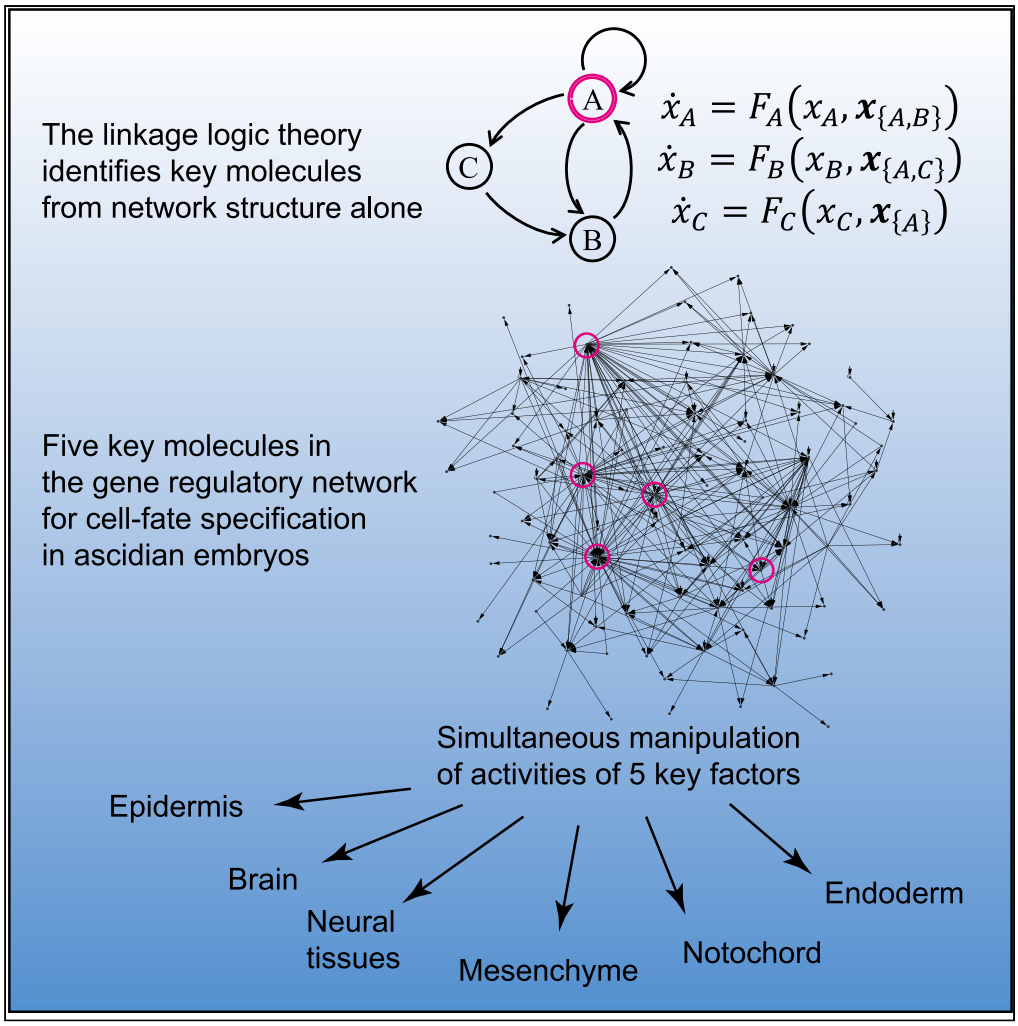
iScience

CellPress

Article

# Controlling Cell Fate Specification System by Key Genes Determined from Network Structure




Kenji Kobayashi, Kazuki Maeda, Miki Tokuoka, Atsushi Mochizuki, Yutaka Satou

mochi@riken.jp (A.M.)
yutaka@ascidian.zool.kyoto-u.ac.jp (Y.S.)

