Gene expression profile during the life cycle of the urochordate *Ciona intestinalis*

Kaoru Azumi[a,b,*], Sorin V. Sabau[c], Manabu Fujie[b], Takeshi Usami[a], Ryo Koyanagi[a], Takeshi Kawashima[d], Shigeki Fujiwara[e], Michio Ogasawara[f], Masanobu Satake[g], Masaru Nonaka[h], Hong-Gang Wang[i], Yutaka Satou[d] and Nori Satoh[b,d]

[a]Division of Innovative Research, Creative Research Initiative "Sousei", Hokkaido University, N21W10 Kita-ku, Sapporo 001-0021, Japan

[b]CREST, Japan Science and Technology Agency, Saitama 330-0012, Japan

[c]Center for Computer Science and Education, Hokkaido Tokai University, 5-1-1-1 Minami-sawa, Minami-ku, Sapporo 005-8601,

[d]Department of Zoology, Graduate School of Science, Kyoto University, Kitashirakawa Oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan

[e]Department of Materials Science, Faculty of Science, Kochi University, 2-5-1 Akebono-cho, Kochi 780-8520, Japan

[f]Department of Biology, Graduate School of Science, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan

[g]Institute of Development, Aging and Cancer, Tohoku University, Seiryo-cho 4-1, Aoba-ward, Sendai 980-8575, Japan

[h]Department of Biological Sciences, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Tokyo 113-0033, Japan,

[i]Department of Interdisciplinary Oncology, University of South Florida, 12902 Magnolia

Drive, Tampa, Florida 33612, USA

*Corresponding author. Fax: +81 11 706 9205.

E-mail address: azumi@cris.hokudai.ac.jp (K. Azumi)

Abstract

Recent whole-genome studies and in-depth expressed sequence tag (EST) analyses have identified most of the developmentally relevant genes in the urochordate, *Ciona intestinalis*. In this study, we made use of a large-scale oligo-DNA microarray to further investigate and identify genes with specific or correlated expression profiles, and we report global gene expression profiles for about 66% of all the *C. intestinalis* genes that are expressed during its life cycle. We succeeded in categorizing the dataset into 5 large clusters and 49 sub-clusters based on the expression profile of each gene. This revealed the higher order of gene expression profiles during the developmental and ageing stages. Furthermore, a combined analysis of microarray data with the EST database revealed the gene groups that were expressed at a specific stage or in a specific organ of the adult. This study provides insights into the complex structure of ascidian gene expression, identifies co-expressed gene groups and marker genes, and makes predictions for the biological roles of many uncharacterized genes. This large-scale oligo-DNA microarray for *C. intestinalis* should facilitate the understanding of global gene expression and gene networks during the development and ageing of a basal chordate. (192 words)

Keywords: Ascidian • *Ciona intestinalis* • Gene expression • DNA microarray • Clustering • Development • Ageing • Organ-specific gene

Introduction

Ascidians are basal chordates, sharing a common ancestor with vertebrates. Ascidians are effective experimental model organisms, not only holding a unique position in evolution, but also having advantages for the investigation of developmental mechanisms (Corbo et al., 2001; Satoh, 1994, 2001; Satoh et al, 2003). *Ciona intestinalis* is a ubiquitous cosmopolitan species that spawns all year round and is used by researchers worldwide. Recently, a draft genome of *C. intestinalis* has been published (Dehal et al., 2002), and its ~160 Mbp genome is estimated to contain 15,852 protein-coding genes, similar to the number in other invertebrates, but only half of that found in vertebrates. It reveals that the ascidian has a basic, non-duplicated set of a chordate-type genome. In addition, an in-depth cDNA analysis including the generation of expressed sequence tags (ESTs), the collection and sequencing of a non-redundant set of cDNAs, and the clarification of their expression profiles in *C. intestinalis* was carried out (Satoh et al., 2003; Satou et al., 2002a). EST analysis was performed using six cDNA libraries made from different developmental stages, including fertilized eggs, cleaving embryos, gastrulae/neurulae, tailbud embryos, larvae and juveniles. In the analysis libraries made from testis, ovary, heart, endostyle, neural complex and blood cells of mature adults (Kawashima et al., 2005) were used. Finally, over 450,000 ESTs were compiled and assembled into 20,616 independent clusters. All of the *C. intestinalis* cDNA information can be viewed at the website http://ghost.zool.kyoto-u.ac.jp/indexrl.html.

DNA microarray analysis is a very powerful technique for the life sciences,

allowing the measurement of the transcriptional levels of thousands of genes simultaneously. Based on the above-mentioned molecular biological characteristics of *C. intestinalis*, we first established a large-scale cDNA microarray, and then developed this to an oligo-DNA microarray analysis. Previously, cDNA microarray or oligo-DNA microarray studies have been performed in embryos of various organisms, including *Caenorhabditis elegans* (Hill et al., 2000), *Drosophila* (Arbeitman et al., 2002; Furlong et al., 2001; White et al., 1999), zebra fish (Lo et al., 2003; Ton et al., 2002), mice (Carter et al., 2003; Miki et al., 2001; Tanaka et al., 2000), *Xenopus* (Baldessari et al., 2005) and sea urchin (Wei et al., 2006). In particular, a panoramic view of the gene expression profiles throughout the life stages has been depicted in *Drosophila* (Arbeitman et al., 2002) and *Xenopus* (Baldessari et al., 2005). This kind of analysis is important for understanding the differential gene expression mechanism differences between embryos and the adult stages.

In ascidians, global analysis of gene expression profiles in embryogenesis has been carried out using large-scale cDNA microarrays or oligo-DNA microarrays (Azumi et al., 2003a; Ishibashi et al., 2003, 2005; Yamada et al., 2005). In the present study, we applied the DNA microarray analysis to delineate transcriptional profiles for 65.7% (10,415) of the *C. intestinalis* genes (15,852) throughout its life cycle, beginning at fertilization, and containing 13 embryonic stages, the larval and juvenile stages, and 4 adult stages. We revealed basic and characteristic patterns of gene expression in *C. intestinalis*, and co-expressed and functionally related gene clusters. The present analysis provides an expansive view of gene expression in *C. intestinalis*, which allows for

predictions of gene function.

Materials and methods

Ascidian eggs, embryos, juveniles and adults

C. intestinalis were cultivated at the Maizuru Fisheries Research Station of Kyoto University, Maizuru, Kyoto, Japan. Adults were maintained under constant light to induce oocyte maturation. Eggs and sperm were obtained surgically from the gonoduct. After insemination, eggs were washed and maintained in seawater at room temperature (18-20˚C). Larvae hatched after about 17 hours of development. After metamorphosis, juveniles that adhered to trays were cultured with the diatom *Chaetoceros gracillis* as their food source. Two weeks after metamorphosis, juveniles adhered to trays were suspended in Maizuru Bay. These juveniles were grown to adulthood on the trays and survived for up to several months.

Microarray design

For hybridization, we used the custom-made oligo-DNA microarray, *C. intestinalis* Oligoarray ver. 1, which was manufactured by Agilent Technologies (Santa Clara, CA). It is loaded with 21,939 probes consisting of 21,617 independent 60-mer oligonucleotides derived from 22,445 cDNA/EST sequences, selected using the sequence

information from over 450,000 *C. intestinalis* ESTs and 4,062 cDNAs (Satou et al., 2002b). More details of the array are described in ref. (Yamada et al., 2005).

RNA preparation, labeling, hybridization and scanning

To prepare total RNA from *C. intestinalis* eggs, embryos and juveniles, the acid guanidine thiocyanate-phenol/chloroform (AGPC) method (Chomczynski and Sacchi, 1987) was used. For total RNA preparation from *C. intestinalis* adults, tunics were removed and the inner whole bodies of three individuals per stage were homogenized and RNA isolated by ultracentrifugation (Azumi et al., 2005). Poly $(A)^+$RNA was purified using an mRNA Purification Kit (Amersham Pharmacia Biotech, Piscataway, NJ). The RNA quality was verified by electrophoresis using an Agilent 2100 Bioanalyzer (Agilent Technologies). Two micrograms of poly $(A)^+$RNA was labeled with either Cy3 or Cy5 using an Agilent Fluorescent Linear Amplification Kit (Agilent Technologies), mixed, and hybridized with the *C. intestinalis* Oligoarray ver. 1. Hybridization and washing protocols were according to the manufacturer's instructions. The microarrays were scanned with a GenePix 4000B DNA Microarray Scanner (Axon Instruments, Foster City, CA). The resulting fluorescence intensity for each spot was quantified using GenePix Pro4.0 microarray analysis software (Axon Instruments) (Azumi et al., 2003).

