Gene-regulatory networks in the Ciona embryos

Atsushi Kubo, Kaoru S. Imai and Yutaka Satou

Abstract

Ascidians belong to the subphylum Urochordata or Tunicata, which is the sister group of the vertebrates. The simple architecture of the ascidian larva represents the basic chordate body plan. Recent analyses have shown many instances of developmental mechanisms conserved during evolution, while these studies have also revealed a much larger number of instances of divergence. However, to precisely determine the degree of conservation and divergence, that is, how many ways are used to make tadpole-like larvae, we need a systems-level understanding of development. Because animal development is organized by the genome and the minimal functional unit of development is a cell, comprehensiveness and single-cell resolution are necessary for a systems-biological understanding of the development. In the ascidian Ciona intestinalis, gene-regulatory networks responsible for the embryonic development have been studied on a genome-wide scale and at single-cell resolution. The simplicity and compactness of the genome facilitates genome-wide studies. In the Ciona genome, only ~670 transcription factor genes are encoded, and their expression profiles during the embryonic development have been analyzed. Gene-knockdown analyses of the transcription factor genes expressed during the embryonic development have been performed. The simplicity of the embryo permits these analyses to be done at single-cell resolution. Actually, these simple embryos are now being modeled in the computer, which allows us to understand the gene-regulatory networks very precisely in three dimensions.

Keywords: gene-regulatory networks; genome; Ciona intestinalis

INTRODUCTION

The ascidian belongs to the phylum Chordata, which also includes vertebrates and cephalochordates (amphioxus) (Figure 1A). The ascidian tadpole larva represents a body plan shared among animals in this phylum. In the tail, it has a dorsal nerve cord and a central notochord, which is flanked laterally by muscle cells and ventrally by an endodermal strand (Figure 1B). The ascidian has a smaller genome and shows a simpler mode of development than vertebrates. The genome of Ciona intestinalis, which is one of the most extensively studied species, is ~160 Mb and encodes almost 16 000 genes [1, 2]. The ascidian larva consists of about 2600 cells, and gastrulation begins at the 110-cell stage. This means that the developmental mechanisms shared among the chordates can be dissected in the ascidian experimental system with a simple genome and a simple embryo [3, 4].

Many studies of ascidian embryogenesis have been done at single-cell resolution, because most of the cells in the early embryo can be identified under the microscope. Since cell lineages in the embryo have been extensively determined, identified single cells can be traced along a timeline [5, 6].

Because of the smaller number of genes in the ascidian genome compared to vertebrate genomes, genome-wide analyses can be relatively easily performed in the Ciona system. First, the number of genes that should be analyzed is small. This is an especially big advantage for comprehensive assays of regulatory interactions, since the number of possible interactions will become much smaller. Second, the smaller number of genes in the ascidian genome

Corresponding author: Yutaka Satou, Department of Zoology, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto, 606-8502, Japan. Tel: +81-75-753-4095, Fax: +81-75-753-4095, E-mail: yutaka@ascidian.zool.kyoto-u.ac.jp

Atsushi Kubo is a final year PhD student at Kyoto University.

Kaoru S. Imai is an assistant professor at the Department of Biodiversity, Graduate School of Science of Kyoto University.

Yutaka Satou is an associate professor at the Department of Zoology, Graduate School of Science of Kyoto University.

© The Author 2009. Published by Oxford University Press. For permissions, please email: journals.permissions@oxfordjournals.org
is partly due to a smaller number of paralogs. This means that gene-knockdown studies, which are achieved with morpholino antisense oligonucleotides, can be performed straightforwardly, because functional redundancies between paralogous genes do not need to be considered in most cases.

The small size of the genome is also partly due to shorter intergenic regions. For example, the cis-regulatory regions required for proper gene expression are typically 3–5 kb in length, and in an extreme example only a 103-bp region is sufficient for promoting a reporter gene to recapitulate the endogenous expression of the gene [7].

We review the current understanding of the gene-regulatory networks underlying ascidian development in the following sections detailing how they are analyzed at the single cell level and genomewide. Both points of view are important for future systems-biological understandings of animal development, because all developmental processes proceed using cells as the most basic functional units and because all developmental processes proceed as a result of coordinated regulation of the genome.