## HIGHLIGHTS

The linkage logic theory identifies key nodes in regulatory networks from structures

It identified five key molecules in the gene regulatory network (GRN) of an ascidian embryo

The network dynamics was controlled by manipulating activities of these key molecules

Kobayashi et al., iScience 4, 281–293
June 29, 2018 
https://doi.org/10.1016/j.isci.2018.05.004

iScience

CellPress

Article

# Controlling Cell Fate Specification System by Key Genes Determined from Network Structure

Kenji Kobayashi,[1,2] Kazuki Maeda,[2,3] Miki Tokuoka,[1,2] Atsushi Mochizuki,[2,4,5,*] and Yutaka Satou[1,2,6,*]

## SUMMARY

**Network structures describing regulation between biomolecules have been determined in many biological systems. Dynamics of molecular activities based on such networks are considered to be the origin of many biological functions. Recently, it has been proved mathematically that key nodes for controlling dynamics in networks are identified from network structure alone. Here, we applied this theory to a gene regulatory network for the cell fate specification of seven tissues in the ascidian embryo and found that this network, which consisted of 92 factors, had five key molecules. By controlling the activities of these key molecules, the specific gene expression of six of seven tissues observed in the embryo was successfully reproduced. Since this method is applicable to all nonlinear dynamic systems, we propose this method as a tool for controlling gene regulatory networks and reprogramming cell fates.**

## INTRODUCTION

Network systems produce dynamics of molecular activity in organisms, and such dynamics are thought to be the origin of biological functions (Alon, 2007; Oda et al., 2005; Peter and Davidson, 2016). A variety of cell types originate in the diversity of steady states of gene expression. We recently developed a new theoretical framework (linkage logic theory) (Fiedler et al., 2013; Mochizuki, 2008; Mochizuki et al., 2013), with which key nodes for controlling nonlinear dynamics are identified only from network structure without assuming quantitative details, such as functional forms, parameters, or initial states. According to this theory, the dynamics of a system is controllable to converge on any solution by controlling a subset of nodes called a feedback vertex set (FVS). Therefore, if the dynamics of a GRN has multiple steady states, we should be able to reproduce them and control the dynamics of the system by manipulating the activities of FVS molecules alone.

In the present study, we applied the linkage logic theory to a GRN that specifies cell fates in embryos of the ascidian *Ciona intestinalis* (type A; also called *Ciona robusta*). The network structure for the specification of cell fate has been determined by a genome-wide gene knockdown assay for regulatory genes that are expressed during embryogenesis (Imai et al., 2006) and was recently updated using data that had been accumulated after the initial construction (Satou and Imai, 2015). Hence, if the fate decision is based on the steady states of this network, cell-type-specific gene expression patterns should be reproduced by manipulating the activities of FVS in the network. Here, we show that the minimum FVSs of this network contain only five factors and that the dynamics of the GRN is indeed controllable by these five FVS factors.

## RESULTS

### Controlling Nonlinear Dynamics of Networks

First, we show the linkage logic theory is applicable to GRNs. A GRN is represented by a directed graph $\Gamma = (V, E)$, consisting of a node set $V$ and an edge set $E$, where nodes represent genes and edges represent regulatory linkages. The dynamics of gene activities is modeled by a system of ordinary differential equations. We assume that gene activities, measured in terms of the concentrations of mRNAs or proteins, decay in the absence of supply or synthesis. Suppose that the dynamics of activity $x_n$ of gene $n \in V$ is written in the form:

$$\begin{aligned} \dot{x}_n &= F_n(\boldsymbol{x}) \\ &= F_n(x_n, \boldsymbol{x}_{I_n}) \end{aligned} \qquad \text{(Equation 1)}$$

[1]Department of Zoology, Graduate School of Science, Kyoto University, Sakyo, Kyoto 606-8502, Japan

[2]CREST, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

[3]Department of Mathematical Sciences, School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda, Hyogo 669-1337, Japan

[4]Theoretical Biology Laboratory, RIKEN, Wako, Saitama 351-0198, Japan

[5]Institute for Frontier Life and Medical Sciences, Kyoto University, Sakyo, Kyoto 606-8507, Japan