Statistical data analysis

Each gene probe was characterized using corresponding cDNA/EST sequences. Annotation information of these sequences was obtained from the non-redundant amino acid (nr-aa) database of NCBI released on December 2005 and June 2006, using the BLAST X algorithm. A filtering analysis of the array dataset was carried out, whereby it was divided into three groups based on the expression level. We also performed a combination analysis of microarray data and EST counts of each clone obtained from the database http://ghost.zool.kyoto-u.ac.jp/index.html. We used cluster analysis in order to reveal the similarities and higher order structure present in the dataset. Hierarchical and k-means clustering were carried out using GeneSpring software (Agilent Technologies). The measure used in both cases was Pearson correlation. The details of data filtering, differential expression analysis and clustering analysis are described in Supplementary methods in Appendix A. The microarray data in this article have been deposited in the National center for Biotechnology Information Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo [accession nos. GSE6308 (series)]

Results

Microarray experiments using C. intestinalis Oligoarray ver. 1

We attempted to obtain global gene expression profiles during the life cycle of *C. intestinalis* using the Oligoarray ver. 1 for this organism. We chose oligo-DNA

microarrays for this analysis rather than cDNA microarrays, by reason of their strict specificity and the feasibility for a larger number of genes. We collected 15 test samples for the developmental stages, comprising fertilized eggs (FE), 2-cell, 4-cell, 8-cell, 16-cell, 32-cell, and 64-cell embryos, and early gastrulae (EG), late gastrulae (LG), early neurulae (EN), embryos at the initial tailbud (ITB), early tailbud (ETB), middle tailbud (MTB), late tailbud (LTB) stages, and early larvae (LV) after hatching. For the adult, we collected samples from 4 different stages, comprising juveniles (JN), 1.5 month-old (1.5M) adults, 2.5 month-old (2.5M) adults and 4.0 month-old (4.0M) adults. Juveniles were collected 2 weeks after metamorphosis. At that time, the juveniles have a small body size but most of the adult organs, except for the ovary and testis (Chiba et al., 2004). Adults of 1.5M are pre-maturation stage; they do not yet have mature gonads. Adults of 2.5M are in the period of reproduction, and have mature gonads, eggs and sperm. Adults of 4.0M still have gonads but have lost their reproductive capacity (Satoh, 1994).

Two different RNA targets for microarray hybridization were used. Namely, for the developmental stages, RNA from fertilized eggs was used as a reference sample and RNA from the 14 stages of embryogenesis was used as test sample (Supplementary Fig. S1a in Appendix A), because it was easy to trace the expression pattern of each gene during development. In the adult stages, we were not able to use egg RNA as a reference, because the expression levels of most of the genes at the adult stage were widely different from those in eggs. We therefore carried out array analysis using adjacent samples, such as larvae versus juveniles, juveniles vs. 1.5M adults, 1.5M vs. 2.5M adults and 2.5M vs. 4.0M adults (Supplementary Fig. S1b in Appendix A). In all of the RNA combinations,

the hybridization was performed twice by reversing dyes. A set of 16,554 of the 21,939 clones (probes) of the *C. intestinalis* oligoarray were rated as producing useful signals. To draw an expression profile for each gene during the life cycle, we computed (see Supplementary data) the ratio of expression of genes in adult stages by defining the expression level of each gene in fertilized eggs as "1" (0 in log scale). Consequently, the expression patterns described in this manuscript are comprised of successively fluctuating patterns of expression levels of each gene, from fertilized eggs to 4.0M adults. It should be noted that the ratio of our array data indicates a relative expression level compared with that in fertilized eggs.

Clustering analysis

The entire dataset of 18 stages was subjected to hierarchical clustering of both genes and stages (Fig. 1). A comparative analysis of gene expression changes over the life stages suggests the division of the gene expression data into some groups of stages that contain genes with similar gene profiles, such as the 2-, 4-, and 8-cell stage (initial embryonic stages) group, the 16-, 32-, and 64-cell stage (early embryonic stages) group, the EG and LG group, the EN, ITB and ETB (middle embryonic stages) group, the MTB, LTB and LV (late embryonic stages) group, and the JN, 1.5M, 2.5M, and 4.0M adult (adult stages) group. These findings indicate that the 16-cell, EG, MTB and JN stages show clear changes in gene expression. Even though the expression patterns of each gene throughout the life cycle may be quite different from each other, two large groups of

genes (the up- and down-regulated gene groups) are easily detectable. On the basis of the results of hierarchical clustering, we performed further detailed analysis using filtering and k-means clustering. Recall that the dataset (16,554 clones) was divided into three groups by filtering based on expression levels. We termed them the up-regulated gene group (8,086 clones), the stably-expressed gene group (1,930 clones), and the down-regulated gene group (6,538 clones), respectively. Using k-means clustering, the up-regulated gene group was subsequently divided into 38 sub-clusters based on the expression pattern and categorized into 3 major clusters: namely, embryonic gene cluster A, embryonic and adult gene cluster B, and adult gene cluster C. We kept the stably-expressed genes in one cluster, namely, the stably-expressed gene cluster D. The down-regulated gene group was divided into 10 sub-clusters, forming the maternal gene cluster E. In this way, the total dataset was divided into 49 sub-clusters and 5 major clusters as described in Fig. 2.

Overview of the gene expression profile during the life cycle

Our data set of 16,554 clones contained 10,415 independent Kyoto grail gene models. We summarized the gene expression profiles for the life cycle of *C. intestinalis* in Fig. 2 and Fig. 3a. The number of genes in the sub-clusters was listed in Supplementary Table S1 in Appendix A. The largest cluster is the maternal gene cluster E containing 38.8% (4,041) of all genes (10,415) detected in our oligoarray. They were expressed in the egg and early embryonic stages, and decreased in the following stages, but returned to

the initial levels during adulthood. The embryonic gene cluster A contains 13.0% (1,357) of the genes, which were expressed at high levels during the embryonic stages but at low levels in adulthood. The embryonic and adult gene cluster B contains 19.9% (2,070) of the genes, highly expressed during the middle of the embryonic stage and maintaining their expression levels during adulthood. The adult gene cluster C contains 15.8% (1,646) of the genes, which were highly expressed in adulthood but expressed at low levels in the egg and embryonic stages. This last cluster contains several unique sub-clusters, such as the organ-specific gene clusters (c6 and c9 sub-clusters, total 165 genes, 1.6%) that contained genes expressed only in the 1.5M and 2.5M adult stages or only in the 2.5M adult stage, and juvenile-specific gene clusters (c3 and c4 sub-clusters, total 312 genes, 3.0%) containing genes expressed only at the juvenile stage. It also includes other sub-clusters containing genes for which the expression level increased according to age, namely the ageing-related gene clusters (c10 and c12 sub-clusters, total 287 genes, 2.8%). The stably-expressed gene cluster D contains 12.5% (1,301) genes for which expression did not significantly change during the life cycle.

Furthermore, we have found that the proportion of non-significantly changed genes (change rate is within ± two-fold) was different between the embryonic stages and adult stages. Non-significantly altered expression of a gene through all the stages (from egg to 4.0M adult) was noted for 12.5% of 10,415 genes, and genes for which the expression level was changed (up- or down-regulated over two-fold) compared with the level at eggs at least once in any stage during that period represented 87.5%. As shown in Fig. 3b, during the embryonic stages, from eggs to larvae, genes without significant

changes in expression represented 27.6% and those with significant changes comprised 72.4%. In contrast, during the adult stages, genes for which the expression level did not significantly change from the juvenile to 4.0M adult stages compared with that at juvenile represented 61.4%, and those with significant changes represented 38.6%. Furthermore, genes for which the expression level was not significantly altered from the 1.5M to 4.0M adult stages compared with that at the 1.5M adult stage comprised 82.6%, and those with significant changes represented 17.4%.

Next, in order to find the characteristics of each cluster, we searched for functionally related gene categories in the 49 clusters from three points of view (Fig. 4). Namely, (i) using the annotation information of each gene, we searched for genes that encoded proteins conserved in humans, human hypothetical (function unknown) proteins, and those bearing no similarity to known proteins. We found 6,443 genes having high similarity to human genes. The distribution pattern of human conserved genes in clusters A-E was similar to that of all genes. However, among them, genes encoding functionally unknown proteins such as hypothetical proteins or those lacking a known protein domain (900 genes) were found mainly (48% of the functionally unknown genes) in cluster E. We found 2,951 genes for which the predicted products were assigned as having “no similarity” to any known protein; these can be anticipated to be ascidian-specific proteins. These were at relatively high proportions in the B (24% of the no similarity genes) and C (19% of the no similarity genes) clusters. (ii) Using EST counting information of embryonic stages and adult organs for each gene, we searched for adult organ-specific genes and juvenile-specific genes (details are given in the following section). In the case

of organ- or juvenile-specific genes, their distribution was markedly disproportionate to that of total genes; they belonged mainly to clusters C (around 49% for each), and B (23% and 27%, respectively). (iii) In order to find functionally related genes, we focused on gene families related to embryogenesis and morphogenesis, or to immunity, since these genes had already been searched for in the *C. intestinalis* genome and details were reported in ref. (Chiba et al., 2003; Kawashima et al., 2003; Hino et al., 2003; Sasakura et al., 2003a, b; Satou et al., 2003a, b; Wada et al., 2003; Yagi et al., 2003; Yamada et al., 2003a) and ref. (Azumi et al., 2003b), respectively. These genes are listed in Supplementary Tables S2 and S3 in Appendix A. We found that embryogenesis- and morphogenesis-related genes (377 genes) belonged mainly to the B (37%) and also E (31%) clusters, while immunity-related genes (133 genes) were mainly found in the C (26%) as well as B (24%) clusters. These results were reasonable from the aspect of the biological relevance of these genes. Gene lists of the 49 sub-clusters including gene ID, annotation information and organ specificity are provided in Supplementary Tables S4 to S8 in Appendix A. We further describe details of the characteristics of each cluster below.

