior most part of the embryo) and epithelial sensory cells. Endoderm-specific expression was observed for 29 genes, the number of mesenchyme-specific genes was 112, notochord-specific expression was detected for 32 genes, and 31 genes were expressed specifically in muscle cells. Several examples of specific gene expression for each tissue will be described in the following section.

Table 4 and Fig. 1 show the relationships between the genes with single tissue-specific expression and classes obtained

Table 3. Overall view of specific expression patterns of genes in *Ciona intestinalis* tailbud embryos

	Genes specifically expressed in a single tissue		Genes specifically expressed in multiple tissues	
Tissues	Number of clusters	Number of clones	Number of clusters	Number of clones
Epidermis	149	688	187	357
Nervous system	34	73	239	1126
Neuron	1*	1*	41*	82*
Brain	6*	11*	196*	1000*
Papilla	8*	9*	153*	274*
Nerve cord	4*	14*	165*	368*
Endoderm	29	49	195	319
Endodermal strand	0	0	88	188
Mesenchyme	112	230	196	337
Trunk lateral cells	0	0	57	83
Trunk ventral cells	0	0	48	83
Notochord	32	86	151	249
Muscle	31	156	82	818
Not clear	329	609		
Not detected	148	210		
(Multiple tissues)	349	1322		
Total	1213	3423		

*The number of genes that are expressed in these cell types specifically. This number is included in the number given for the nervous system.

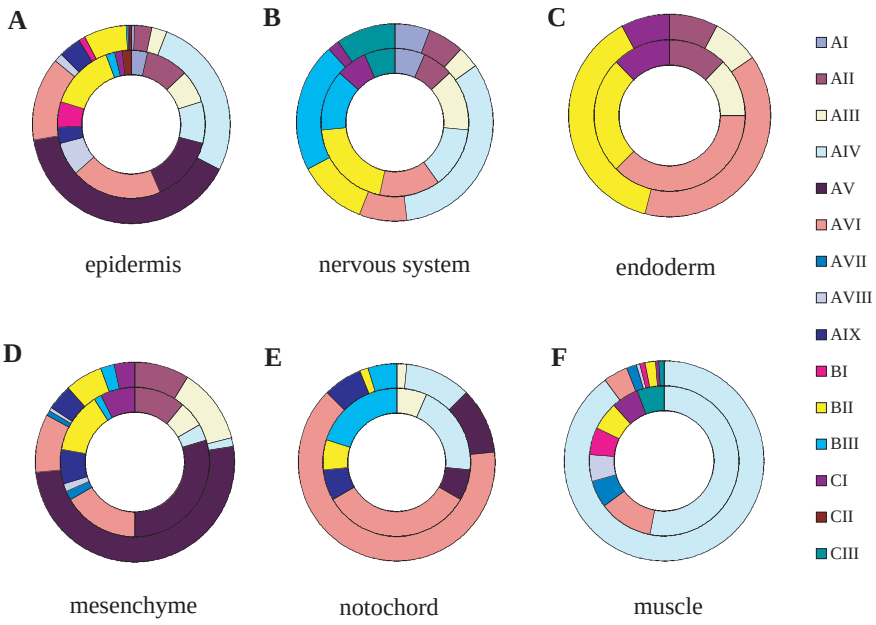

Fig. 1. Proportion of clones (the outer circle) and clusters (the inner circle) for the different classes of genes that are specifically expressed in epidermis (A), nervous system (B), endoderm (C), mesenchyme (D), notochord (E) or muscle (F). The classification of genes is shown in Table 1. Real numbers of genes are listed in Table 4.

Table 4. Relationship between number of specifically expressed genes and their classes

Number of clusters	AI	AII	AIII	AIV	AV	AVI	AVII	AVIII	AIX	BI	BII	BIII	CI	CII	CIII	DI	DII
Epidermis	2	5	4	5	8	11	0	4	2	3	8	1	1	1	0	27	67
Nervous system	1	1	2	2	0	2	0	0	0	0	3	2	1	0	1	3	16
Neuron	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	1*	0*	0*	0*	0*
Brain	0*	0*	0*	0*	0*	1*	0*	0*	0*	0*	2*	0*	0*	0*	0*	2*	1*
Papilla	0*	0*	0*	0*	0*	1*	0*	0*	0*	0*	0*	1*	0*	0*	0*	1*	5*
Nerve cord	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	1*	0*	0*	0*	0*	3*
Endoderm	0	1	1	0	0	3	0	0	0	0	2	0	1	0	0	5	16
Mesenchyme	0	6	3	2	16	9	1	1	4	0	7	1	4	0	0	20	38
Notochord	0	0	1	3	1	5	0	0	1	0	1	3	0	0	0	5	12
Muscle	0	0	0	9	0	2	1	1	0	1	1	0	1	0	1	3	11

Number of clones	AI	AII	AIII	AIV	AV	AVI	AVII	AVIII	AIX	BI	BII	BIII	CI	CII	CIII	DI	DII
Epidermis	2	9	9	90	135	46	0	5	12	4	23	1	1	1	0	224	126
Nervous system	3	3	2	17	0	4	0	0	0	0	6	11	1	0	5	4	17
Neuron	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	1*	0*	0*	0*	0*
Brain	0*	0*	0*	0*	0*	3*	0*	0*	0*	0*	5*	0*	0*	0*	0*	2*	1*
Papilla	0*	0*	0*	0*	0*	1*	0*	0*	0*	0*	0*	1*	0*	0*	0*	2*	5*
Nerve cord	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	0*	10*	0*	0*	0*	0*	4*
Endoderm	0	1	1	0	0	5	0	0	0	0	5	0	1	0	0	9	27
Mesenchyme	0	13	18	2	74	14	1	1	6	0	9	3	5	0	0	32	52
Notochord	0	0	1	7	7	41	0	0	4	0	1	3	0	0	0	8	14
Muscle	0	0	0	118	0	5	2	1	0	1	2	0	1	0	1	9	16

