Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 299 (2002) 74–84
BBRC
www.academicpress.com

EST analysis of mRNAs expressed in neurula of Chinese amphioxus

Chun-yan Mou,[a] Shi-cui Zhang,[b] Jiang-hai Lin,[a] Wen-li Yang,[a] Wen-yan Wu,[a] Jian-wen Wei,[a] Xiao-kun Wu,[a] Jing-chun Du,[a] Zhi-yan Fu,[a] Lan-ting Ye,[a] Yang Lu,[a] Xiao-jin Xie,[a] Yi-liang Wang,[a] and An-long Xu[a,*]

[a] *Department of Biochemistry, State Open Laboratory for Marine Functional Genomics, Guangzhou Center for Bioinformatics, College of Life Sciences, Zhongshan (Sun Yat-Sen) University, 135 Xingangxi Road, Guangzhou 510275, PR China*
[b] *Department of Marine Biotechnology, Ocean University of Qingdao, 5 Yushan Road, Qingdao 266003, PR China*

Received 15 October 2002

Abstract

Amphioxus, a cephalochordate, is the closest living relative to the vertebrates. In order to investigate the molecular mechanisms of the early embryogenesis of amphioxus, we constructed a neurula embryo cDNA library of Chinese amphioxus (*Branchiostoma belcheri tsingtauense*) and generated 5235 expressed sequenced tags in the present study. The initial ESTs consisted of 638 clusters and 1855 singletons, which revealed approximately 2493 unique genes in the data set. Of these sequences, 35.52% ESTs matched to known genes, 12.76% matched to other ESTs, and 51.71% had no match to any known sequences in GenBank. Interestingly we found homologous genes related to neural development and human disease. Bioinformatic analysis showed the direct evidence that the gene homologue found only in vertebrates in previous studies also exists in the amphioxus genome. This study provides a preliminary view of the gene information involved in the development of neurula embryos of Chinese amphioxus and helps our understanding of vertebrate evolution at gene level.

Keywords: Chinese amphioxus; Neurula embryo; Expressed sequence tag

In early embryogenesis of animals, the neurulation is a character of the development of deuterostomes above the level of echinoderms. It is driven by a combination of morphogenetic events within the neural plate ectoderm and nearby areas of epidermal ectoderm and results in the formation of the dorsal, hollow nerve cord [1,2]. In most craniates, neurulation is essentially an invagination of the neural plate while that of others like lampreys and teleosts an cavitation of a solid keel [3]. Though neurulation is presently regarded as one of the best understood examples of vertebrates morphogenesis, leading to the formation of a major organ rudiment, the molecular basis of neurulation remains largely unknown [4].

Amphioxus, a cephalochordate, is thought to be the closest living invertebrate relatives of the vertebrates [5,6]. Its body plan is similar to but simpler than vertebrates. Both have pharyngeal gill slits, a dorsal, hollow neural tube, a notochord, and segmented axial muscles, but amphioxus lacks neural crest and an axial skeleton. The study of amphioxus at gene level will certainly provide evidence to elucidate the evolution of vertebrates.

Amphioxus embryology is well understood from classical approaches [7,8] and neurular development has been characterized as representing the common and generalized embryonic form of a vertebrate ancestor [9].

In the neurula of amphioxus, the nonneural ectoderm detaches from the edges of the neural plate and fuses dorsally via lamellipodia over the forming neural tube. The notochord and the mesoderm are formed by folding from the dorsomedian and the dorsolateral wall of the archenterons. Subsequently, the neural plate rolls up into the neural tube which encloses a neural canal. At last a typical triploblastic embryo is formed, consisting of

* Corresponding author. Fax: +86-20-84038377.
E-mail address: ls36@zsu.edu.cn (A.-l. Xu).

0006-291X/02/$ - see front matter © 2002 Elsevier Science (USA). All rights reserved.
PII: S0006-291X(02)02582-2

definitive ectoderm, neural tube, notochord, endoderm, and mesoderm [10]. This is possibly the primitive type of neurulation in deuterostomes [3]. The molecular investigation of the neurula of amphioxus can provide important clues to understand the development of vertebrate nervous system because the expression pattern of some amphioxus genes has been found to be similar to that of the vertebrate counterpart in previous study. Among many available techniques of molecular investigation, e.g., serial analysis of gene expression [11], differential display [12], subtraction cDNA library [13], cDNA microarray, and chip technologies [14,15], we chose cDNA library construction and expressed sequence tag (EST) sequencing to reveal the genes expressed in the neurula stage of amphioxus. Here we report the characterization of ESTs derived from neurula embryo cDNA library of Chinese amphioxus. A total of 5235 ESTs composed of 2493 clusters were generated in the present study. Analysis of the ESTs may help understanding the development of embryogenomics of this model organism, particularly a global understanding of early development of chordate nervous system.

Materials and methods

cDNA library construction. Matured adults of a Chinese amphioxus, *Branchiostoma belcheri tsingtauense*, were obtained from Kioachow Bay near Qingdao, China, about a week before the onset of breeding seasons in 2000. Animals were kept in large earthen vessels with aeration and supplied with fresh seawater and plankton daily [10]. Females swimming up to lay eggs from the sand were caught with a net and immediately put into a large petri dish containing naturally inseminated seawater [15]. The neurulae were collected and quickly frozen in liquid nitrogen for cDNA library construction. Total RNA was extracted using TRIzol reagent (Gibco-BRL) according to supplier's method. cDNA was prepared using SMART cDNA library construction kit (Clontech) following manufacturer's instructions. The cDNA was ligated into pcDNA3.0 and electroporated into *Escherichia coli* DH5α cells using a Gene Pulser II electroporation system under standard conditions. The library contained 6×10^5 independent clones. A total of 10,000 independent cDNA clones were picked randomly and stored at −80 °C for further analysis.

EST sequencing. Clones were thawed, inoculated directly into 96-well plates containing 1 ml LB broth, and cultured overnight. The DNAs were extracted using Vitagene 96-easy plasmid Miniprep Kit (Vitagene Biochemical Technique). The 5′-end sequencing of each cDNA was conducted in an automated ABI PRISM 3700 sequencer (Perkin–Elmer), using ABI PRISM Big-Dye Terminator v3.0 Ready Reaction Cycle Sequencing Kit (Applied Biosystems) and T7 primer.

Sequence data analysis. Sequences were edited by using programs to remove vector sequences and ambiguous regions, and then assembled into groups of sequences (clusters). The consensus sequence of each cluster was used as query sequences to search against GenBank DNA sequences with TFASTA and TBLASTX [16]. The original sequence data and analytic results were kept in our database at website (http://192.168.0.111). Annotations of possible protein-coding genes were performed and assembled for future study.