The embryonic gene cluster

The embryonic gene cluster A contains 10 sub-clusters (a1-a10) for which the gene transcription levels were high during embryogenesis, and low during the juvenile phase and adulthood (Fig. 2A). The sub-cluster a1 showed dramatically increased gene expression in 8- or 16-cell embryos and the expression level was maintained at the middle

stages of embryogenesis, before decreasing at the late stages and returning to the original level in the juvenile and adult stages. The expression level of genes belonging to a2 increased at the 32-cell stage and decreased at later stages, while that of a3 increased at EG, a4 at EN, a5 at ETB, and a6 at MTB. The expression level of genes in each sub-cluster also decreased at later embryonic stages to become low in the juvenile stages. The expression level of other sub-clusters increased more generally and up-regulation levels were not high. The A cluster contains basic embryogenesis-related genes such as transcription factors; *Hox* (in the a2 sub-cluster), *Fox* (a2), *Pim* (a2), *MyoD* (a3), and notochord-related genes such as *ZicL* (a1), *Brachyury* (a2) and *Brachyury*-downstream genes (a3 and a4). The a5 and a6 sub-clusters seem to be 'tailbud embryo-specific' gene clusters, because these genes were expressed only at the tailbud stage. They contain genes related to cell junction and adhesion, such as *claudin*, *cadherin*, *filamin*, *Notch* and *Selectin P*. The a5 sub-cluster also contains 10 genes of the transporter or solute carrier family.

The embryonic and adult gene cluster

The embryonic and adult gene cluster B was the second-largest cluster and contains 16 sub-clusters (b1-b16) (Fig. 2B). There were two main patterns of gene expression in this cluster. The first pattern was of a high expression level from the middle or late embryonic stage until the juvenile stage and then a decrease at the adult stage, such as in the b1 to b7 sub-clusters. In the second pattern, expression started at the middle or

late embryonic stage and high expression was maintained during all the adult stages, such as in the b8 to b14 sub-clusters. The B cluster contains many genes related to embryogenesis and morphogenesis, especially expressed in muscle cells and epithelial cells, such as myosins, actins, collagens, and protein biosynthesis genes such as those of ribosomal proteins. Sub-clusters b1, b2 and b10 contain a relatively large number of embryogenesis and morphogenesis (EM)-related genes, so-called 'EM-related gene clusters'; b1 contains genes of the receptor tyrosine kinase (RTK) family, Wnt family and cell junction and adhesion family, b2 contains many muscle-related genes, and b10 contains some Hox family genes and several collagen genes. There are some other characteristic sub-clusters; b4 has several cell junction and adhesion-related genes, b5 contains transforming growth factor β (TGFβ) family genes, b7 contains some synapse-related genes, and b8 and b14 are protein biosynthesis gene clusters, since they contain many genes encoding ribosomal proteins. Furthermore, the B cluster is secondarily a large cluster for both organ-specific (containing 23% of the organ-specific genes) and juvenile-specific genes (29%); b8 and b14 particularly contain a relatively high percentage of organ-specific and juvenile-specific genes, respectively.

The adult gene cluster

The adult gene cluster C contains 12 sub-clusters (c1-c12) for which the expression level is low at the egg and embryonic stages, and high in the adult (Fig. 2C). This cluster contains unique characteristic sub-clusters that show high expression peaks

in only one or two specific stages, such as a peak at 1.5M through 2.5M (c6, c7 and c8), 2.5M only (c9), or mainly at the juvenile stage (c3 and c4). Furthermore, there were two sub-clusters, c10 and c12, for which the expression level increased only at the latest stage (4.0M) in the life cycle of this animal, so these genes may be functionally related to the ageing process. Many of the immunity-related genes belong to cluster C; in particular, several complement component genes were found in the c1 and c2 sub-clusters. The c3 and c4 sub-clusters were juvenile-specific, since a sharp expression peak was detected at the juvenile stage in these sub-clusters. The c6, c7 and c8 sub-clusters contain genes that were expressed mainly in the 1.5M-2.5M adults, and they contain a high percentage of organ-specific genes. Among them, c6 is a unique sub-cluster; 70% of the genes in c6 were organ-specific genes and most of them were genes with no similarity to known genes. The c9 is a small cluster, however, expression of its genes was detected solely at the 2.5M adult stage, and more than 60% of the genes in c9 were organ-specific genes.

Stably-expressed gene cluster

Genes of the stably-expressed gene cluster D are ubiquitous and abundantly expressed at all the stages, with an expression level that was stable from fertilized eggs to 4.0M adults (Fig. 2D). This cluster contains some PCP (planar cell polarity) pathway genes, WASP pathway genes, 28 genes that are related to protein biosynthesis such as ribosomal protein, translation initiation factor and transcription elongation factor, and 11 ubiquitin-related genes.

Maternal gene cluster

The maternal gene cluster E is the largest cluster, containing 10 sub-clusters (e1-e10) (Fig. 2E). These sub-clusters are characterized by genes that were expressed maternally in eggs and during early embryonic development. It should be noted that the gene expression gradually decreased during later embryonic stages. In many of the sub-clusters, the expression level of genes returned to the original level at the 1.5M or 2.5M adult stage, except for e6, where the gonad contained mature eggs through the 1.5M to 2.5M adult stages (oogenesis). These expression patterns correspond well with the tissue organization of *C. intestinalis*. Many of the genes of this cluster encode proteins, such as human conserved and hypothetical proteins, and proteins functioning in RNA processing, mitochondria, the cell cycle, chromatin organization, the respiratory chain and protein biosynthesis. Of the 10 sub-clusters of cluster E, many of them share a functional relationship to another sub-cluster, so that the characteristic aspects of each sub-cluster are not entirely distinct. For example, the mitochondrial gene group is found in the e1, e2, e7 and e10 sub-clusters, the respiratory chain-related genes are in e2 and e3, and protein biosynthesis-related genes are in e2 and e7. Furthermore, the E cluster secondarily contains a large number of embryogenesis- and morphogenesis-related genes, with, in particular, PI3K (phosphoinositide 3-kinase) pathway genes and PCP pathway genes being detected in e1 and e5, respectively.

Distribution of organ-specific genes

Array data regarding *C. intestinalis* adult organs has not yet been obtained, however, in-depth EST analysis suggested organ specificity of gene expression. We have obtained EST statistical data for 6 organs (ovary, testis, heart, neural complex, blood cells and endostyle) from the *C. intestinalis* EST database, and searched for genes for which more than 80% of those EST clones were found in one or more adult organ library. We defined 1,021 of 10,415 genes as “organ-specific genes”. In the same way, we defined 370 of 10,415 genes as “juvenile-specific genes” for which more than 80% of those EST clones were found in only a juvenile library. Next, we searched for the cluster that those genes belonged to. The largest proportions of organ-specific genes (42% of 1,021 genes) and juvenile-specific genes (43% of 370 genes) were found in the adult-expressed gene cluster (C); c6 and c4, in particular, contained the largest number of organ-specific and juvenile-specific genes, respectively (See Supplementary Table S6 in Appendix A). Furthermore, the organ-specific genes were classified into two major groups. One group of genes was predominantly expressed in a single organ. The other group of genes was expressed in two or more organs but not in embryos and juveniles. The one-organ-specific genes were also primarily found in cluster C (Fig. 5), however, 38% of the ovary-specific genes were additionally detected in cluster E. Among the sub-clusters of cluster C, testis-specific genes were found in c6 (66% of the genes in c6) and c7 (15%), ovary-specific genes were in c2 (33%), neural complex-specific genes were detected in c7 (28%) and c2 (19%), heart-specific genes were in c1 (28%), and blood cell-specific

genes were in c1 (21%) and c2 (17%). In other words, c6 is a testis-specific gene cluster, and c7 contains both neural complex-specific and testis-specific genes. The gene expression pattern of c6 and c7 was at a high level only at the 1.5M and 2.5M adult stages. The c2 sub-cluster contains ovary-specific, neural complex-specific and blood cell-specific genes, and c1 contains heart-specific and blood cell-specific genes. The c1 and c2 sub-clusters contain genes for which expression increases at the juvenile stage and the expression level is maintained during the adult stage. Juvenile-specific genes were detected in c3 and c4, which contain genes expressed at a high level only at this stage.

Reliability of the microarray data

With the aim of confirming the reliability of our array data, we compared the array data with previously published results, such as EST counts or in situ hybridization data. We calculated the ratio (percentage) of the numbers of EST clones detected in the embryonic stages (eggs, cleaving embryos, gastrulae, tailbud embryos and larvae) and adult stages (juveniles and adult tissues) for each embryogenesis- and morphogenesis-related gene in Table S2, and each immunity-related gene in Table S3, respectively. Next, we calculated the average percentages of the EST counts in both the embryonic and adult stages in each of clusters A, B, C, D and E (Table S9). As shown in Table S9, in the case of immunity-related genes, the EST count data showed good correlations with our array data. For example, cluster A contained embryo-specific genes and their average EST counts were significantly larger in embryonic stages than in adult

stages. Furthermore, cluster C contained mainly adult-specific genes, and their average EST counts were significantly larger in adult stages than in embryonic stages. Clusters B, D and E contained embryonic and adult genes, stably-expressed genes and maternal genes, respectively, and their EST counts were therefore detected in both the embryonic and adult stages. In the case of embryogenesis- and morphogenesis-related genes, the EST count data were mostly correlated with the array data, although the rate of correlation was lower than that for immunity-related genes. Next, we compared the array data with in situ hybridization data. Table S10 shows the previously published in situ hybridization data for 51 embryogenesis- and morphogenesis-related genes. The in situ hybridization data revealed local information of gene expression rather than the level of gene expression. However, in Table S10, genes belonging to cluster A were detected in relatively early stages of embryogenesis, while many genes belonging to cluster B were detected in late stages of embryogenesis. These findings were well correlated with our array data. To further assess the biological significance of our array data, we describe the notochord-related genes and complement genes in the following paragraphs.