*The number of genes that are expressed in these cell types specifically. This number is included in the number given for the nervous system.

from sequence data. This analysis highlighted a characteristic gene expression profile for each type of tissue (see Fig. 2). Nearly three-quarters of the epidermis-specific genes with distinct functions are in Class A (Fig. 1A). When this ratio is calculated against the clone number, nearly 90% of mRNAs expressed in the epidermis are categorized into Class A or genes that encode for proteins associated with functions that many cell types use (Fig. 1A). Of the genes that are specifically expressed in the nervous system, more than 50% are categorized into Class A (Fig. 1B). When they are examined by clone number, genes for cytoskeletal proteins (subclass AIV) and extracellular matrix (subclass BIII) are predominant (Fig. 1B).

Fig. 1C shows that, as was expected, many endoderm-specific genes are associated with basic cell metabolism (subclass AVI). Interestingly, it was also noticed that many genes for intracellular signaling molecules (subclass BII) are also expressed in the endoderm (Fig. 1C). Many mesenchyme-specific genes are associated with protein synthesis (subclass AV), and more than half of their mRNAs are for that function (Fig. 1D). Fig. 1E shows that the most abundant mRNAs of genes that are expressed specifically in the notochord are associated with metabolism (subclass AVI), suggesting an active metabolism of ascidian notochord cells. As was expected, many muscle-specific genes are associated with cytoskeleton proteins (subclass AIV), and this category of gene accounted for nearly 90% of the clones expressed there (Fig. 1F).

In addition, the present analysis demonstrates that 349 of the 1213 clusters are expressed in multiple tissues simultaneously (Table 3, Fig. 2). This multi-tissue expression pattern was conspicuous in genes that are expressed in TLCs or TVCs (Table 3). Although the present whole-mount in situ hybridization analysis failed to identify genes that are exclusively expressed in either of these two cell types, 57 genes are expressed in TLCs and other tissues, and 48 genes are expressed in TVCs and other tissues. A tissue that showed the simultaneous expression with TLCs was mesenchyme (Table 5); three genes are specifically expressed in TLCs and mesenchyme, 51 genes are expressed simultaneously in TLCs, mesenchyme and other tissues, and only six genes in TLCs and tissues other than mesenchyme. For example, 00141 cDNA encodes oxidoreductase (*P* is 3E-22; Table 5) and is expressed in TLCs and mesenchyme cells (Fig. 2N). 00597 cDNA is another example, which encodes M-TAXREB107 (*P* is 2E-72; Table 5) and is expressed in both TLCs and mesenchyme cells. A recent cell lineage analysis has demonstrated that embryonic mesenchyme cells give rise to tunic cells (free living cells within the tunic) of the adult, while embryonic TLCs give rise to coelomic cells and other mesodermal cells of the adult (Hirano and Nishida, 1997). Therefore, it is likely that both cell types express the same genes in the tailbud embryos.

In contrast, a tissue that showed the simultaneous expression with TVCs was muscle (Table 5); five genes are specifically expressed in TVCs and muscle, 35 genes are expressed simultaneously in TVCs, muscle and other tissues, and 13 genes in TVCs and tissues other than muscle. Four examples are listed in Table 5. 00327 cDNA encodes a *Ciona* counterpart of *Halocynthia* embryonic muscle myosin heavy chain (*P*=6E-96; Table 5), which is expressed in TVCs and muscle cells. 02049 cDNA is another example; it encodes Ca^{2+} transporting