Results and discussion

Overview of ESTs from the Chinese amphioxus early embryos

A cDNA library was constructed from early neurula embryos of Chinese amphioxus. The average length of the insert cDNA fragments is 1 kb, ranging from 0.5 to 3 kb. The cDNA clones were randomly selected and partially sequenced to generate ESTs. Sequences shorter than 100 bases were removed and discarded. Seven ESTs consisting of vector sequences were subtracted from the initial ESTs prior to analysis. A total of 5235 ESTs were examined in the present study. The occurrences of the readable sequence lengths are shown in Fig. 1. A large fraction of the sequences ranged from 450 to 750 bp. The average length of the reads on which the following analysis was based was 524.88 bp. The initial ESTs were grouped into 2493 consensus sequences, consisting of 638 clusters that contained more than two ESTs each cluster and 1855 singletons. Distribution of those ESTs is as follows:

(1) Sixteen clusters that comprised of more than 20 ESTs each cluster contain the most abundant transcripts. They constituted 0.64% of the total contig (16 of 2493 clusters) and included 20.68% of total ESTs (1083 of 5235 ESTs). Nine of them are genes already identified in amphioxus such as cytochrome c oxidase subunits I, II, and III, NADH dehydrogenase subunits 1, 4, and 5, actin, and intermediate filament. The other six clusters are unknown genes that show no significant similarity to the genes in GenBank. It is very interesting that cluster C00444 consists of 614 clones. The consensus sequence of this gene is 1400 bp. ORF finder (open reading frame finder) analysis identified frame 2 from 551 to 652 bp and deduced 34 amino acid codons. Why the transcript of this gene is so abundant and what the possible role

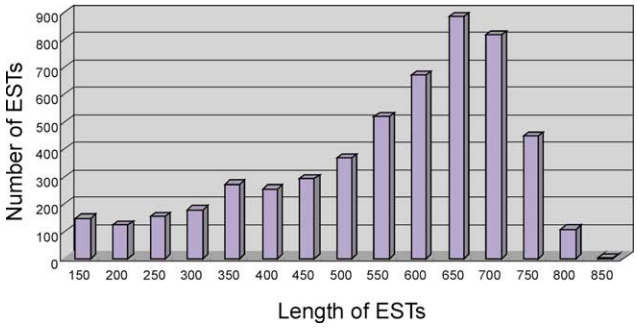

Fig. 1. The distribution of the readable length of the ESTs. A total of 5235 ESTs were generated in the present study. The ESTs consisting of vector sequences and sequences shorter than 100 bp and ambiguous regions were discarded before analysis. The readable sequence lengths ranged from 100 to 850 bp. The average length of the reads on which the following analysis was based was 524.88 bp. Abscissa (50 bp) is the length of sequences, while the *Y*-coordinate is the number of ESTs. A large fraction of ESTs is between 450 and 750 bp.

implicates in this developing stage are not clear. Experiments are being performed to identify its function. The highest expressed genes in assembled clusters are shown in Table 1.

(2) Thirty-four clusters consisting of 10–19 ESTs each cluster consisted of 1.36% of total clusters (34 of 2493) and 8.46% of total ESTs (443 of 5235). They are represented by 12 clusters of unknown genes and recognized ones such as creatine kinase, calmodulin, NADH dehydrogenase subunit 2, catechol-*O*-methyltransferase, and SEC61. They are second abundant mRNA transcripts in the early developing embryos.

(3) Five hundred and eighty eight clusters contained 2–9 ESTs each cluster representing 23.58% of clusters (588 of 2493 clusters) and 35.41% of ESTs (1854 of 5235 ESTs) such as calcium-binding gene, elongation factor TU, calmodulin, DNA directed RNA polymerase II, myogenic determination factor, selenium donor protein. They are medium-sized clusters with low prevalence.

(4) One thousand eight hundred and fifty five clusters were unique sequences consisting 74.40% of clusters (1855 of 2493 clusters) and 35.43% of ESTs (1855 of 5235 ESTs). The occurrence rate for these clusters is only once in the present study. They contain regulatory proteins, transcription factors, and genes without similarities to the public databases. The distribution of the cluster sizes is shown in Fig. 2.

The cDNA library we used is a non-normalized primary one without amplification. The clone abundance or the cluster size will reflect the relative mRNA population. About one-third of the total clones belong to singleton, representing the complexity of the mRNA population of this developmental stage.

Analysis of the ESTs by matching with GenBank

The consensus sequences of each cluster were submitted to public databases to search sequence similarities. According to the blast results, 5235 ESTs were divided into three categories: (I) 35.53% of total ESTs (1860/5235) belonging to recognized protein-coding sequences with strong matches ($P < 10^{-12}$); (II) 12.76% (668/5235) being similar to functionally unidentified ESTs or cosmid sequences; and (III) 51.71% (2707/5235) having no significant similarity to any known sequences including sequences with matches higher than 10^{-12} as described in Fig. 2.

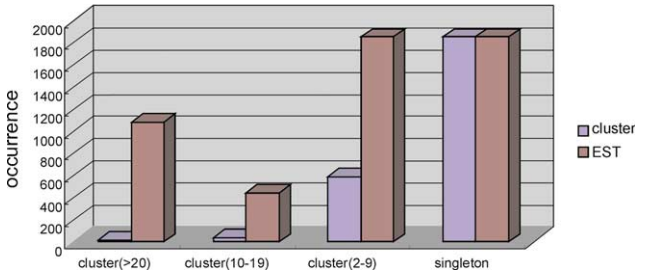

Fig. 2. Prevalence distribution of the cluster size. The initial 5235 ESTs were grouped into 2493 clusters, consisting of 16 clusters (469 of 5235 ESTs) that comprised of more than 20 ESTs each cluster, 34 clusters contained 10–19 ESTs (443 of 5235 ESTs), 589 clusters comprised 2–9 ESTs each cluster (2468 of 5235 ESTs), and 1855 unique sequences. Most ESTs belonged to medium-sized clusters and singletons representing the complexity of the mRNA stored in the neurula of Chinese amphioxus.