Gene cascade of notochord-related genes

In ascidians, *ZicL*, *Brachyury* and *Brachyury*-downstream genes are essential for larval notochord formation. It has been established using antisense morpholino and *in situ* hybridization techniques that their gene expression is regulated in the fashion of a gene cascade, such as *ZicL* activates *Brachyury*, and subsequently, *Brachyury* activates

several downstream genes (Hotta et al., 2000; Yagi et al., 2004). The expression patterns of *ZicL*, *Brachyury* and downstream genes in our dataset are shown in Fig. 6a. The expression of *ZicL* started at the 32-cell stage, that of *Brachyury* at the 64-cell stage, and expression of downstream genes commenced at 64-cell stage or the later stages. The array data corresponded well to the *in situ* data previously reported (Table S11). In addition, *ZicL*, *Brachyury* and its downstream genes were also co-expressed with other genes in respective sub-clusters. For example, *ZicL* was found in a1, *Brachyury* in a2, *fibrin* and *pk2* in a3, *Not1* and *Not4* in a4, *ASAK* in a5, *netrin* in b1, and *Not2* in b12.

Gene expression profile of immunity-related genes

It has been generally believed that in both vertebrates and invertebrates, immunity-related genes are expressed mainly in adult organs, such as the liver and circulatory system, to produce protective proteins against exotic microorganisms and viruses. Our microarray data show that the gene expression pattern of 133 immunity-related genes was categorized mainly into three clusters, B (24% of the immunity-related genes), C (26%) and E (27%) clusters. The genes belonging to cluster C consist of many complement component genes such as *complement C3*, *factor B*, *mannose-binding protein (MBP)*, and *mannan-binding lectin-associated serine protease (MASP)*, and a Toll-like receptor gene. The genes belonging to cluster E consisted of not only immune-response genes, but also transcription factors such as *interferon regulatory factor 4* (*IRF4*), *Myb* and *Oct*, and genes concerned with signal

transduction such as *mitogen-activated protein kinase* (*MAPK*), *ZAP-70*, *Syk* and *tumour necrosis factor receptor* (*TNFR*). In the case of genes belonging to cluster B, we found some transcription factors related to embryogenesis and morphogenesis, such as *Sox*, *Pax*, *EBF* and adhesion molecules. Furthermore, we found some possible complement component genes in cluster B. Fig. 6b shows the expression patterns of possible complement component genes found in clusters B and C. *C3*, *MASP*, *factor B* (*Bf*), *MBP* and two genes of complement C6 were expressed only at the adult stages. On the other hand, *C1q-like* and two other genes of complement C6 were expressed in the middle of the embryonic stages, and maintained their expression level during the adult stages. Similar results were found for other immunity-related genes; *calmodulin*, *leucine-rich repeat-containing receptor* and *TNFR-associated factor*. These findings strongly suggested that these genes have different functions during the life cycle in *C. intestinalis*.

Discussion

In this report, we presented the transcriptional profiles for about 66% of the whole predicted genes throughout the *C. intestinalis* life cycle, from fertilized eggs to ageing adults. Based on the expression patterns, using a combination of filtering method and k-means algorithm, we succeeded in categorizing the genes into 5 large clusters and 49 sub-clusters. The expression pattern of most of the *C. intestinalis* genes changed broadly over the life cycle. This was clearly different from the gene expression patterns in the fruit fly *Drosophila*, since it has been reported that genes of *Drosophila* are expressed

in two waves during its life cycle, with embryonic expression patterns recapitulated in pupae and larval patterns recapitulated in adults (Arbeitman et al., 2002). In contrast, in *C. intestinalis*, detected expression peaks were singular in most genes except for those in the b15 sub-cluster. However, there were appreciable gene expression changes at the 16-cell, EG, ITB and juvenile stages (see Fig. 2, the diagram of hierarchical clustering). Since the juvenile stage in the ascidian is known to be the stage in which adult organogenesis advances, it is reasonable to assume that the expression pattern at the juvenile stage is more related to that at the adult stages than at the embryonic stages.

We have in this study presented an overview of the gene expression patterns in *C. intestinalis*. The characteristics of gene expression in this animal were as follows: (i) the maternal gene cluster is the largest cluster, containing 38.8% of the genes. Similar results have been recorded in *Drosophila* and *Xenopus*, where the proportions of maternal genes were found to be around 30% in each species (Arbeitman et al., 2002; Baldessari et al., 2005). (ii) At the embryonic stages, expression of genes fluctuates rapidly, and about 72% of whole genes were up- or down-regulated, while only 28% were not significantly changed. In contrast, at the adult stages, the proportion of non-significantly changed gene expression was much greater. For example, expression of 61% of genes was not significantly changed from the juvenile to 4.0M adult stage, and expression of 83% was not significantly changed from the 1.5M to 4.0M adult stage. These findings suggest that the gene expression changed quickly in accordance with the intrinsic program of embryogenesis, while in the adult, expression of most of the genes was stable and did not significantly change during the process of normal ageing. However, gene expression in

the adult will respond to physiological and biological stimulations, such as physiological stress, exposure to pollutant chemicals, and bacterial or viral infections. In fact, we have found that the expression of some of these “non-significantly changed genes” was dramatically increased when ascidians were exposed to a pollutant chemical compound at the adult stage (Azumi et al., 2004).

In order to confirm the reliability and assess the biological significance of our array data, we compared the array data with previously published EST counts or in situ hybridization data. The in situ hybridization data mainly showed local information of the expressed genes, while the EST counts and microarray data indicated the mRNA level of each gene. Within an EST count dataset containing a total of 230,000 EST clones, we found that 57% of our 10,415 genes had less than 10 EST counts and 36% of our 10,415 genes had less than 5 EST counts during all stages and tissues. Despite the limited information of the EST counts and in situ hybridization data, our array data results showed apparently corresponding trends of the gene expression patterns with these data. Furthermore, we have extracted the expression patterns of *ZicL*, *Brachyury* and *Brachyury*-downstream genes from the array data. The expression patterns of these genes corresponded well to the previously obtained *in situ* data (Hotta et al., 2000; Yagi et al., 2004). Furthermore, we found that many of the possible complement component genes belonged to cluster C and that they were expressed mainly in the adult. It is reported that the complement system is very important in ascidian immunity as the major opsonic system in the hemolymph. At least two complement components, MASP and complement C3, are shown to be involved in opsonization, enhancing phagocytosis by

hemocytes through the integrin-type complement receptors on their surface (Nonaka and Yoshizaki, 2004). The expression pattern of the complement genes that we observed corresponds well to the localization of their proteins. These results strongly suggest that our microarray data has biological significance.

Using annotation information, we examined the expression profiles of human-related genes and ascidian-specific genes in our dataset. We found that 6,443 genes of *C. intestinalis* were similar to human genes, and they include 900 genes of human hypothetical proteins for which functions are unknown. Those genes that have counterparts in the human genome were mainly categorized into the maternal gene cluster E. This result suggests that some ascidian genes, which are expressed maternally and function during embryogenesis, are related to functional human genes. This finding is reasonable, because the molecular mechanisms of embryogenesis in ascidians have been conserved in vertebrates including humans. In contrast, we found 2,951 genes that were assigned as “not similar” to known proteins, i.e., ascidian-specific genes. These genes were relatively high in number in clusters B and C. This indicates that the ascidian-specific genes contribute to the morphogenesis and functions of adult organs in ascidians.

The next assessment was whether genes with related biochemical functions were coordinately expressed at the embryonic and adult stages. We found that genes within a functional group tend to be expressed at similar stages. For example, most cell-cycle genes are expressed at relatively high levels in the egg (maternal) and at the embryonic stages when cell division is rapid. In contrast, most protein biosynthesis genes

are expressed at their highest levels at the middle embryonic stage and during the larval and adult phases. The muscle group genes encoding skeletal muscle-specific proteins such as myosins, tropomyosins, skeletal troponins and actins are highly expressed at the embryonic and adult stages, and at low levels in eggs and early embryos. Interestingly, some of the immunity-related genes are expressed not only in adults but also during embryogenesis. As described above, the 5 large clusters obviously contained functionally related genes, and the additional 49 sub-clusters further characterized the gene details. We found embryogenesis- and morphogenesis-related gene clusters (B2 and b10), juvenile-specific gene clusters (C3 and c4), organ-specific gene clusters (c6 and c9), ageing-related gene clusters (c10 and c12) and so on. These sub-clusters contain many genes of unknown function and further characterization of these sub-clusters will be useful for estimating the functions of these genes. Furthermore, our preliminary analysis using a *Ciona* oligo DNA microarray revealed that the genes up- or down-regulated after exposure to chemical pollutants belonged to specific or limited sub-clusters. Our unique classification method for *C. intestinalis* genes will be of benefit for further studies.