Table 5. Examples of genes that show specific expression

ClusterID	Class	Accession number	Database entry name	Organism	Probability
Epidermis					
00014	AVI	AAC50786	FX. (GDP-4-keto-6-deoxy-D-mannose epimerase-reductase)	*Homo sapiens*	9E-75
00104	DII				
00224	DII				
00333	AVI	BAA10929	Cytochrome P450 like TBP	*Nicotiana tabacum*	9E-18
00605	DII				
01020	AVIII	BAA82365	Chymotrypsinogen 1	*Paralichthys olivaceus*	4E-25
01071	AV	CAA55946	40S ribosomal protein S12	*Sus scrofa*	2E-51
01493	DII				
Nervous system					
00086	AIV	BAA22381	β-tubulin	*Halocynthia roretzi*	E-129
00124	DII				
00218	AVI	AAA21654	ARL3 (ADP-ribosylation factor -like 3)	*Homo sapiens*	2E-66
00271	AIV	AAA28987	α-tubulin 3	*Drosophila melanogaster*	E-118
00531	CIII	JC6179	Dorsal switch protein 1	*Drosophila melanogaster*	3E-47
00607	BII	AAD39365	cGMP-phosphodiesterase delta subunit	*Canis familiaris*	1E-65
00659	AIV	P18258	Tubulin α-1 chain	*Paracentrotus lividus*	7E-40
01087	AII	BAA95118	Etr-1	*Danio rerio*	9E-33
01427	DII				
01880	CI	CAB64386	SoxNeuro	*Drosophila melanogaster*	3E-35
01993	DI	BAA93705	MIWI (piwi)	*Mus musculus*	1E-50
02008	AVI	AAB24700	Cystathionine γ-lyase, cystathionase	*Homo sapiens*	2E-76
02027	DII				
Endoderm					
00194	AVI	CAA55404	Glutathione transferase protein	*Rattus norvegicus*	3E-23
00453	BII	CAA66179	ARFAPTIN 2 (partner of RAC1) (POR1 protein)	*Homo sapiens*	3E-53
00783	DII				
Mesenchyme					
00013	BII	AAC53193	WW domain binding protein 6; WBP6/SRPK-1	*Mus musculus*	8E-76
00127	AV	AAC15655	60S ribosomal protein L37A	*Cryptochiton stelleri*	1E-27
00138	AVI	BAA32086	Natural killer cell enhancing factor	*Cyprinus carpio*	1E-85
00142	DII				
00345	BIII	P41104	60S ribosomal protein L22 (heparin binding protein HBP15)	*Sus scrofa*	4E-27
00389	AIX	CAA37662	Glycoprotein 25L	*Canis familiaris*	3E-50
00693	AV	A47151	Ethionine adenosyltransferase (EC 2.5.1.6)	*Mus musculus*	2E-44
01897	AV	P48588	40S Ribosomal protein S25	*Drosophila melanogaster*	1E-27
Notochord					
00012	AVI	BAA34673	cFKBP/SMAP	*Gallus gallus*	3E-54
00045	AV	CAA40592	Ribosomal protein S1a	*Xenopus laevis*	3E-99
00051	AIV	BAB00623	Tropomyosin-like protein	*Ciona intestinalis*	E-111
00347	AIX	BAB00626	Fibrinogen-like protein	*Ciona intestinalis*	8E-79
00407	AIV	CAA39588	Talin	*Mus musculus*	4E-51
00586	AIV	AAA48986	Nonmuscle myosin heavy chain	*Gallus gallus*	2E-32
01205	AIII	BAA76333	DUF87	*Xenopus laevis*	6E-17
01222	AVI	BAA13447	Mitochondrial thioredoxin	*Bos taurus*	7E-28
01809	DI	AAB61158	Cysteine-rich intestinal protein	*Homo sapiens*	2E-26
02050	BII	AAA48773	Protein tyrosine kinase	*Gallus gallus*	2E-29
Muscle					
00007	DI	AAF26737	Neuroendocrine differentiation factor	*Homo sapiens*	3E-36
00031	AIV	AAA59890	Myosin regulatory light chain	*Homo sapiens*	2E-52
00069	DII				
00173	AIV	AAD09271	Troponin I	*Ciona intestinalis*	5E-92
00189	AIV	AAA86910	Fast-twitch myosin light chain 1	*Bos taurus*	1E-37
00435	AVI	CAA25465	Creatine kinase-M	*Gallus gallus*	1E-28
00488	CIII	Q08945	Structure-specific recognition protein (SSRP1)	*Homo sapiens*	3E-61
00608	AIV	CAA45469	Tropomyosin	*Ciona intestinalis*	1E-49
00645	AVI	AAC78657	Guanosine monophosphate reductase	*Rattus norvegicus*	9E-58
00660	AIV	AAB36559	Fast myotomal muscle tropomyosin	*Salmo salar*	6E-61
01307	AVII	AAA18335	Heat shock protein HSP27	*Mus musculus*	2E-44
01345	BI	AAB59944	Nicotine acetylcholine receptor δ subunit precursor	*Gallus gallus*	4E-21
01458	DII				
01464	AIV	BAA07799	Myosin binding protein C	*Gallus gallus*	9E-44
01633	DII				
01826	CI	CAA62353	Polybromo 1 protein	*Gallus gallus*	3E-18
01868	BII	BAA89462	Caveolin-1 β isoform	*Mus musculus*	4E-25
Mesenchyme and trunk lateral cells					
00141	AVI	AAD36074	Oxidoreductase, aldo/keto reductase family	*Thermotoga maritima*	3E-22
00597	AV	CAA57513	M-TAXREB107	*Mus musculus*	2E-72

Table 5. Continued

ClusterID	Class	Accession number	Database entry name	Organism	Probability
Brain and endodermal strand					
00794	AI	AAF35832	Zinc transporter hZIP2 protein	*Homo sapiens*	7E-25
Notochord, endoderm, trunk ventral cells and muscle					
00152	DI	AB41244	*Homo sapiens* hypothetical protein	*Homo sapiens*	2E-30

ClusterID	Class	Accession number	Protein description	Organism	Probability
Muscle and trunk ventral cells					
00119	AVI	BAA00931	Long-chain acyl-CoA synthetase	*Homo sapiens*	2E-66
00327	AIV	BAA08111	Embryonic muscle myosin heavy chain	*Halocynthia roretzi*	6E-96
01964	AVI	AAD09820	Sarco(endo)plasmic reticulum-type calcium ATPase	*Heliothis virescens*	1E-22
02049	AI	AAA40991	Ca^{2+} transporting ATPase	*Rattus norvegicus*	1E-88

ATPase (*P*=1E-88; Table 5) and is expressed in both TVCs and muscle. Because TVCs are responsible for the formation of adult muscle tissues (Hirano and Nishida, 1997), it is conceivable that the same genes are expressed in muscle and TVCs of the tailbud embryos.