Our analysis showed that 45 genes out of our dataset were previously identified in amphioxus. The remaining clusters were new genes and could be very interesting for further analysis. In category (I) 1860 ESTs, consisting of 661 distinct coding proteins, were homologous to the previously identified proteins with the average number of hits/protein 2.81 (1860/661). According to the previous classification [17], 661 proteins can be divided into three major classes: 557 proteins consisting of 1591 ESTs belonging to class A (AI–AX), which comprised a large fraction of the recognized protein-coding mRNA transcripts, containing structural and enzymatic housekeep-

Table 1
Assembled clusters that correspond to the highest expressed genes

Cluster	Gene description	Organism	ESTs
C00029	NADH dehydrogenase subunit 4	*Branchiostoma lanceolatum*	20
C00414	Unknown	Unknown	21
C00101	Intermediate filament protein D1	*Branchiostoma lanceolatum*	21
C00102	Unknown	Unknown	24
C00078	Unknown	Unknown	24
C00015	Hypothetical protein	Mouse	26
C00116	Unknown	Unknown	27
C00124	Cytoplasmic actin	*Branchiostoma floridae*	28
C00104	NADH dehydrogenase subunit 5	*Branchiostoma lanceolatum*	30
C00111	Cytochrome *c* oxidase subunit I	*Branchiostoma lanceolatum*	32
C00117	Unknown	*Caenorhabditis elegans*	33
C00136	NADH dehydrogenase subunit 1	*Branchiostoma lanceolatum*	43
C00144	Cytochrome *c* oxidase subunit II	*Branchiostoma floridae*	46
C00139	Cytochrome *c* oxidase subunit III	*Branchiostoma lanceolatum*	45
C00209	ATP synthase F0 subunit 6	*Branchiostoma floridae*	49
C00444	Unknown	Unknown	614

ing proteins associated with the functions of many different cells; class B (BI–BIII), consisting of 68 proteins (about 178 ESTs) associated with cell–cell communication; class C (CI–CIII), including 36 transcription factors and other gene regulatory proteins (91 ESTs altogether) as described in detail in Table 2. In class DI, 668 clones represented by 363 clusters (14.56% of the total clusters) were matched to ESTs or hypothetical proteins (mostly from *Homo sapiens*). The remaining clusters (66.14% of total clusters) were found to be no matches to the known sequences and further study should be carried out to reveal their functions. The number of partial mRNA transcripts represented in each category is listed in Table 3.

ESTs relevant to development

The most abundant genes reported here were found to be involved in intermediary synthesis and catabolism of enzymes implying the rapid growth of the embryo body and the formation of the ectoderm, mesoblast, endoblast, definitive notochord, and neural tube, which is a typical feature for neurula embryo. In addition, neurula embryos display distinct morphological polarity along the dorsoventral and anteroposterior axis, which is controlled by a series of genes involving BMP signaling pathway regulated by twisted gastrulation gene (Tsg). We found that a cluster homologous to Tsg, a component of BMP signaling pathway, which plays an important role to regulate embryonic dorsal–ventral patterning in flies, frogs, and fish, also influences dorsoventral polarity of the neural tube [18,19]. Identification of this gene in amphioxus suggested the conservative role of this gene in patterning formation of embryos. We also found a cluster similar to short form protein of one-eyed pinhead, which belongs to the EGF-CFC gene family essential for nodal signaling, functioned on organizer patterning. The EGF-CFC gene family plays important roles in germ layer formation, correct positioning of anterior–posterior (A–P) and left–right (L–R) axis [20,22]. Mutation of this gene resulted in cyclopia and defects in formation of endoderm, prechordal plate, and ventral neuroectoderm during zebrafish embryonic development [21]. The expressing pattern of this gene in developing embryos may identify the real function on A–P or L–R patterning in this invertebrate.

Table 2
The distribution of the genes

	Classes of gene function and gene anotation	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	28	116
AII	RNA processing, polymereizing, splicing, and binding proteins, and enzymes	39	64
AIII	Cell replication, histones, cyclins and kinase, DNA polymerase, topoisomerases, DNA modification	42	103
AIV	Cytoskeleton and membrane proteins	79	200
AV	Protein synthesis co-factors, tRNA synthetases, ribosomal proteins	90	264
AVI	Intermediary synthesis and catabolism enzymes	138	587
AVII	Stress response, detoxification, and cell defense proteins	38	79
AVIII	Protein degradation and processing, proteases	45	81
AIX	Apoptosis-related	12	17
AX	Transportation and binding proteins for proteins and other macromolecules	46	80
	Total	557	1591
B	Cell–cell communication		
BI	Signaling recepors, including cytokine and hormane receptors, and signal intermediates	6	12
BII	Intercellular signal transduction pathway molecules including kinase and signal intermediates	54	148
BIII	Extacellular matrix proteins and cell adhesion	8	18
	Total	68	178
C	Transcription factors and toher gene regulatory proteins		
CI	Sequence-specific DNA-binding proteins	24	61
CII	Non-DNA-binding proteins that perform positive or negative roles	12	30
	Total	36	91
D	Others		
DI	Not enough information to classify	363	668
DII	No significant similarities to known proteins	1649	2707
	Total	2493	5235

Table 3
EST sequence similarities, gene description and probability of occurrence by chance