### THE ASCIDIAN EMBRYO CAN BE ANALYZED AT THE SINGLE-CELL LEVEL AND ON A GENOME-WIDE SCALE

In the draft genome sequence of *Ciona intestinalis* [1], genes-encoding sequence-specific DNA-binding transcription factors with known DNA-binding motifs have been identified comprehensively (~670 genes) [8–11]. This includes genes encoding zinc finger proteins, a fraction of which are transcription factors. The expression patterns of these transcription factors were revealed by *in situ* hybridization from eggs until the stage just before hatching [12, 13]. Up to the gastrula stage, individual cells within the embryo can be identified under the microscope, and therefore the expression patterns of these transcription factor genes in the early embryos can be described at single-cell resolution (Figure 2A and B). Similarly, genes that are involved in major signaling pathways were comprehensively annotated and their embryonic-expression profiles were described [14, 15].

The early embryos were reconstructed virtually in computers based on confocal microscopic observations [16]. These virtual embryos allow us to estimate the volume and shape of each cell, spatial relationships between cells and contact surfaces between two neighboring cells. Each cell within the embryo has a unique name according to its lineage. Accordingly, cells can be traced in the 4D space.

The combinatorial use of comprehensive expression profiles of transcription factor genes and 4D information of the embryos predicts possible regulatory interactions among transcription factor genes and signaling genes [17, 18]. As far as the regulatory genes are concerned, possible interactions can be predicted genome-widely with this method.

Many studies have examined and confirmed such regulatory interactions experimentally by gene-knockdown and overexpression methods followed by *in situ* hybridization [19–29], and most of the predicted interactions have been experimentally examined [17]. The reconstituted ‘provisional’ networks can explain why given genes are expressed in specific cells in most cases.
For example, Otx is expressed in two pairs of cells of the animal hemisphere at the 32-cell stage as a result of induction by FGF9/16/20 (an FGF gene orthologous for vertebrate FGF9, FGF16 and FGF20), as demonstrated by the fact that Otx expression is down-regulated in FGF9/16/20-knockdown embryos [24, 30]. This FGF is thought to be secreted from five pairs of vegetal cells, because its mRNA is detected in these cells [20]. The sums of the contact surfaces between the vegetal cells expressing FGF9/16/20 and each of the two cells with Otx expression are much larger than those between cells expressing FGF9/16/20 and each of the other animal cells, according to the 3D-virtual embryo. An experiment in which the contact surfaces were forced to change showed that this difference of contact surface is actually important for the induction [16].

Although embryos in later stages have not yet been reconstructed virtually, most of the cells can be identified under the microscope. Direct lineage trace experiments can also be achieved easily by Dil-labeling of a single blastomere [23]. The ascidian embryo develops a dorsal neural tube, the anterior part of which differentiates into the brain and the posterior part of which differentiates into the dorsal nerve cord. Because the cell lineages of the cells in the ascidian central nervous system have been traced in detail [6], and because each cell can be identified under the fluorescence microscope after DAPI staining of nuclei, regulatory genes and regulatory interactions among them have been examined in the same way in the nervous system as in the early embryos. The medio-lateral axis of the dorsal nerve cord is determined by signaling of Nodal, which is secreted from cells forming the dorsal row of the nerve cord at the late gastrula stage [25, 31]. Although the ascidian brain lacks a clear midbrain-equivalent region, the brain is regionalized by FGF8/17/18, which is reminiscent of the function of FGF8 in the isthmus organizer of the vertebrates [32]. This FGF is expressed in a pair of cells that give rise to four pairs of cells in the visceral ganglion of the tailbud embryo, and specifies two anterior neighbors as the neck region; without this induction, the presumptive neck cells become a part of the sensory vesicle (Figure 2C).
**cis-REGULATORY MECHANISMS IN THE CIONA EMBRYO**

The binding of multiple transcription factors to DNA has been shown to be a combinatorial input at the level of the *cis*-regulatory mechanism. Thus, determining whether given interactions between transcription factors and their downstream genes are direct or indirect is important for understanding the architecture of the networks. DNA constructs can easily be introduced into fertilized eggs by electroporation [33] and microinjection, and therefore analysis of *cis*-regulatory mechanisms can be performed relatively easily in *Ciona* embryos. Another advantage is the dense encoding of genes in the *Ciona* genome (16 000 genes/160 Mb = 1 gene/10 kb). Moreover, the *cis*-regulatory elements that are known so far are located proximal to the transcription start sites of genes.