Simultaneous expression of specific genes in multiple tissues was also evident for every tissue. For example, the present analysis demonstrated that not only are there 34 nervous system-specific genes, but 239 genes are also expressed in the nervous system and in other tissues (Table 3); 29 genes are specifically expressed in the endoderm, but an additional 195 genes are expressed in the endoderm and other tissues (Table 3). Two interesting examples are shown in Table 5 and Fig. 2: 00794 cDNA encodes a zinc transporter hZIP2-like protein (*P*=7E-25), the gene for which is expressed specifically in a subset of brain cells and endodermal strand cells (Fig. 2L); 00152 cDNA encodes a protein with no significant sequence information, but this gene is expressed in endoderm, notochord, TVCs and muscle cells (Fig. 2T).

Specific expression patterns

The following are descriptions of representative genes that show tissue-specific expression pattern, although the spatial expression pattern for every gene appears at http://ghost.zool.kyoto-u.ac.jp

Epidermis

A single layer of 800 epidermal cells constitutes the outer most layer of the ascidian tailbud embryo. In *Halocynthia* embryos, cDNA clones for eight different epidermis-specific genes have been isolated and characterized (Ueki et al., 1991; Ueki and Satoh, 1995; Ishida et al., 1996), whereas *Cs-Epi1* and *Cs-Epi2* are two genes that are specifically expressed in *C. savignyi* embryonic epidermal cells (Chiba et al., 1998).

The present analysis of gene expression profiles demonstrate that 149 genes are expressed specifically in the epidermal cells, and a wide variety of other genes in relation to the other functional classes are also expressed there (Table 4; Fig. 1A). However, as is evident in Table 4, 67 of the 149 genes encode for proteins that are not in the database, suggesting a need to further characterize these genes to understand their function in the development of the ascidian epidermis. Eight examples of epidermis-specific genes are listed in Table 5. Among the 149 genes, five were matched to *Halocynthia* epidermis-specific genes. 00014 clone encodes for a protein similar to human GDP-4-keto-6-deoxy-D-mannose epimerase-reductase (*P*=9E-75). As shown in Fig. 2A, gene 00104 is expressed in epidermal cells, except for those along the midline of the embryo; epidermal cells along the midline of the tailbud embryo are thought to be specialized to form the larval fin (Ishida et al., 1996). Interestingly, 00224 is expressed in epidermal cells along the dorsal and ventral midlines (Fig. 2B). This pattern of gene expression was not detected using *Halocynthia* epidermis-specific genes (Ishida et al., 1996). In addition, gene 01493, which shows no sequence similarity with any other genes, was expressed in epidermal cells that reside in the posterior half or tail region of the embryo (Fig. 2C).

Ishida et al. have reported that spatio-temporal expression patterns of eight epidermis-specific genes of *Halocynthia* embryos are categorized into four groups (Ishida et al., 1996). All four patterns were observed using *Ciona* epidermis-specific genes: 133 of the 149 genes are expressed in the entire surface of the embryo, 11 genes are only expressed in the tail epidermis, four are expressed in the midline, and one gene is expressed in all epidermal cells, except for midline cells.

Nervous system

The formation of nervous system of *Ciona* larva has been studied in detail (Katz, 1983; Nicol and Meinertzhagen, 1991; Takamura, 1998). Altogether, the nervous system consists of the central nervous system (CNS) and peripheral sensory nervous system, which consists of approximately 350 cells. The anterior-most of the larva is an adhesive organ or palps (papilla) in which several neuronal cells are differentiated. In the mid-dorsal trunk region is located a brain vesicle with two sensory organs called the otolith and ocellus. The CNS extends into the tail nerve cord, which consists of glial ependymal cells. Peripheral epithelial neuronal cells are scattered over the entire surface of the larva. Although several marker genes for the CNS have been characterized in *Halocynthia* embryos, only one monoclonal antibody has been used to monitor neuronal differentiation in *Ciona* embryos (Takamura, 1998).

The present comprehensive analysis of gene expression profiles in *Ciona* tailbud embryos revealed 34 genes that are expressed exclusively in the nervous system (Table 3). For example, 01087 cDNA encodes for a Etr-1 homolog (Ci-Etr-1; *P*=9E-33; Table 5), and this gene is expressed in almost all of the neuronal cells including palps, brain, nerve cord and epithelial sensory neurons (Fig. 2D). The expression of *Ci-Etr-1* may be used as a pan neuronal differentiation marker in future studies. In addition, the present study revealed genes that are specifically expressed in various subpopulations of nervous system cells: six genes in the brain, eight genes in the palps

and four genes in the nerve cord (Table 3). 00531 cDNA encodes for *Drosophila* dorsal switch protein 1 homolog (*P*=3E-47;Table 5), and the expression of this gene is evident only in the CNS (Fig. 2E). 01993 is *PIWI* homolog (*P*=1E-50; Table 5) and is expressed in a subset of brain cells (Fig. 2F). Two cDNAs, 00659 (Fig. 2G) and 02027 (Fig. 2H) are expressed only in palps, the former encodes for tubulin α-1 chain and the latter for a protein with no sequence similarity (Table 5). The nerve cord consists of four rows of ependymal cells (upper, two lateral and lower) and functions in the guidance of axons from neuronal cell bodies situated in the brain. To date, only one gene (*HrWnt-7*) of *Halocynthia* was identified to be expressed exclusively in the nerve cord (Sasakura and Makabe, 2000), although several genes are reported to be expressed in the nerve cord and other tissues. 00124 encodes a protein with no sequence similarity, but this gene is expressed in the upper and lower rows of nerve cord cells only (Fig. 2I). There is no database information for 01427; however, this gene is expressed in the nerve cord cells only (Fig. 2J).