Class	Cluster ID	Accession No.	Database entry name	Organisms	Probability
AI	Transportation and binding proteins for ions and other small molecules				
	C01599	P21282	V-ATPase C subunit	*Bos taurus*	1.00E – 50
	C00377	NP_062278	Caltractin	*Mus musculus*	1E – 78
	C00209	NP_007760	ATP synthase F0 subunit 6	*Branchiostoma floridae*	2E – 83
	C00707	NP_001976	Electron-transfer-flavoprotein	*Homo sapiens*	5E – 85
	C00386	AAF73513	Voltage-dependent anion channel	*Gallus gallus*	1E – 73
	C01637	AAF26679	Voltage-dependent calcium channel β1B subunit	*Bos taurus*	6E – 61
	C01399	P49946	Frih_salsaferritin	*Salmo salar*	1E – 55
	C00038	AAG37428	ISCU2	*Homo sapiens*	6.00E – 60
AII	RNA processing, polymeizing, splicing and binding proteins, and enzymes				
	C00468	XP_059317	Polymerase (RNA) II	*Homo sapiens*	8E – 40
	C01095	XP_035490	Testis enhanced gene transcript	*Homo sapiens*	1E – 51
	C01612	AAH07742	Cleavage and polyadenylation specific factor 2	*Homo sapiens*	2E – 66
	C01524	NP_057018	Nucleolar protein NOP5/NOP58	*Homo sapiens*	1E – 79
	C01346	NP_008937	Cleavage and polyadenylation specific factor 5	*Homo sapiens*	e – 103
	C00546	NP_005863	Breast carcinoma amplified sequence 2	*Homo sapiens*	5E – 83
	C00345	NP_002687	DNA-directed RNA polymerase II polypeptide G	*Homo sapiens*	3E – 93
	C00183	A42811	Nuclear RNA helicase	*Rattus norvegicus*	4E – 85
	C00818	A32618	DNA-directed RNA polymerase	*Homo sapiens*	6E – 75
	C00514	BAB62225	Hu/elav class neuron-specific RNA-binding protein	*Branchiostoma belcheri*	4E – 60
	C00397	I59377	Template activating factor-I	*Homo sapiens*	2E – 55
AIII	Cell replication, histones, cyclins and kinase, DNA polymerase, topoisomerases, and DNA modification				
	C00368	XP_049960	SWI/SNF-related, matrix-associated	*Homo sapiens*	3E – 50
	C01291	Q91684	DNA polymerase gamma subunit 1	*Xenopus laevis*	5E – 67
	C01096	P56520	Histone deacetylase 3	*Gallus gallus*	e – 112
	C00100	P08991	Histone H2A variant	*Strongylocentrotus purpuratus*	2E – 51
	C01237	NP_004520	Myeloid/lymphoid or mixed-lineage leukemia	*Homo sapiens*	1E – 56
	C01062	NP_112538	Mitogen-activated protein-binding protein interacting protein	*Mus musculus*	2E – 48
	C01120	NP_191019	Histone H2A.F/Z	*Arabidopsis thaliana*	2E – 45
	C00582	NP_478143	Histone H4	*Drosophila melanogaster*	1E – 39
	C00336	P50532	Chromosome assembly protein XCAP-C	*Xenopus laevis*	2.00E – 63
	C00212	NP_009037	Ubiquitin-like 3	*Homo sapiens*	7E – 49
	C001474	NP_002085	G1 to S phase transition 1	*Homo sapiens*	8E – 51
	C01474	CAC20564	PD2 protein	*Homo sapiens*	1E – 71
AIV	Cytoskeleton and membrane proteins				
	C00101	CAA11446	Intermediate filament protein D1	*Branchiostoma lanceolatum*	e – 176
	C00858	XP_006242	Integral membrane protein 1	*Homo sapiens*	9E – 92
	C00124	Q93131	Actin, cytoplasmic	*Branchiostoma floridae*	e – 127
	C001784	Q27203	4-Hydroxyphenylpyruvate dioxygenase	*Tetrahymena thermophila*	6E – 56
	C00315	P28287	Tubulin α chain	*Oxytricha granulifera*	e – 115
	C001296	P12716	Actin, cytoplasmic	*Pisaster ochraceus*	4E – 98
	C001299	P02578	Actin 1	*Acanthamoeba castellanii*	e – 115
	C00845	NP_523517	COP9 complex homolog subunit 2	*Drosophila melanogaster*	1E – 88
	C00145	NP_523366	Cyclophilin 1	*Drosophila melanogaster*	1E – 67
	C01627	NP_476489	Septin 2	*Rattus norvegicus*	3E – 65
	C00615	NP_291024	Myosin regulatory light chain	*Homo sapiens*	2E – 81
	C00590	NP_031617	Calreticulin	*Mus musculus*	2E – 79
	C00052	BAA94967	ϵ1-COP	*Bos taurus*	1E – 82

Table 3 (*continued*)

Class	Cluster ID	Accession No.	Database entry name	Organisms	Probability
AV	Protein synthesis co-factors, tRNA synthetases, ribosomal proteins				
	C00734	Q9GR88	Eukaryotic peptide chain release factor subunit 1	*Polyandrocarpa misakiensis*	3E – 72
	C00426	Q26481	60S ribosomal protein L5	*Styela clava*	2E – 79
	C00014	P29520	Elongation factor 1-α	*Bombyx mori*	5E – 92
	C00224	NP_230016	Elongation factor TU	*Vibrio cholerae*	2.00E – 87
	C00316	O61231	60S ribosomal protein L10	*Drosophila melanogaster*	e – 106
	C01373	NP_243976	Phenylalanyl-tRNA synthetase β subunit	*Bacillus halodurans*	1.00E – 23
	C00272	O01727	40S ribosomal protein S6	*Branchiostoma floridae*	e – 104
	C01712	NP_036182	Ribosomal protein S3	*Mus musculus*	2E – 89
	C00980	NP_005042	Glutaminyl-tRNA synthetase	*Homo sapiens*	1E – 61
	C00966	NP_001003	Ribosomal protein S8	*Homo sapiens*	1E – 91
	C01436	NP_231888	Ribosome recycling factor	*Vibrio cholerae*	4.00E – 42
	C00013	BAB63215	EF-1a	*Branchiostoma floridae*	3E – 61
	C01467	BAA92160	Eukaryotic polypeptide chain release factor 3	*Oryctolagus cuniculus*	2E – 92
AVI	Intermediary synthesis and catabolism enzymes				
	C01561	XP_037869	Pyruvate dehydrogenase kinase	*Homo sapiens*	1E – 76
	C00007	Q27238	ADP, ATP carrier protein	*Anopheles gambiae*	5E – 70
	C00595	P56966	Geranylgeranyl pyrophosphate synthetase	*Bos taurus*	2E – 82
	C00579	P49088	Asparagine synthetase	*Rattus norvegicus*	1E – 66
	C00677	P46472	26S protease regulatory subunit 7	*Xenopus laevis*	4E – 80
	C00200	P42026	NADH-ubiquinone oxidoreductase	*Bos taurus*	3E – 84
	C00601	P26990	ADP-ribosylation factor 6	*Gallus gallus*	6E – 72
	C00538	P10658	Phosphoserine aminotransferase	*Oryctolagus cuniculus*	3E – 90
	C01757	NP_488456	Cystathionine β-synthase	*Nostoc* sp. PCC 7120	6E – 75
	C00318	NP_036379	Selenophosphate synthetase	*Homo sapiens*	6E – 61
	C00144	NP_007758	Cytochrome c oxidase subunit II	*Branchiostoma floridae*	e – 108
	C01754	CAA04917	Phenylalanine hydroxylase	*Branchiostoma floridae*	6E – 83
	C01615	A24050	Ribonucleoside-diphosphate reductase	*Mus musculus*	1E – 80
AVII	Stress response, detoxification, and cell defense proteins				
	C01127	XP_001655	HSPCO34 protein	*Homo sapiens*	4E – 33
	C01104	S37284	Cytochrome P450 2D	*Bos taurus*	1E – 34
	C00463	P46633	HSP 90-ALPHA	*Cricetulus griseus*	2E – 38
	C00281	BAB70508	Prostate cancer antigen-1	*Homo sapiens*	5.00E – 84
	C01305	P08108	Heat shock cognate 70	Oncorhynchus mykiss	6E – 83
	C00423	NP_112478	Tissue specific transplantation antigen P35B	*Mus musculus*	e – 103
	C01632	NP_062266	Polymyositis/scleroderma autoantigen 1	*Mus musculus*	1E – 49
	C00683	NP_032857	Prohibitin	*Mus musculus*	1E – 45
	C00534	CAC38780	Allograft inflammatory factor 1	*Suberites domuncula*	8E – 39
	C01713	CAB89875	CD81 protein	*Saguinus oedipus*	3E – 27
	C00917	AAH00864	Sjogren*s syndrome nuclear autoantigen 1	*Homo sapiens*	1E – 38
AVIII	Protein degradation and processing, proteases				
	C00217	S38529	Multicatalytic endopeptidase complex	*Xenopus* sp.	2E – 91
	C01725	Q03168	Lysosomal aspartic protease precursor	*Aedes aegypti*	2E – 30
	C01570	P30759	Arginase	*Xenopus laevis*	2E – 43
	C00415	NP_064335	Signal peptidase complex (18 kDa)	*Mus musculus*	7.00E – 86
	C00996	NP_071720	Cathepsin Z	*Mus musculus*	5E – 87
	C01052	NP_001152	Nudix	*Homo sapiens*	1.00E – 40
	C00794	NP_005330	Huntingtin interacting protein 2	*Homo sapiens*	4.00E – 72
	C01336	NP_003336	Ubiquitin-conjugating enzyme E2I	*Homo sapiens*	3E – 82
	C01821	CAB50892	Ubiquitin fusion protein	*Kluyveromyces lactis*	2E – 58