Conclusions

In this study, we have collected more than 270,000 articles of gene expression data, and consequently we have obtained in-depth gene expression profiles during the life cycle of *C. intestinalis*. The 5 large clusters and 49 sub-clusters we obtained will greatly benefit the characterization of genes identified in this animal. The lists of genes within

sub-clusters having potential stage- and organ-specific expression will facilitate the identification of novel markers and novel functions of uncharacterized genes, both those conserved with humans and those specific to ascidians. We conclude that our studies using *C. intestinalis* large-scale oligo-DNA microarrays, in addition to providing functional annotation of the *C. intestinalis* genome, should prove to be very useful toward a complete description of the genetic networks that control embryogenesis and ageing in both invertebrates and vertebrates (Imai et al., 2006). In future studies, it will be necessary to undertake systematic loss-of-function analyses of genes suggested in this study to have novel functions. This can now be readily tested in experiments using injection of antisense morpholino oligos (Yamada et al., 2003b) and large scale use of whole-mount *in situ* hybridization (WISH) (Ogasawara et al., 2006). Although we used EST data for identifying organ-specific genes, we are preparing further array analysis of adult tissues to confirm both the genuine single organ-specific genes and the genes that are expressed in multiple organs of *C. intestinalis*. We are also constructing the array database so as to be publicly available in the near future.

Acknowledgments

This work was a part of a CREST project “Development, Differentiation and Regeneration”, CREST, JST, Japan (N.S. and K.A.). We wish to thank Kenichiro Kanda, Yasufumi Miki, Kazuko Hirayama and Chikako Imaizumi for their technical support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version.

References

Arbeitman, M.N., Furlong, E.E.M., Imam, F., Johnson, E., Null, B.H., Baker, B.S., Krasnow, M.A., Scott, M.P., Davis, R.W., White, K.P., 2002. Gene expression during the life cycle of *Drosophila melanogaster*. Science 297, 2270-2275.

Azumi, K., Takahashi, H., Miki, Y., Fujie, M., Usami, T., Ishikawa, H., Kitayama, A., Satou, Y., Ueno, N. Satoh, N., 2003a. Construction of a cDNA microarray derived from the ascidian *Ciona intestinalis*. Zool. Sci. 20, 1223-1229.

Azumi, K., De Santis, R., De Tomaso, A., Rigoutsos, I., Yoshizaki. F., Pinto, M.R., Marino, R., Shida, K., Ikeda, M., Ikeda, M., et al., 2003b. Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot". Immunogenetics 55, 570-581.

Azumi, K., Fujie, M., Usami, T., Miki, Y., Satoh, N., 2004. A cDNA microarray technique applied for analysis of global gene expression profiles in tributyltin-exposed ascidians. Mar. Env. Res. 58, 543-546.

Azumi, K., Sasaki, T., Okochi, K., Yamasaki, S., Saito, T., Takayama, H., Yokosawa, H., 2005. Differential display analysis reveals the expression of glutathione S-transferase

omega and novel genes through an ITAM-containing receptor in ascidian immunocytes. Immunogenetics 57, 444-452.

Baldessari, D., Shin, Y., Krebs O., Konig, R., Koide, T., Vinayagam, A., Fenger, U., Mochii M., Terasaka, C., Kitayama, A., et al., 2005. Global gene expression profiling and cluster analysis in *Xenopus laevis*. Mech. Dev. 122, 441-475.

Carter, M.G., Hamatani, T., Sharov, A.A., Carmack, C.E., Qian, Y., Aiba, K., Ko, N.T., Dudekula, D.B., Brzoska, P.M., Hwang, S.S. et al., 2003. In situ-synthesized novel microarray optimized for mouse stem cell and early developmental expression profiling. Genome Res. 13, 1011-1021.

Chiba, S., Awazu, S., Itoh, M., Chin-Bow, S. T., Satoh, N., Satou, Y., Hastings, K.E.M., 2003. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. IX. Genes for muscle structural proteins. Dev. Genes Evol. 213, 291-302.

Chiba, S., Sasaki, A., Nakayama, A., Takamura, K., Satoh, N., 2004. Development of *Ciona intestinalis* juveniles (Through 2nd ascidian stage). Zool. Sci. 21, 285-298.

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

Corbo, J.C., Di Gregorio, A., Levine, M., 2001. The ascidian as a model organism in developmental and evolutionary biology. Cell 106, 535-538.

Dehal, P., Satou, Y., Campbell, R.K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D. et al., 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 298,

2157-2167.

Furlong, E.E.M., Andersen, E.C., Null, B., White, K.P., Scott, M.P., 2001. Patterns of gene expression during *Drosophila* mesoderm development. Science 293, 1629-1633.

Hill, A.A., Hunter, C.P., Tsung, B.T., Tucker-Kellogg, G., Brown, E.L., 2000. Genomic analysis of gene expression in *C. elegans*. Science 290, 809-812.

Hino, K., Satou, Y., Yagi, K., Satoh, N., 2003. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. VI. Genes for Wnt, TGFbeta, Hedgehog and JAK/STAT signaling pathways. Dev. Genes Evol. 213, 264-272.

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

Imai, K.S., Levine, M., Satoh, N., Satou, Y., 2006. Regulatory blueprint for a chordate embryo. Science 312, 1183-1187.

Ishibashi, T., Nakazawa, M., Ono, H., Satoh, N., Gojobori, T., Fujiwara, S., 2003. Microarray analysis of embryonic retinoic acid target genes in the ascidian *Ciona intestinalis*. Dev. Growth Differ. 45, 249-259.

Ishibashi, T., Usami, T., Fujie, M., Azumi, K., Satoh, N., Fujiwara, S., 2005. Oligonucleotide-based microarray analysis of retinoic acid target genes in the protochordate, *Ciona intestinalis*. Dev. Dyn. 233, 1571-1578.

Kawashima, T., Tokuoka, M., Awazu, S., Satoh, N., Satou, Y., 2003. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. VIII. Genes for PI3K

signaling and cell cycle. Dev. Genes Evol. 213, 284-290.

Kawashima, T., Satou, Y., Murakami, S.D., Satoh, N., 2005. Dynamic changes in developmental gene expression in the basal chordate *Ciona intestinalis*. Dev. Growth Differ. 47, 187-199.

Lo, J., Lee, S., Xu, M., Liu, F., Ruan, H., Eun, A., He, Y., Ma, W., Wang, W., Wen, Z. et al., 2003. 15000 unique zebrafish EST clusters and their future use in microarray for profiling gene expression patterns during embryogenesis. Genome Res. 13, 455-466.

Miki, R., Kadota, K., Bono, H., Mizuno, Y., Tomaru, Y., Carninci, P., Itoh, M., Shibata, K., Kawai, J., Konno, H., et al., 2001. Delineating developmental and metabolic pathways in vivo by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays. Proc. Natl. Acad. Sci. U. S. A. 98, 2199-2204.

Nonaka, M., Yoshizaki, F., 2004. Primitive complement system of invertebrates. Immunol. Rev. 198, 203-215.

Ogasawara, M., Satoh, N., Shimada, Y., Wang, Z., Tanaka, T., Noji, S., 2006. Rapid and stable buffer exchange system using InSitu Chip suitable for multicolor and large-scale whole-mount analyses. Dev. Genes Evol. 216, 100-104.

Sasakura, Y., Yamada, L., Takatori, N. Satou, Y., Satoh, N., 2003a. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. VII. Molecules involved in the regulation of cell polarity and actin dynamics. Dev. Genes Evol. 213, 273-283.

Sasakura, Y., Shoguchi, E., Takatori, N., Wada, S., Meinertzhagen, I.A., Satou, Y., Satoh, N., 2003b. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. X. Genes for cell junctions and extracellular matrix. Dev. Genes Evol.

213, 303-313.

Satoh, N., 1994. In: Developmental biology of ascidians. Cambridge Univ Press, NY.

Satoh, N., 2001. Ascidian embryos as a model system to analyze expression and function of developmental genes. Differentiation 68, 1-12.

Satoh, N., Satou, Y., Davidson, B., Levine, M., 2003. *Ciona intestinalis*: an emerging model for whole-genome analyses. Trends Genet. 19, 376-381.

Satou, Y., Takatori, N., Fujiwara, S., Nishikata, T., Saiga, H., Kusakabe, T., Shin-i, T., Kohara, Y., Satoh, N., 2002a. *Ciona intestinalis* cDNA projects: expressed sequence tag analyses and gene expression profiles during embryogenesis. Gene 287, 83-96.

Satou, Y., Yamada, L., Mochizuki, Y., Takatori, N., Kawashima, T., Sasaki, A., Hamaguchi, M., Awazu, S., Yagi, K., Sasakura, Y. et al., 2002b. A cDNA resource from the basal chordate *Ciona intestinalis*. Genesis 33, 153-154.

Satou, Y., Imai, K. S., Levine, M., Kohara, Y., Rokhsar, D., Satoh, N., 2003a. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. I. Genes for bHLH transcription factors. Dev. Genes Evol. 213, 213-221.

Satou, Y., Sasakura, Y., Yamada, L., Imai, K. S., Satoh, N., Degnan, B., 2003b. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. V. Genes for receptor tyrosine kinase pathway and Notch signaling pathway. Dev. Genes Evol. 213, 254-263.