The nervous system of *Ciona* provides researchers with a simple and primitive model with which to analyze the functional complexities of the vertebrate CNS. Genes that have been identified in the present study should be very useful in future studies.

Endoderm

The endodermal tissue of the *Ciona* tadpole larva constitutes the trunk endoderm and tail endodermal strand. The endoderm consists of about 500 cells. Because the tadpole of most ascidian species is the dispersal phase of their life cycle, they do not open their mouth before attachment to the substrate, which is an early event of metamorphosis. The histochemical detection of alkaline phosphatase (AP) activity is a conventional way to study endoderm differentiation (Whittaker, 1990), and a cDNA clone for the AP gene was isolated in *H. roretzi* (Kumano and Nishida, 1998), and *C. intestinalis* and *C. savignyi* (Imai et al., 2000). The present analysis demonstrates 29 independent genes that are expressed specifically in endodermal cells (Table 3); three examples are listed in Table 5. For example, 00453, which encodes for AFRAPTIN 2 (*P*=3E-53), and 00783, which encodes for a protein with no sequence similarity, are specifically expressed in the endodermal cells (Fig. 2K).

In *Ciona* early embryos, nuclear accumulation of β-catenin is most probably the first step of endodermal cell specification (Imai et al., 2000). The nuclear localization of β-catenin appears to trigger the activation of many transcription factor genes, as well as genes for signal transduction molecules involved in endodermal cell differentiation (Y. Satou et al.,

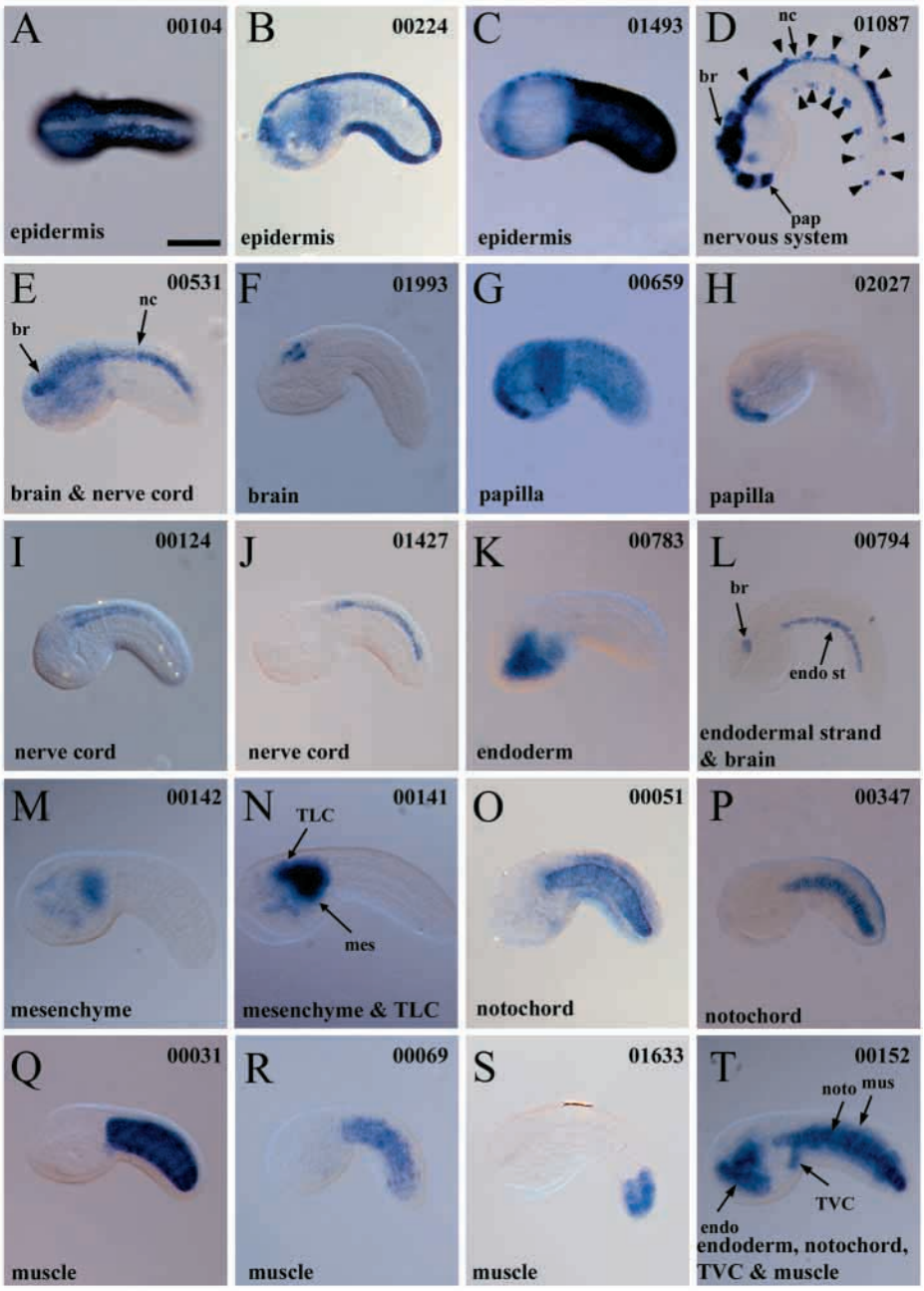

Fig. 2. Whole-mount in situ hybridization showing the gene expression profile specific to organs and/or tissues in *Ciona intestinalis* tailbud embryos. Organs and tissues where the gene is specifically expressed are shown in the bottom left-hand corner, while the clone numbers are indicated in the top right-hand corner. Arrowheads in D indicate epithelial sensory cells. br, brain; endo, endoderm; es, endodermal strand; mes, mesenchyme; mus, muscle; nc, nerve cord; noto, notochord; pap, papilla; TLC, trunk lateral cells; TVC, trunk ventral cells. See text for details. Scale bar: 100 μm.