Table 3 (*continued*)

Class	Cluster ID	Accession No.	Database entry name	Organisms	Probability
AIX	Apoptosis-related				
	C01993	XP_005698	programmed cell death 4	*Homo sapiens*	7E – 30
	C01952	NP_003066	SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily c, member 2; Rsc8	*Homo sapiens*	4E – 92
		XP_035490	testis enhanced gene transcript (BAX inhibitor 1)	*Homo sapiens*	1E – 51
	C00829	JC7093	Fas-associated factor 1	*Homo sapiens*	5E – 32
AX	Transportation and binding proteins for proteins and other macromolecules				
	C01318	XP_034765	Vesicle trafficking protein sec22b	*Homo sapiens*	1.00E – 52
	C01478	XP_010296	Transient receptor potential channel 5	*Homo sapiens*	9E – 23
	C00616	Q9PRL8	Acyl-CoA-binding protein	*Gallus gallus*	4E – 31
	C00537	Q94519	Acyl carrier protein	*Drosophila melanogaster*	5E – 34
	C01639	NP_511039	Solute carrier family 35	*Mus musculus*	1E – 53
	C00566	NP_477137	UbcD2-P1	*Drosophila melanogaster*	4E – 30
	C01268	NP_064654	Solute carrier family 37	*Mus musculus*	2E – 37
	C00431	AAD34970	T-complex polypeptide 1	*Danio rerio*	2.00E – 97
	C00568	NP_058927	Phosphatidylinositol transfer protein	*Rattus norvegicus*	2E – 76
	C01108	NP_055043	Solute carrier family 6, member 7	*Homo sapiens*	1E – 48
	C00446	NP_031664	Chaperonin subunit 7	*Mus musculus*	5.00E – 74
	C00619	NP_002257	Karyopherin α 2	*Homo sapiens*	2E – 41
	C01771	NP_001275	Adaptor-related protein complex 3, sigma 1 subunit	*Homo sapiens*	1E – 82
	C00380	AAK50397	GDP-fucose transporter	*Homo sapiens*	4E – 72
BI	Signaling recepors, including cytokine and hormane receptors, and signal intermediates				
	C00325	AAK27327	Twisted gastrulation protein	*Xenopus laevis*	4E – 27
	C00453	AAC04339	One-eyed pinhead short form protein	*Danio rerio*	6E – 20
	C01386	AAF27548	PTP36-A isoform	*Mus musculus*	2E – 36
BII	Intercellular signal transduction pathway molecules including kinase and signal intermediates				
	C00542	O45040	Guanine nucleotide-binding protein	*Homarus americanus*	e – 172
	C00941	Q05975	RAS-related protein RAB-2	*Lymnaea stagnalis*	1E – 57
	C00842	P79735	GTP-binding nuclear protein Ran	*Danio rerio*	e – 101
	C00550	NP_524175	Neurocalcin	*Drosophila melanogaster*	1E – 72
	C01376	NP_005361	Mel transforming oncogene	*Homo sapiens*	2E – 80
	C00208	AAB97725	Calumenin	*Homo sapiens*	4E – 74
	C01183	AAL31764	cAMP-specific phosphodiesterase isoform	*Rattus norvegicus*	1E – 34
	C00487	AAK29780	Creatine kinase	*Branchiostoma floridae*	e – 125
	C01750	AAG33129	MER receptor tyrosine kinase	*Homo sapiens*	2E – 72
	C00235	NP_509939	14-3-3 protein	*Caenorhabditis elegans*	4E – 89
	C00976	NP_062652	Trk-fused gene	*Mus musculus*	5E – 55
	C00657	AAD12256	GTP-binding protein	*Gallus gallus*	3E – 86
BIII	Extracellular matrix proteins and cell adhesion				
	C01552	P26043	Radixin	*Mus musculus*	2E – 80
	C01160	NP_057725	Membrane interacting protein of RGS16	*Homo sapiens*	4E – 31
	C00526	I51670	Focal adhesion kinase pp125FAK	*Xenopus laevis*	1E – 83
	C02023	AAG00570	Embryonic blastocoelar extracellular matrix protein precursor	*Lytechinus variegatus*	3E – 49
	C00247	NP_033268	Secreted acidic cysteine rich glycoprotein	*Mus musculus*	7E – 55
CI	Sequence-specific DNA-binding proteins				
	C01117	XP_041992	Kruppel-like zinc finger protein	*Homo sapiens*	4E – 31
	C00991	NP_079517	UBX domain-containing 1	*Homo sapiens*	1E – 47
	C00433	NP_056647	Methyl-CpG-binding domain protein 2	*Homo sapiens*	3E – 32

Table 3 (*continued*)

Class	Cluster ID	Accession No.	Database entry name	Organisms	Probability
	C01055	NP_035780	Translin	*Mus musculus*	6E – 39
	C01443	NP_033580	Zinc finger protein 37	*Mus musculus*	9E – 54
	C00249	NP_005639	Transcription elongation factor B	*Homo sapiens*	1E – 56
	C02161	BAA01477	Zinc finger protein	*Mus musculus*	2E – 98
	C00012	XP_067318	Elongation factor 1-α 1	*Homo sapiens*	2E – 95
	C01049	NP_006521	Glioma-amplified sequence-41	*Homo sapiens*	4E – 47
	C00309	NP_037374	Mouse Glt3	*Homo sapiens*	e – 107
CII	Non-DNA-binding proteins that perform positive or negative roles				
	C00436	XP_012145	Small nuclear ribonucleoprotein polypeptide F	*Homo sapiens*	7E – 36
	C00611	NP_035405	Ribonuclease H1	*Mus musculus*	5E – 48
	C00296	NP_004166	Small nuclear ribonucleoprotein D3 polypeptide	*Homo sapiens*	5E – 40
	C00405	NP_003085	Small nuclear ribonucleoprotein polypeptide E	*Homo sapiens*	4E – 35
	C01351	CAC44272	XNop56 protein	*Xenopus laevis*	4E – 19
	C00987	BAA37095	Ribonucleoprotein F	*Rattus norvegicus*	3E – 33