Tanaka, T.S., Jaradat, S.A., Lim, M.K., Kargul, G.J., Wang, X., Grahovac, M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R. et al., 2000. Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA

microarray. Proc. Natl. Acad. Sci. U. S. A. 97, 9127-9132.

Ton, C., Stamatiou, D, Dzau, V.J., Liew, C-C., 2002. Construction of a zebrafish cDNA microarray: gene expression profiling of the zebrafish during development. Biochem. Biophys. Res. Commun. 296, 1134-1142.

Wada, S., Tokuoka, M., Shoguchi, E., Kobayashi, K., Di Gregorio, A., Spagnuolo, A., Branno, M., Kohara, Y., Rokhsar, D., Levine, M. et al., 2003. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. II. Genes for homeobox transcription factors. Dev. Genes Evol. 213, 222-234.

Wei, Z., Angerer, R.C., Angerer, L.M., 2006. A database of mRNA expression patterns for the sea urchin embryo. Dev. Biol. 300, 476-484.

White, K.P., Rifkin, S.A., Hurban, P., Hogness, D.S., 1999. Microarray analysis of *Drosophila* development during metamorphosis. Science 286, 2179-2184.

Yagi, K., Satou, Y., Mazet, F., Shimeld, S. M., Degnan, B., Rokhsar, D., Levine, M., Kohara, Y., Satoh, N., 2003. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. III. Genes for Fox, ETS, nuclear receptors and NFkappaB. Dev. Genes Evol. 213, 235-244.

Yagi, K., Satou, Y., Satoh, N., 2004. A zinc finger transcription factor, ZicL, is a direct activator of Brachyury in the notochord specification of *Ciona intestinalis*. Development 131, 1279-1288.

Yamada, L., Kobayashi, K., Degnan, B., Satoh, N., Satou, Y., 2003a. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. IV. Genes for HMG transcriptional regulators, bZip and GATA/Gli/Zic/Snail. Dev. Genes Evol. 213,

245-253.

Yamada, L., Shoguchi, E., Wada, S., Kobayashi, K., Mochizuki, Y., Satou, Y., Satoh, N., 2003b. Morpholino-based gene knockdown screen of novel genes with developmental function in *Ciona intestinalis*. Development 130, 6485-6495.

Yamada, L., Kobayashi, K., Satou, Y., Satoh, N., 2005. Microarray analysis of localization of maternal transcripts in eggs and early embryos of the ascidian, *Ciona intestinalis*. Dev. Biol. 284, 536-550.

Figure legends

Fig. 1. Hierarchical cluster analysis of the oligoarray dataset of *Ciona intestinalis*. Hierarchical cluster dendrogram of the whole gene expression dataset was performed using the array analysis software GeneSpring. Gene expression levels are shown as the logarithmic ratio (base 2); "Ratio 0" indicates an expression level equivalent to that in fertilized eggs. Each expression level was color-coded as shown in the color gradient on the right.

Fig. 2. Gene expression patterns of *Ciona intestinalis*. The dataset was split into 5 large clusters (A-E) and 49 sub-clusters (a1-10, b1-16, c1-12, d, and e1-10) using a combination of filtering method and k-means cluster algorithm. The number of independent genes in each cluster is indicated. Expression ratios are shown as a logarithmic ratio (base 2); "ratio 0" indicates the expression level in fertilized eggs as a reference of each sample.

Fig. 3. Overview of gene expression in the life cycle of *Ciona intestinalis*. (a) A pattern diagram of the gene groups A to E corresponding to the five large clusters described in Fig. 3, and to the three specific sub-clusters of cluster C. (b) Percentage of non-significantly changed genes (not having significantly differential expression) from a total of 10,415 genes. "Non-significantly changed genes" are those genes having during the indicated period, an expression level changed within ± two-fold compared with the

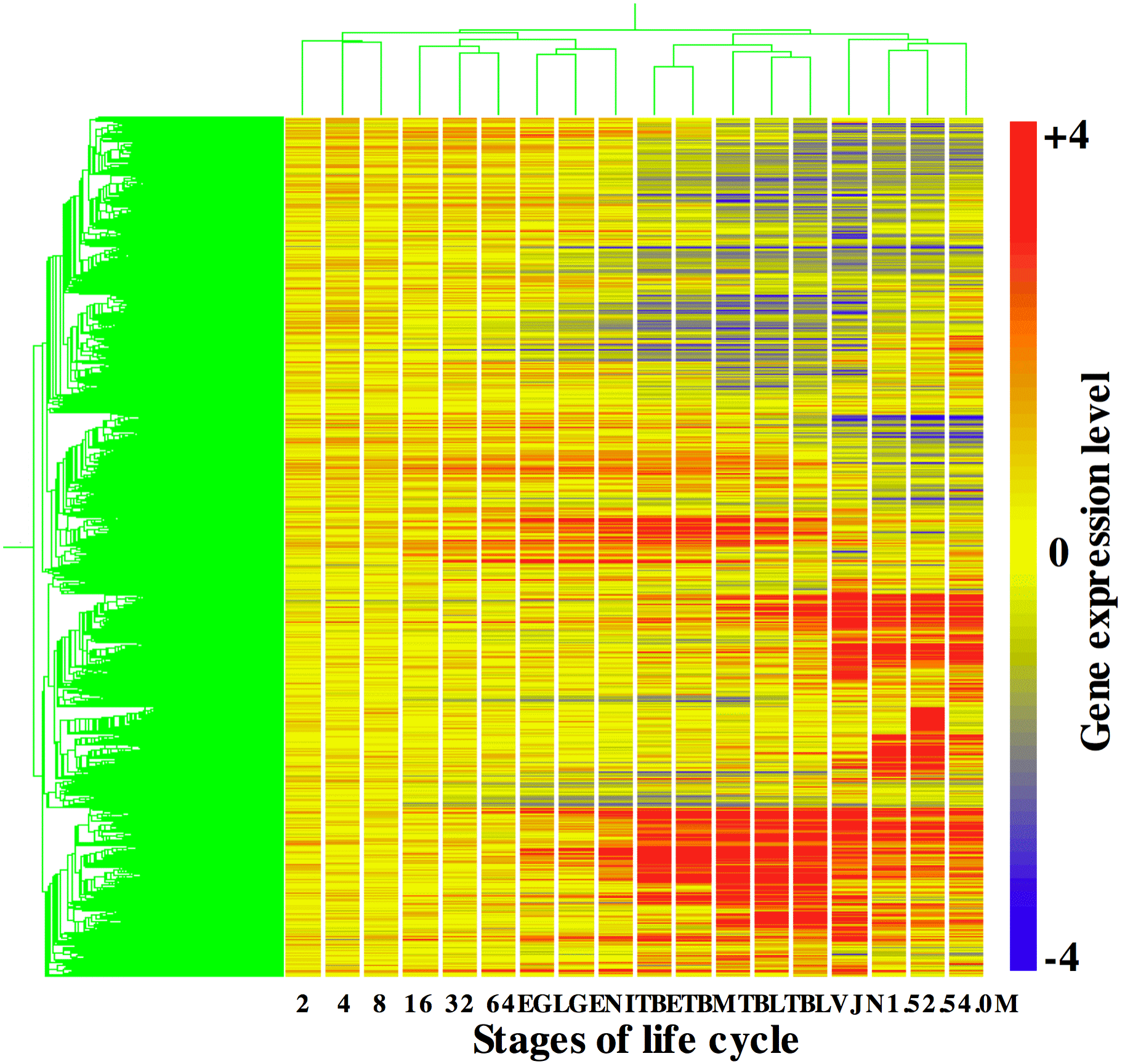
level at the first stage of the period. “Significantly changed genes” comprises the significantly differentially-expressed genes, which are those genes having, during the indicated period, an expression level which increased or decreased more than two-fold compared with the level at the first stage of the period.

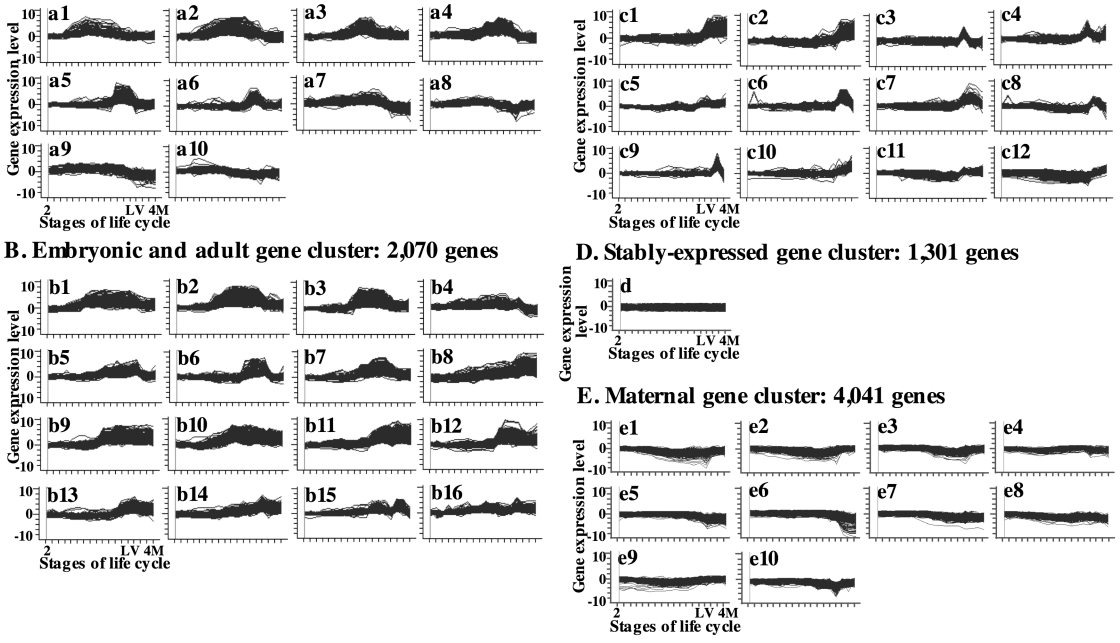
Fig. 4. Proportion of genes belonging to the five large clusters for: all genes (a), human conserved genes, human hypothetical genes and no-similarity genes (b), organ-specific genes and juvenile-specific genes (c), and embryogenesis- and morphogenesis-related genes, and immunity-related genes (d). A, embryonic gene cluster; B, embryonic and adult gene cluster; C, adult gene cluster; D, stably-expressed gene cluster; E, maternal gene cluster. The total number of genes in each category is indicated in brackets in each graph’s legend. The percentage of genes present in each cluster is indicated in brackets in the pie graph.