unpublished). Future studies should ask how these upstream genes control the expression of endoderm-specific genes identified by the present analysis.

Mesenchyme

The *Ciona* tailbud embryo contains four pockets (two pairs) of mesenchyme cells situated in the posterior part of the trunk (Katz, 1983). To date, no mesenchyme-specific genes have been reported in ascidian embryos. Although a cytoplasmic-type actin gene (*HrCA1*) is expressed predominantly in mesenchyme cells in *Halocynthia* embryos, this gene is also expressed in specific neuronal cells in the dorsal trunk region, as well as in notochord cells and in muscle cells (Araki et al., 1996). The present analysis has scored 112 genes that are expressed specifically in mesenchyme cells. As shown in Table 3, many of the different classes of genes are expressed specifically in the mesenchyme cells, and a representative of eight genes are listed in Table 5. For example, 00142 cDNA encodes for a protein with no sequence similarity, which gene is expressed exclusively in mesenchyme cells (Fig. 2M).

It is believed that the mesenchyme of an ascidian embryo is a source of adult mesodermal tissues and therefore this embryonic tissue is rather quiescent in gene activity during embryogenesis. However, as shown in the present study, the second largest number of specific genes were observed in this embryonic tissue, the first largest number of genes were expressed in the epidermis. This suggests that nuclei of embryonic mesenchyme cells are not always quiescent, a new finding that should be explored in future studies.

Notochord cells

The notochord of an ascidian larva is composed of just 40 cells, and the entire lineage has been described completely (Nishida, 1987). Previous studies have revealed that ascidian *Brachyury* genes, *HrBra* (*As-T*) in *H. roretzi* (Yasuo and Satoh, 1993) and *Ci-Bra* in *C. intestinalis* (Corbo et al., 1997), are expressed exclusively in the notochord precursor cells and play a pivotal role in notochord formation (Yasuo and Satoh, 1998). In addition, isolation and characterization of *Ci-Bra* downstream genes has demonstrated that nearly 40 genes are expressed specifically and predominantly in notochord cells (Takahashi et al., 1999). Twenty cDNAs were completely sequenced and found to be ascidian homologs of *Drosophila prickled* genes, genes for antithrombin III-like protein, netrin protein, tropomyosin-like protein, ATP citrate lyase, cdc45, fibrinogen-like protein, pellino-like protein, ATP sulfurylase kinase, tropomyosin-2-like protein, leprecan, ERM, protein tyrosine phosphatase, mad4-like protein and β4-Gal-transferase, and seven other genes had no sequence similarity (Hotta et al., 2000).

The present analysis has revealed that 32 genes are expressed in a notochord-specific manner (Table 3; Fig. 2O,P). For example, 00051 (Table 5; Fig. 2O) is a *Ci-trop* gene that has already been characterized as a target gene of *Ci-Bra* (Di Gregorio and Levine, 1999). 00347 (Table 5; Fig. 2O) encodes a fibrinogen-like protein that was identified as a *Ci-Bra* target gene (Hotta et al., 2000). Interestingly, although 20 of the 32 genes were previously identified as notochord-specific genes, the other 12 genes are newly identified in the present analysis. This raises an intriguing question of whether or not these 12 genes are targets of *Ci-Bra* transcription factor.

Muscle cells

The *Ciona* tailbud embryo develops 18 unicellular, striated muscle cells on each side of the tail (total 36 cells); of them, 28 cells of the anterior and middle part of the tail are derived from B4.1 pair (primary lineage), while four cells at the posterior part and four cells at the tip of the tail originate from the A4.1 and b4.2 pair, respectively (Nishida, 1987). Several muscle-specific genes, including genes for actin (*HrMA4*) and myosin heavy chain (*HrMHC*), have been characterized in *Halocynthia* embryos (reviewed by Satou and Satoh, 1999). In addition, cDNA clones for troponin I (MacLean et al., 1997) and tropomyosin (Meedel and Hastings, 1993) have been isolated from *C. intestinalis.*

The present study has revealed that 31 genes are specifically expressed in muscle cells; 17 examples are listed in Table 5. In addition, 28 of these muscle-specific genes encode structural proteins for muscle cell function; for example, 00031 encodes myosin regulatory light chain (MRLC) and is expressed in all of the muscle cells in the tailbud embryo (Fig. 2Q), 00173 encodes troponin I, 00189 encodes fast-twitch myosin light chain I, 00435 encodes creatine kinase-M and 00608 encodes tropomyosin (Table 5).