Four clusters, homologous to neurocalcin [23], choline cotransporter [24,25], Zic3-binding protein [26], and CYFIP [27], respectively, were first identified in amphioxus in our study. These four genes play important roles in neuronal functions and neural development. NC, a neuron-specific EF-hand Ca^{2+}-binding protein belonging to a novel family of neuronal calcium sensors, was found primarily in vertebrate brain and retina. It was proposed to play a role in calcium-dependent regulation of enzymes in signal transduction pathways and be involved in control of neuron function [23]. CHT is essential to cholinergic transmission and rate limiting in the biosynthesis of acetylcholine. The expression of CHT was restricted to brain regions rich in cholinergic neurons including the putamen, spinal cord, and medulla in human, mouse, and *C. elegans* [24,25]. Zic3-binding protein was expressed mainly in the developing nervous system, with high levels of expression in the proliferating neuroepithelium of the brain and the neural tube, and the dorsal root ganglia during embryonic development in the mouse [26]. CYFIP is shown to interact with the small GTPase Rac1, which devotes to the development and maintenance of neuronal structures [27]. Genes mentioned above are listed in Table 4.

Table 4
Clusters homologous to the genes found in the previous study

Cluster ID	Accession No.	Gene description	Organisms
C00514	BAB62225	Hu/elav class neuron-specific RNA-binding protein	*Branchiostoma floridae*
C01476	AAG33015	Mnx homeodomain protein	*Branchiostoma floridae*
C00971	AAF19841	Bone morphogenetic protein 2/4	*Branchiostoma belcheri*
C01583	BAA78620	AmphiHox1	*Branchiostoma floridae*
C00453	AAC04339	One-eyed pinhead short form protein	*Danio rerio*
C01362	AAG41055	Choline cotransporter	*Limulus polyphemus*
C00550	NP_524175	Neurocalcin	*Drosophila melanogaster*
C01119	AAD22979	Zinc3-binding protein	*Xenopus laevis*
C01535	JC5496	Prox 1 protein 671	Chicken
C00669	AAK64276	Ephrin B2b	*Danio rerio*
C00913	AAG61253	Cyfip	*Danio rerio*
C01828	BAB40596	Ci-META1	*Ciona intestinalis*
C00713	CAB56698	Sec61β protein	*Drosophila melanogaster*
C01786	P97603	Neo1_rat neogenin precursor gb	*Rattus norvegicus*
C02023	AAG00570	Embryonic blastocoelar extracellular matrix protein precursor	*Lytechinus variegatus*
C00434	AAK61351	Translocon-associated protein β	*Danio rerio*
C00325	AAK27327	Twisted gastrulation protein	*Xenopus laevis*
C00235	NP_509939	14-3-3 protein	*Caenorhabditis elegans*
C01117	XP_041992	Kruppel-like zinc finger protein	*Homo sapiens*
C02073	AAC69756	LIM-domain protein	*Branchiostoma floridae*
C02070	AAF19840	Secreted protein Wnt8	*Branchiostoma belcheri*
C00013	BAB63215	EF-1a [*Branchiostoma floridae*]	*Branchiostoma floridae*
C02092	AAK58840	Homeobox amphivent	*Branchiostoma floridae*
C00002	AAL47678	Myogenic determination factor	*Branchiostoma belcheri*

Zinc finger proteins play important roles in embryonic development. We identified 15 clusters of zinc finger proteins such as zfp-37 and REST protein. Protein zfp-37, a structural protein of the neuronal nucleus, is expressed in neurons of the developing and adult CNS of mouse [30,31]. While REST protein is a zinc finger transcription factor to repress the expression of neuron-specific genes in non-neural tissues and undifferentiated neural precursors during early embryogenesis [32].

ESTs homologous to immune or disease-related genes

Bioinformatic analysis showed that 19 distinct disease-related genes were found in the present study (Table 5), in which one (C00238) being already identified in Chinese amphioxus (AAK84394), four similar to those of Mus musculus, 12 relevant to human disease genes, and two homologous to other organisms. Half of these genes have *E*-values lower than 1×10^{-30}, indicating that though they may not be significantly homologous to disease proteins, the related proteins and portions of the proteins existed in amphioxus and conserved in the evolution history. Most of them were syndrome related genes or antigen. They will be useful to investigate the molecular and biochemical activities of human disease proteins and further elucidate the mechanisms of human disease. A most interesting gene (C00522) is pituitary tumor-transforming gene (PTTG)-binding factor precursor. This gene product binds PTTG, an oncogene, facilitates its nuclear translocation, and potentially regulates its transcriptional activation [33].

ESTs similar to unidentified proteins

Of the sequences similar to unidentified proteins, most of them had shown similarities to hypothetical proteins in human and Mus musculus. These results are the direct evidence that homologues of vertebrate gene families probably existed in the amphioxus genome since amphioxus is thought to be the closest living invertebrate relatives of the vertebrates. Furthermore, the molecular information directed early embryo development is remarkably conserved among animals in terms of evolution.

ESTs identifying no significant matches to known genes

Clusters (66.14%) have no significant matches to any known genes in GenBank. We analyzed partial sequences by checking the putative coding sequences one by one. ORF of a total of 737 clusters (29.56% of the examined ESTs) was determined and analyzed according to the stop codons. The lengths of the nucleotides between start and stop codons were classified into 10 bp bins and the number of clusters was calculated accordingly [17], and the results of the statistical analysis are shown in Fig. 3. The data indicate that the peak distribution of the lengths is between 50 and 250 bp. The results are similar to those of the analysis of 7-h sea urchin embryos [17], in which the possible protein-coding sequences were 50–200 bp in their data. A large fraction lies between 90 and 100 bp, represented by more than 60 ESTs. In the present study the occurrence of 90–100 bp is 56. Both showed that the possible protein-coding sequences of no matches to known genes are