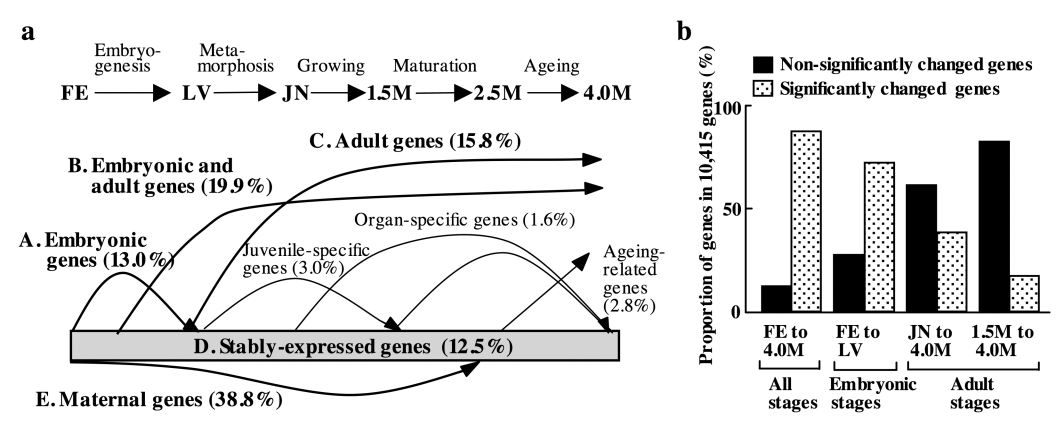
Fig. 5. Distribution of ovary-, testis-, endostyle-, neural complex-, heart- and blood cell-specific genes in five clusters. A, embryonic gene cluster; B, embryonic and adult gene cluster; C, adult gene cluster; D, stably-expressed gene cluster; and E, maternal gene cluster. Total number of each organ-specific gene is in brackets. The percentage in each cluster is indicated in the circle graph.

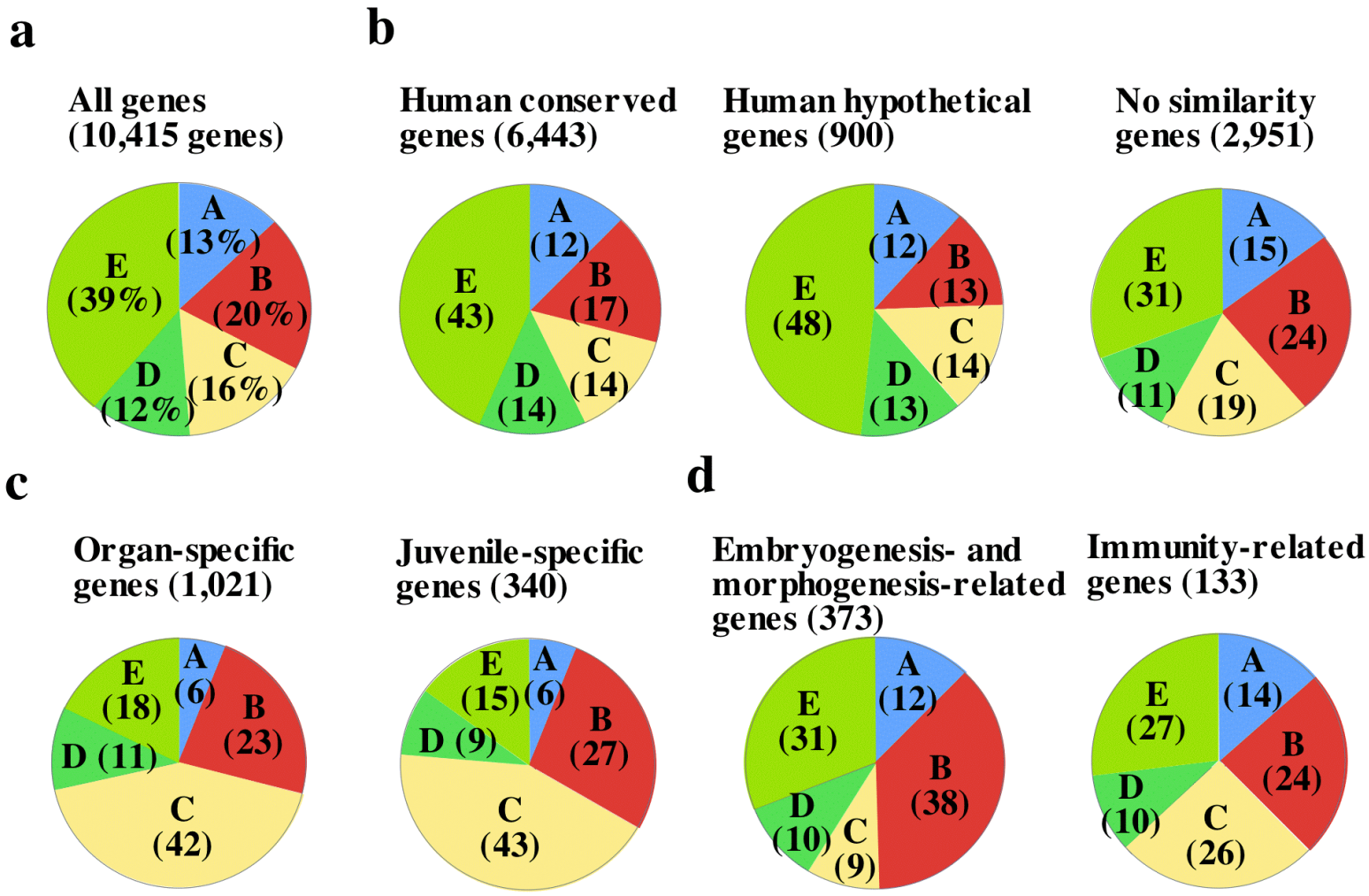
Fig. 6. Expression patterns of *ZicL*, *Brachyury* and *Brachyury*-downstream genes (a) and complement component genes (b). The large white vertical arrows between graphs, inside

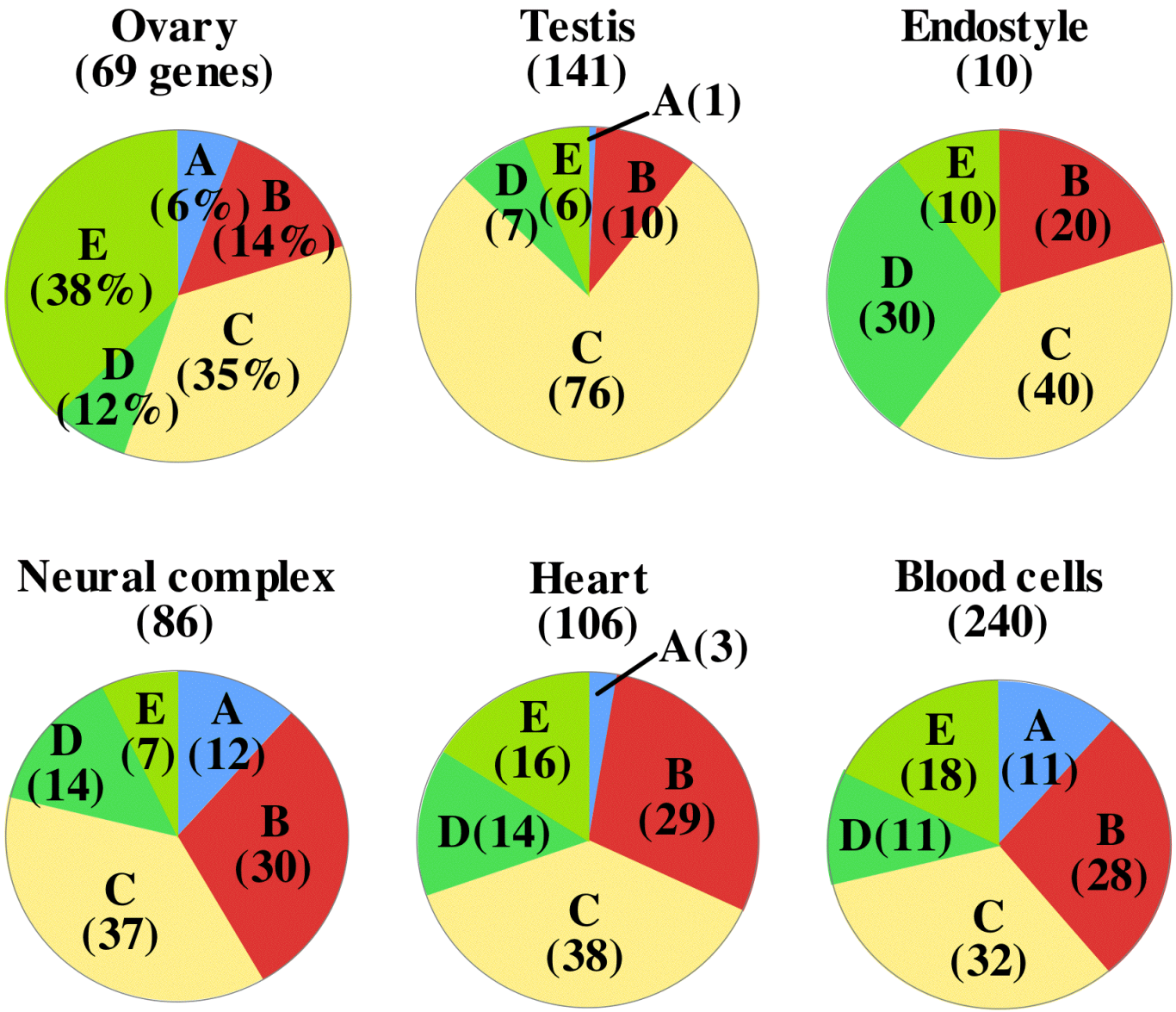
the graphs, indicate the downstream direction of the gene regulation cascade. The thin black vertical arrows indicate the timing of gene activation. Expression levels are shown as logarithmic ratios (base 2).