Two interesting clones are 00069 (Fig. 2R) and 01633 (Fig. 2S). Both genes encode proteins with no sequence similarity. In *Halocynthia* embryos, *HrMA4* and *HrMHC* begin to be expressed as early as the 32-cell stage (Satou et al., 1995), and *HrMRLC* gene is also expressed at the 32-cell stage (Y. Satou et al., unpublished). Therefore, transcripts of these muscle-specific genes are distributed in the entire cytoplasm of muscle cells at the tailbud stage (Fig. 2Q). However, in situ signals of 00069 (Fig. 2R) are seen mainly in nuclei of muscle cells in the tailbud embryo, suggesting that this gene begins to be expressed around the neurula stage or later. 01633 clone seems to be more interesting, because this gene is expressed in only six pairs of muscle cells located in the posterior region of the tail (Fig. 2S). Cell lineage studies suggest that they are two pairs of b-line muscle cells at the tip of the tail, two pairs of A-line cells at the posterior region and two pairs of B-line muscle cells (Nishida, 1987). All of the muscle-specific genes so far identified are expressed in every muscle cell, although some genes show differences in temporal expressions between the primary (B-line) and secondary (A- and b-lines) lineage muscle cells. The spatial expression of 01633 is therefore unique, and determining the molecular mechanisms that underlie the spatial expression of the 01633 gene should be an intriguing goal of future studies.

In conclusion, the present analysis of a set of 1213 independent clusters derived from the *Ciona intestinalis* tailbud embryos reveals that 502 of them show significant matches to reported proteins, while 184 do not have enough information to be categorized and 527 do not show significant similarities to any known proteins. Sequence similarity analyses of the 502 clusters suggest that 390 of them are associated with functions that many types of cells use, 85 with cell-cell communication, and 27 with transcription factor functions and other gene regulatory proteins. Whole-mount in situ hybridization analysis of all of the 1213 clusters demonstrates that a total of 387 clusters show expressions that are specific to a tissue or organ; 149 show epidermis-specific expression, 34 are specific to the nervous system, 29 to endoderm, 112 to mesenchyme, 32 to notochord and 31 to muscle. Many genes are also specifically

expressed in multiple tissues. The present analysis also highlights characteristic gene expression profiles that are dependent on the type of tailbud tissue and provides new information on many molecular markers for every tissue and organ constituting the *Ciona* tailbud embryo. The sequence information will also be used for future whole genome analysis to explore molecular mechanisms involved in the development of one of the most primitive chordate body plans.

We thank Kazuko Hirayama, Chikako Imaizumi and Reiko Nakamura for technical support. This research was supported by a Grant-in-Aid for Priority Area C (No. 12202001) from the Ministry of Education, Science, Sports and Culture, Japan to N. S. The Academia DNA Sequencing Center is supported by a Grant-in-aid for Scientific Research on Priority Area C (No. 12201001) from the Ministry of Education, Science, Sports and Culture, Japan to Y. K.

REFERENCES

Araki, I., Tagawa, K., Kusakabe, T. and Satoh, N. (1996) Predominant expression of a cytoskeletal actin gene in mesenchyme cells during embryogenesis of the ascidian *Halocynthia roretzi*. *Dev. Growth Differ*. **38**, 401-411.

Chiba, S., Satou, Y., Nishikata, T. and Satoh, N. (1998). Isolation and characterization of cDNA clones for epidermis-specific and muscle-specific genes in *Ciona savignyi* embryos. *Zool. Sci*. **15**, 239-246.

Chomczynski, P. and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159.

Conklin, E. G. (1905). The organization and cell lineage of the acidian egg. *J. Acad. Nat. Sci*. **13**, 1-119.

Corbo, J. C., Levine, M. and Zeller, R. W. (1997). Characterization of a notochord-specific enhancer from the *Brachyury* promoter region of the ascidian, *Ciona intestinalis*. *Development* **124**, 589-602.

Di Gregorio, A. and Levine, M. (1998). Ascidian embryogenesis and the origins of the chordate body plan. *Curr. Opin. Genet. Dev*. **8**, 457-463.

Di Gregorio, A. and Levine, M. (1999). Regulation of *Ci-tropomyosin-like*, a Brachyury target gene in the ascidian, *Ciona intestinalis. Development* **126**, 5599-5609.

Hirano, T. and Nishida, H. (1997). Developmental fates of larval tissues after metamorphosis in ascidian *Halocynthia roretzi*. I. Origin of mesodemal tissues of the juvenile. *Dev. Biol*. **192**, 199-210.

Hotta, K., Takahashi, H., Asakura, T., Saitoh, B., Takatori, N., Satou, Y. and Satoh, N. (2000). Characterization of *Brachyury*-downstream notochord genes in the *Ciona intestinalis* embryo. *Dev. Biol*. **224**, 69-80.

Imai, K., Takada, N., Satoh, N. and Satou, Y. (2000). β-catenin mediates the specification of endoderm cells in ascidian embryos. *Development* **127**, 3009-3020.

Ishida, K., Ueki, T. and Satoh, N. (1996). Spatio-temporal expression patterns of eight epidermis-specific genes in the ascidian embryo. *Zool. Sci*. **13**, 699-709.

Jeffery, W. R. (2001). Determinants of cell and positional fate in ascidian embryos. *Int. Rev. Cytol*. **203**, 3-62.

Katz, M. J. (1983). Comparative anatomy of the tunicate tadpole, *Ciona intestinalis*. *Biol. Bull*. **164**, 1-27.

Kumano, G. and Nishida, H. (1998). Maternal and zygotic expression of the endoderm-specific alkaline phosphatase gene in embryos of the ascidian, *Halocynthia roretzi*. *Dev. Biol*. **198**, 245-252.