Table 5
Putative disease-related genes

Cluster ID	Accession No.	Description	Organisms	Probability
C00546	NP_005863	Breast carcinoma amplified sequence 2	*Homo sapiens*	5.00E – 83
C01428	NP_113601	Cat eye syndrome chromosome region	*Homo sapiens*	2.00E – 51
C01491	NP_004928	Cystinosis, nephropathic	*Homo sapiens*	1.00E – 25
C00645	NP_004690	XAP-5 protein	*Homo sapiens*	5.00E – 15
C00778	NP_149105	MADP-1 protein	*Homo sapiens*	8.00E – 15
C00868	NP_036347	Meningioma expressed antigen 5	*Homo sapiens*	6.00E – 34
C00281	BAB70508	Prostate cancer antigen-1	*Homo sapiens*	5.00E – 84
C01750	AAG33129	MER receptor tyrosine kinase	*Homo sapiens*	2.00E – 72
C00522	NP_004330	Pituitary tumor-transforming protein 1-interacting protein precursor	*Homo sapiens*	2.00E – 29
C02016	NP_003711	Down syndrome critical region protein 2	*Homo sapiens*	1.00E – 21
C00490	AAD47291	Amyloid precursor protein homolog HSD-2	*Homo sapiens*	5.00E – 18
C01233	NP_000544	Werner syndrome protein	*Homo sapiens*	1.00E – 23
C00257	NP_062369	AMMECR1 protein	*Mus musculus*	1.00E – 67
C00976	NP_062652	Trk-fused gene	*Mus musculus*	5.00E – 55
C00784	BAB27938	Sjogren*s syndrome/scleroderma autoantigen 1	*Mus musculus*	3.00E – 31
C01632	NP_062266	Polymyositis/scleroderma autoantigen 1	*Mus musculus*	1.00E – 49
C00238	AAK84394	Translationally controlled tumor protein	*Branchiostoma belcheri*	4.00E – 94
C00913	AAG61253	Cyfip	*Danio rerio*	2.00E – 45
C01154	Q17103	Myc protein (C-MYC)	*Asterias vulgaris*	2.00E – 23

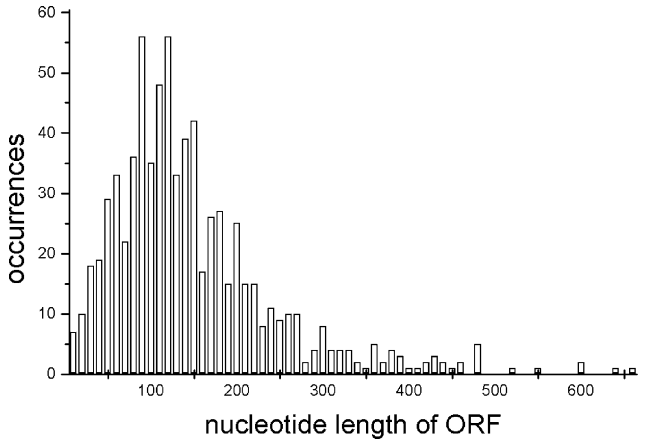

Fig. 3. Statistical analysis of the sequences showing no matches to known genes. We analyzed 786 consensus sequences of the clusters without matches to known genes in GenBank by checking the putative coding sequences one by one. The largest ORF was determined and the lengths of the nucleotides between start and stop codons were calculated and classified into 10 bp bins and the number of clusters was calculated accordingly. Of those, 49 sequences having no stop codons were not shown in this histogram and 737 sequences were statistically estimated. The peak distribution of the lengths is between 50 and 250 bp.

around 100 bp. Analysis indicates that about 68% of the total no-match ESTs are the codogenic sequences of mRNAs, similar to those of early sea urchin embryos (65–80%). We count the distribution of the sequence length mostly ranging from 400 to 750 bp and exclude the possibility that too short sequences lead to no hits. The high abundance of these sequences may correspond to the complexity of mRNA transcripts stored in early embryos, and why is complicated is not very clear by now (see Fig. 4).

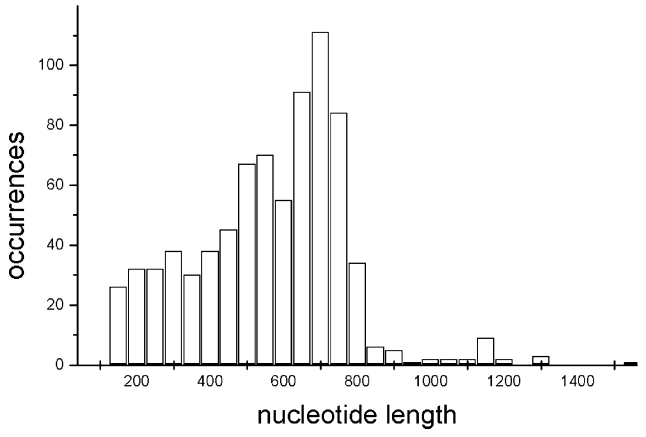

Fig. 4. Length distribution of the sequences of no matches to known genes. A total of 786 consensus sequences showing no matches to known genes in GenBank consisting of 119 clusters and 667 singletons were calculated to obtain the histogram. Abscissa (50 bp) is the readable length of sequences, while the *Y*-coordinate is the occurrences of the ESTs. The statistical estimation showed that the readable length is mostly 400–705 bp excluding the possibility that too short sequences lead to no hits.

Amphioxus is genomically and structurally simpler than vertebrates since it has not undergone the genome duplications that occurred early in vertebrate evolution [34]. For example, earlier studies have suggested that amphioxus contains comparable regions to vertebrate diencephalons and hindbrain and partly midbrain except isthmocerebrellar region [35]. The further genetic studies of amphioxus will provide simpler terms not only on neural development but a global molecular mechanism comparing to the complexity of vertebrate.

In short, a total of 5235 ESTs consisting of 2493 clusters derived from neurula of Chinese amphioxus were examined in the present study. Identification of the genes related to diseases and neural development provides valuable insights into the molecular mechanism of nervous development and possible explanation for diseases. The ESTs homologous to unidentified proteins and sequences showing no similarities to known genes were also discussed. Although the ESTs we generated are derived from a non-normalized cDNA library, and only a part collection are sequenced and analyzed, our data still can provide interesting information for understanding molecular basis of neural development. The study of ESTs in this paper provides a global understanding of the mRNA transcripts that existed in neurula of amphioxus embryos and a resource for further investigation of embryogenomics of amphioxus. Furthermore, this study provides further insights into the evolution of the chordates at gene level. However, further functional analysis of those EST clusters (or genes) with or without matches to known sequences in GenBank will certainly be needed to elucidate the functions of interesting genes.

Acknowledgments

We thank Guijun Zhao, Ning Zhang, Bing Xu, Huihui Kong, and Ting Jia for their help with the data analysis. This research was supported by a grant from State High-Tech Development Project (863) of Ministry of Science and Technology of China (2001AA626010), a key grant of National Natural Science Foundation (69935020), a key grant of Ministry of Education (Key 0107), and key grants of Commission of Science and Technology of Guangdong Province and Guangzhou City.

References

[1] G.C. Schoenwolf, J.L. Smith, Mechanisms of neurulation: traditional viewpoint and recent advances, Development 109 (1990) 243–270.

[2] J.D. Moury, G.C. Schoenwolf, Cooperative model of epithelial shaping and bending during avian neurulation: autonomous movements of the neural plate, autonomous movements of the epidermis, and interactions the neural plate/epidermis transition zone, Dev. Dynam. 204 (1995) 323–337.

[3] N.D. Holland, L.H. Grace Panganiban Erika, L.Z. Holland, Sequence and developmental expression of AmphiDII, an amphioxus Distal-less gene transcribed in the ectoderm, epidermis and nervous system: insights into evolution of craniate forebrain and neural crest, Development 122 (1996) 2911–2920.