Precise mapping of transcription start sites is important for analyzing *cis*-regulatory mechanisms of gene transcription. A significant fraction of *Ciona* gene transcripts (~50%) are subjected to trans-splicing, as a result of which they have 16-bp of 5′-spliced-leader sequences at their 5′-termini [2, 34, 35]. Although determining the precise transcription start sites of these genes is difficult, most of their *trans*-splicing-acceptor sites have been determined [2]. The transcription start sites (TSSs) of about 800 genes whose transcripts are not usually *trans*-spliced were also determined by the oligocapping method [2, 34].

Although the 5′-ends (5′-*trans*-splicing-acceptor sites and TSSs) of a significant part of the *Ciona* genes have been identified, there is an additional feature that should be noted: some *trans*-spliced genes are encoded in operons, in which two or more genes are transcribed as a single molecule. The operonic transcripts are processed into individual functional messages by *trans*-splicing between spliced-leader RNA and downstream genes within operons. In the *Ciona* genome, about one-fifth (~3000) of all genes are encoded as operons in the genome [2]. Although other chordates do not appear to have this type of genomic architecture, not a small number of non-chordate animals have operonic genes in their genomes. The uniqueness of the *Ciona* operons is that there are no intergenic regions between members of each operon [34]. This suggests that it is not possible to resolve both of upstream and downstream genes into functional messages at the same time, because the *trans*-splice-acceptor site branch-point of the downstream mRNA must necessarily reside within the upstream mRNA and the upstream mRNA would presumably be released as a branched nucleic-acid structure [34]. It has also been found that each member within the single operons is not always expressed in the same tissue/cells and at the same time. Therefore, there should be a regulatory mechanism that determines which genes within the single operons are converted into functional messages, as well as a regulatory mechanism that activates the expression of operons. This subject is important for understanding the gene-regulatory networks of the ascidian embryo, because about one-fifth of the genes are encoded in operons. However, these regulatory mechanisms remain to be determined.

The *cis*-regulatory elements of almost 100 ascidian genes have been examined in detail [36]. In most cases, *cis*-regulatory elements directing specific expression are located within 5-kb upstream from transcription start sites. Conservation between the *Ciona intestinalis* genome and the genome of its close relative *Ciona savignyi* has been used for predicting functional *cis*-regulatory elements [37]. By taking advantage of the large amount of gene-expression data [38], efforts have been made to identify functional *cis*-regulatory elements shared by genes with similar expression patterns [39]. These approaches have actually succeeded in determining the functional elements of several genes. This success is partly due to the compactness of the genome, because this compactness facilitates comparisons between two different genomes and between different genes with similar expression patterns.

Comparative methods have advanced our understanding of the genome. However, these approaches could miss some functional elements that are not well conserved between two different genomes or between two different genes or that are difficult to predict for technical reasons (Figure 3). Recent studies in other animals have detected redundant sets of *cis*-regulatory elements [40] and numerous binding sites of transcription factors to DNA, the biological importance of which is still unknown [41]. This means that if a given element flanked by a reporter gene can recapitulate the endogenous gene expression, there may be more such elements for the same gene. Accordingly, we should note that comparative methods may bias our understanding of *cis*-regulatory mechanisms, in spite of the usefulness of the comparative approach for finding functional elements.
CONCLUSIONS

Gene-regulatory networks in *C. intestinalis* have been examined at single-cell resolution and on a genome-wide scale. These networks can explain why particular regulatory genes are expressed in specific cells in many cases. However, the networks are not yet well understood at the level of *cis*-regulatory mechanisms. Unbiased and genome-wide approaches including the chromatin-immunoprecipitation method will be required for understanding the architecture of the networks. Understanding the regulatory mechanisms of operonic gene expression will also be required. Lineage-specific gene knockdowns will be required for detailed dissection of networks in later stages, but such a technique is not well developed currently. Experiments addressing these problems will help to reveal the complete regulatory network for tunicates in every cell and at every stage of development.

There are at least 80 miRNA loci in the *Ciona* genome [42]. There is no evidence that these small RNAs function in the regulatory network, but because some of them are developmentally regulated [42], it is possible that these molecules have developmental roles.

While precise positioning of blastomeres in the embryo is an important factor in order for the gene-regulatory networks to work properly, the architecture of the embryo is directly or indirectly regulated by the gene-regulatory networks. The simple architecture of the *Ciona* embryo gives us an opportunity for analyzing how the regulatory networks control the shape or movement [43]. We may be close to a systems-biological understanding of development in the *Ciona* system.

### References