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

MacLean, D. W., Meedel, T. H. and Hastings, K. E. (1997). Tissue-specific alternative splicing of ascidian troponin I isoforms. Redesign of a protein isoform-generating mechanism during chordate evolution. *J. Biol. Chem*. **272**, 32115-32120.

Meedel, T. H. and Hastings, K. E. (1993). Striated muscle-type tropomyosin in a chordate smooth muscle, ascidian body-wall muscle. *J. Biol. Chem*. **268**, 6755-6764.

Nakatani, Y. and Nishida, H. (1994). Induction of notochord during ascidian embryogenesis. *Dev. Biol*. **166**, 289-299.

Nakatani, Y., Moody, R. and Smith, W. C. (1999). Mutations affecting tail and notochord development in the ascidian *Ciona savignyi*. *Development* **126**, 3293-3301.

Nicol, D. R. and Meinertzhagen, I. A. (1988). Development of the central nervous system of the larva of the ascidian, *Ciona intestinalis* L. II. Neural plate morphogenesis and cell lineages during neurulation. *Dev. Biol*. **130**, 737-766.

Nicol, D. and Meinertzhagen, I. A. (1991). Cell counts and maps in the larval central nervous system of the ascidian *Ciona intestinalis* (L.). *J. Comp. Neurol*. **309**, 415-429.

Nishida, H. (1987). Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to the tissue restricted stage. *Dev. Biol*. **121**, 526-541.

Nishida, H. (1997). Cell fate specification by localized cytoplasmic determinants and cell interactions in ascidian embryos. *Int. Rev. Cytol*. **176**, 245-306.

Nishida, H. and Satoh, N. (1983). Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. I. Up to the eight-cell stage. *Dev. Biol*. **99**, 382-394.

Nishino, A. and Satoh, N. (2001). The simple tail of chordates: phylogenetic significance of appendicularians. *Genesis* **29**, 36-45.

Ortolani, G. (1955). The presumptive territory of the mesoderm in the ascidian germ. *Experientia* **11**, 445-446.

Sasakura, Y. and Makabe, K. W. (2000). Ascidian *Wnt-7* gene is expressed exclusively in the tail neural tube of tailbud embryos. *Dev. Genes Evol*. **210**, 641-643.

Satoh, N. (1994). *Developmental Biology of Ascidians*. New York: Cambridge University Press.

Satoh, N. (1999). Cell fate determination in the ascidian embryo. In *Cell Lineage and Fate Determination* (ed. S. A. Moody), pp. 59-74. Academic Press.

Satoh, N. and Jeffery, W. R. (1995). Chasing tails in ascidians: developmental insights into the origin and evolution of chordates. *Trends Genet*. **11**, 354-359.

Satou, Y., Kusakabe, T., Araki, I. and Satoh, N. (1995). Timing of initiation of muscle-specific gene expression in the ascidian embryo precedes that of developmental fate restriction in lineage cells. *Dev. Growth Differ*. **37**, 319-327.

Satou, Y. and Satoh, N. (1999). Development gene activities in ascidian embryos. *Curr. Opin. Genet. Dev*. **9**, 542-547.

Simmen, M. W., Leitgeb, S., Clark, V. H., Jones, S. J. M. and Bird, A. (1998). Gene number in an invertebrate chordate, *Ciona intestinalis*. *Proc. Natl. Acad. Sci. USA* **95**, 4437-4440.

Sordino, P., Heisenberg, C-P., Cirino, P., Toscano, A., Giuliano, P., Marino R., Pinto, M.R. and De Santis, R. (2000). A mutational approach to the study of development of the protochordate Ciona intestinalis (Tunicata, Chordata). *Sarsia* **85**, 173-176.

Tabara, H., Motohashi, T. and Kohara, Y. (1996) A multi-well version of in situ hybridization on whole mount embryos of *C. elegans*. *Nucleic Acids Res.* **24**, 2119-2124.

Takahashi, H., Hotta, K., Erives, A., Di Gregorio, A., Zeller, R. W., Levine, M. and Satoh, N. (1999). *Brachyury* downstream notochord differentiation in the ascidian embryo. *Genes Dev*. **13**, 1519-1523.

Takamura, K. (1998). Nervous network in larvae of the ascidian *Ciona intestinalis*. *Dev. Genes Evol*. **208**, 1-8.

Ueki, T., Makabe, K. W. and Satoh, N. (1991). Isolation of cDNA clones for epidermis-specific genes of the ascidian embryo. *Dev. Growth Differ*. **33**, 579-586.

Ueki, T. and Satoh, N. (1995). Sequence motifs shared by the 5′ flanking regions of two epidermis-specific genes in the ascidian embryo. *Dev. Growth Differ*. **37**, 597-604.

Wada, H. and Satoh, N. (2001). Patterning the protochordate neural tube. *Curr. Opin. Neurobiol*. **11**, 16-21.

Whittaker, J. R. (1990). Determination of alkaline phosphatase expression in endodermal cell lineages of an ascidian embryo. *Biol. Bull*. **178**, 222-230.

Yasuo, H. and Satoh, N. (1993). Function of vertebrate *T* gene. *Nature* **364**, 582-583.

Yasuo, H. and Satoh, N. (1998). Conservation of the developmental role of *Brachyury* in notochord formation in a urochordate, the ascidian *Halocynthia roretzi*. *Dev. Biol*. **200**, 158-170.