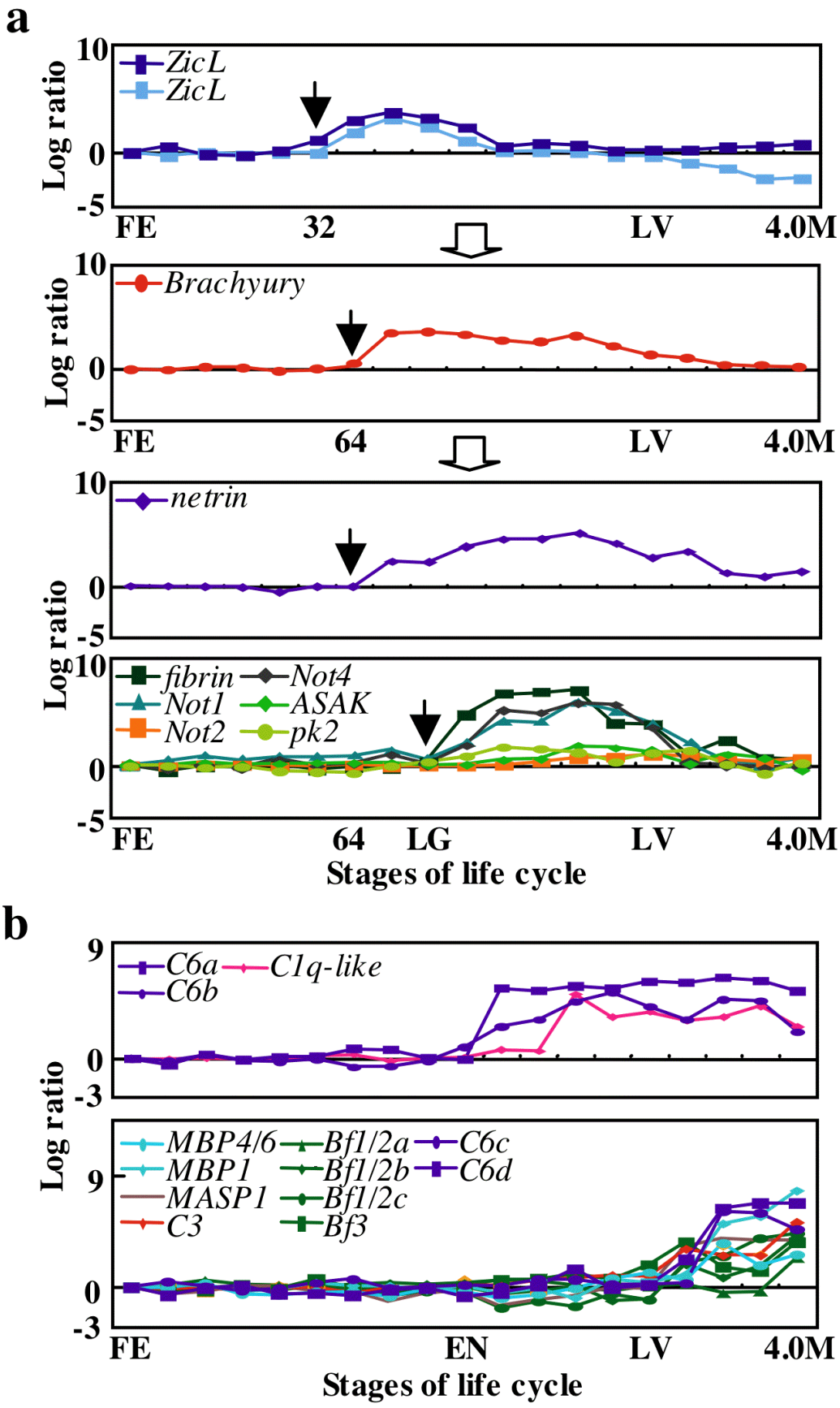

+4
0
-4
Gene expression level
2
4
8
16
32
64
EG
LG
EN
ITB
ET
BM
TB
LT
BL
V
J
N1
52
54
.0
M
Stages of life cycle


A. Embryonic gene cluster: 1,357 genes
a1
a2
a3
a4
a5
a6
a7
a8
a9
a10
Gene expression level
10
0
-10
2
LV
4M
Stages of life cycle
B. Embryonic and adult gene cluster: 2,070 genes
b1
b2
b3
b4
b5
b6
b7
b8
b9
b10
b11
b12
b13
b14
b15
b16
C. Adult gene cluster: 1,646 genes
c1
c2
c3
c4
c5
c6
c7
c8
c9
c10
c11
c12
D. Stably-expressed gene cluster: 1,301 genes
d
E. Maternal gene cluster: 4,041 genes
e1
e2
e3
e4
e5
e6
e7
e8
e9
e10


a
Embryo-genesis
Meta-morphosis
Growing
Maturation
Ageing
FE
LV
JN
1.5M
2.5M
4.0M
C. Adult genes (15.8%)
B. Embryonic and adult genes (19.9%)
Organ-specific genes (1.6%)
A. Embryonic genes (13.0%)
Juvenile-specific genes (3.0%)
Ageing-related genes (2.8%)
D. Stably-expressed genes (12.5%)
E. Maternal genes (38.8%)
b
Non-significantly changed genes
Significantly changed genes
Proportion of genes in 10,415 genes (%)
100
50
0
FE to 4.0M
FE to LV
JN to 4.0M
1.5M to 4.0M
All stages
Embryonic stages
Adult stages


a
All genes
(10,415 genes)
A
(13%)
B
(20%)
C
(16%)
D
(12%)
E
(39%)
b
Human conserved
genes (6,443)
A
(12)
B
(17)
C
(14)
D
(14)
E
(43)
Human hypothetical
genes (900)
A
(12)
B
(13)
C
(14)
D
(13)
E
(48)
No similarity
genes (2,951)
A
(15)
B
(24)
C
(19)
D
(11)
E
(31)
c
Organ-specific
genes (1,021)
A
(6)
B
(23)
C
(42)
D (11)
E
(18)
Juvenile-specific
genes (340)
A
(6)
B
(27)
C
(43)
D (9)
E
(15)
d
Embryogenesis- and
morphogenesis-related
genes (373)
A
(12)
B
(38)
C
(9)
D
(10)
E
(31)
Immunity-related
genes (133)
A
(14)
B
(24)
C
(26)
D
(10)
E
(27)


Ovary
(69 genes)
A
(6%)
B
(14%)
C
(35%)
D
(12%)
E
(38%)
Testis
(141)
A(1)
E
(6)
B
(10)
C
(76)
D
(7)
Endostyle
(10)
E
(10)
B
(20)
C
(40)
D
(30)
Neural complex
(86)
A
(12)
B
(30)
C
(37)
D
(14)
E
(7)
Heart
(106)
A(3)
B
(29)
C
(38)
D(14)
E
(16)
Blood cells
(240)
A
(11)
B
(28)
C
(32)
D(11)
E
(18)


a
10
0
-5
Log ratio
ZicL
ZicL
FE
32
LV
4.0M
Brachyury
64
netrin
fibrin
Not4
Not1
ASAK
Not2
pk2
LG
Stages of life cycle
b
9
-3
C6a
C1q-like
C6b
MBP4/6
Bf1/2a
C6c
MBP1
Bf1/2b
C6d
MASP1
Bf1/2c
C3
Bf3
EN