[4] Jean-franccois cloas, C.S. Gary, Towards a cellular and molecular understanding of neurulation, Dev. Dyn. 221 (2001) 117–145.
[5] H. Wada, N. Satoh, Details of the evolutionary history from invertebrates to vertebrates, as deduced from the sequences of 18S rDNA, Proc. Natl. Acad. Sci. USA 91 (1994) 1801–1804.
[6] J.M. Turbeville, J.R. Schulz, R.A. Raff, Deuterostome phylogeny and the sister group of the chordates: evidence from molecules and morphology, Mol. Biol. Evol. 11 (1994) 648–655.
[7] B. Hatschek, The Amphioxus and its Development, Swan Sonnenschein, London, 1893.
[8] E.G. Conklin, The embryology of amphioxus, J. Morphol. 54 (1932) 69–151.
[9] O.E. Nelson, Comparative Embryology of the Vertebrates, McGraw-Hill, New York, 1953.
[10] R. Hirakow, N. Kajita, Electron microscopic study of the development of amphioxus, *Branchiostbelcheri belcheri tsing- tauense*: the neurula and larva, Acta Anat. Nippon. 69 (1994) 1–13.
[11] Velculescu, V.E. Lin Zhang, V. Bert, W.K. Kenneth, Serial analysis of gene expression, Science 270 (1995) 484–487.
[12] P. Liang, A.B. Pardee, Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction, Science 257 (1992) 967–971.
[13] T.D. Sargent, I.B. David, Differential gene expression in the gastrula of *Xenopus laevis*, Science 222 (1983) 135–139.
[14] D.J. Lockhart, E.A. Winzeler, Genomics, gene expression and DNA arrays, Nature 405 (2000) 827–836.
[15] T.R. Hughes, D.D. Shoemaker, DNA microarrays for for expression profiling, Curr. Opin. Chem. Biol. 5 (2001) 21–25.
[16] D.A. Benson, I.I. Karsch-Mizrachi, D.J. Ipman, J. Ostell, B.A. Rapp, D.L. Wheeler, Nucleic Acids Res. 28 (2000) 15–18.
[17] Y.-H. Lee, G.M. Huang, R.A. Cameron, G. Graham, E.H. Davidson, L. Hood, R.J. Britten, EST analysis of gene expression on early cleavage-stage sea urchin embryos, Development 126 (1999) 3857–3867.
[18] Y.J. Jeffre, S. Ross Osamu, V. Peter, Annapetryk, K. Hyon, G. Karin, H. Spencer, C.E. Stephen, B.O. Michael, J.L. Marsh, Twisted gastrulation is a conserved extracellular BMP antagonist, Nature 410 (2001) 479–483.
[19] P. Vilmos, K. Gaudenz, Z. Hegedus, J.L. Marsh, The twisted gastrulation family of proteins, together with the IGFBP and CCN families, comprise the TIC superfamily of cysteine rich secreted factors, Mol. Pathol. 54 (5) (2001) 317–323.
[20] R.N. Bamford, E. Roessler, R.D. Burdine, U. Saplakoglu, J. dela Cruz, M. Splitt, J.A. Goodship, J. Towbin, P. Bowers, G.B. Ferrero, B. Marino, A.F. Schier, M.M. Shen, M. Muenke, B. Casey, Loss-of-function mutations in the EGF-CFC gene CFC1 are associated with human left-right laterality defects, Nat. Genet. 26 (2000) 365–369.
[21] J. Zhang, W.S. Talbot, A.F. Schier, Positional cloning identifies zebrafish one-eyed pinhead as a permissive EGF-related ligand required during gastrulation, Cell 92 (1998) 241–251.
[22] J. Ding, L. Yang, Y.T. Yan, A. Chen, N. Desai, A. Wynshaw-Boris, M.M. Shen, Cripto is required for correct orientation of the anterior-posterior axis in the mouse embryo, Nature 395 (1998) 702–707.
[23] J.R. Dyer, W.S. Sossin, M. Klein, Cloning and characterization of aplycalcin and *Aplysia neurocalcin*, two new members of the calmodulin superfamily of small calcium-binding proteins, J. Neurochem. 67 (1996) 932–942.
[24] T. Okuda, T. Haga, Y. Kanai, H. Endou, T. Ishihara, I. Katsura, Identification and characterization of the high-affinity choline transporter, Nat. Neurosci. 3 (2000) 120–125.
[25] S. Apparsundaram, S.M. Ferguson, R.D. Blakely, Molecular cloning and characterization of a murine hemicholinium-3-sensitive choline transporter, Biochem. Soc. Trans. 29 (Pt 6) (2001) 711–716.
[26] K. Vetter, W. Wurst, Expression of a novel mouse gene 'mbFZb' in distinct regions of the developing nervous system and the adult brain, Mech. Dev. 100 (1) (2001) 123–125.
[27] A. Schenck, B. Bardoni, A. Moro, C. Bagni, J.L. Mandel, A highly conserved protein family interacting with the fragile × mental retardation protein (FMRP) and displaying selective interactions with FMRP-related proteins FXR1P and FXR2P, Proc. Natl. Acad. Sci. USA 98 (15) (2001) 8844–8849.
[28] Deleted in proofs.
[29] Deleted in proofs.
[30] N. Mazarakis, D. Michalovich, A. Karis, F. Grosveld, N. Galjart, Zfp-37 is a member of the KRAB zinc finger gene family and is expressed in neurons of the developing and adult CNS, Genomics 33 (1996) 247–257.
[31] E. Payen, T. Verkerk, D. Michalovich, S.D. Dreyer, A. Winterpacht, B. Lee, C.I. De Zeeuw, F. Grosveld, N. Galjart, The centromeric/nucleolar chromatin protein ZFP-37 may function to specify neuronal nuclear domains, J. Biol. Chem. 273 (1998) 9099–9109.
[32] K. Palm, M. Metsis, T. Timmusk, Neuron-specific splicing of zinc finger transcription factor REST/NRSF/XBR is frequent in neuroblastomas and conserved in human, mouse and rat, Brain Res. Mol. Brain Res. 72 (1999) 30–39.
[33] W. Chien, L. Pei, A novel binding factor facilitates nuclear translocation and transcriptional activation function of the pituitary tumor-transforming gene product, J. Biol. Chem. 275 (2000) 19422–19427.
[34] P.W.H. Holland, J. Garcia-Fernandez, N.A. Williams, A. Sidow, Gene duplications and the origins of vertebrate development, Development Suppl. 120 (1994) 125–133.
[35] L.Z. Holland, N.D. Holland, Chordate origins of the vertebrate central nervous system, Curr. Opin. Neurobiol. 9 (1999) 596–602.