[6]Lead Contact

*Correspondence: mochi@riken.jp (A.M.), yutaka@ascidian.zool.kyoto-u.ac.jp (Y.S.)

https://doi.org/10.1016/j.isci.2018.05.004

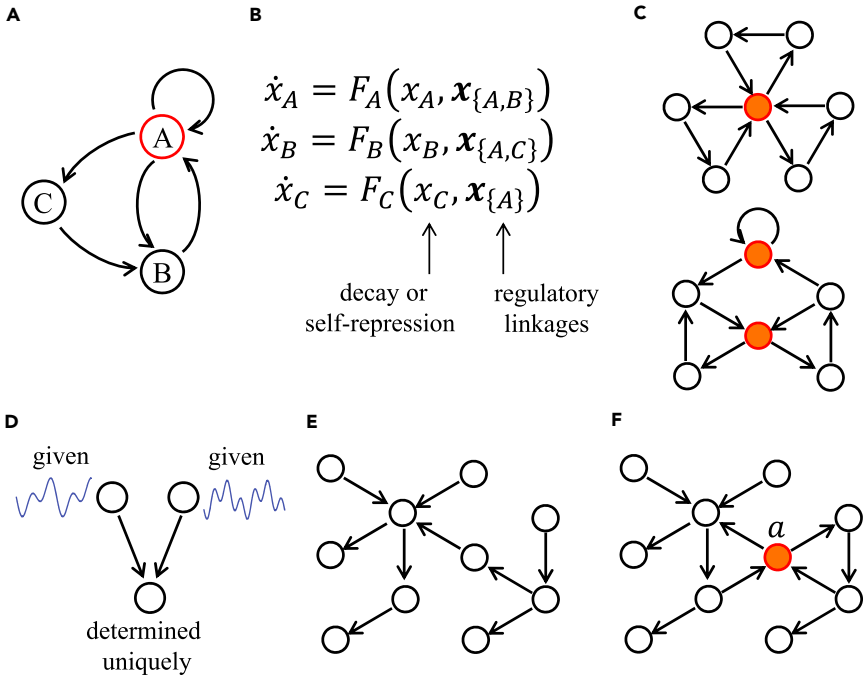



**Figure 1. Controlling Network Dynamics by FVS**

(A) A GRN containing three nodes and three cycles ($A \rightarrow A$, $A \rightarrow B \rightarrow A$, $A \rightarrow C \rightarrow B \rightarrow A$). Regulatory interaction is positive for the self-loop in *A* and either positive or negative for the others. The minimum FVS includes node *A*.

(B) An ODE system corresponding to the network in (A). The second argument sets in $F_n$s specify regulatory linkages.

(C) Two GRNs containing minimum FVSs marked in red.

(D) The dynamics of regulating nodes determines the dynamics of the regulated node uniquely.

(E) A network without cycles has an empty FVS.

(F) A network with two cycles has an FVS including node *a* only.

with the "decay condition":

$$\partial_1 F_n(x_n, \boldsymbol{x}_{I_n}) < 0. \qquad \text{(Equation 2)}$$

The set $I_n \subseteq V$ is the input set of $n$, a subset of molecules that regulate molecule $n$, that is, $I_n = \{i | (i \rightarrow n) \in E\}$. The notation $\partial_1$ implies the first partial derivative with respect to the first argument. The set $I_n$ includes $n$ ($n \in I_n$), that is, it is a self-regulatory loop, if $\partial F_n / \partial x_n$ is "not always negative." Note that even if $\partial F_n / \partial x_n$ is not negative, we can make the system satisfy the decay condition (2) by adding a positive term indicating a self-regulatory loop. The sets of $I_n$ ($\forall n \in V$) directly represent the graphical structure of the regulatory network. An example of a hypothetical network consisting of three genes is shown in Figures 1A and 1B.

Under formulations (1) and (2), we proved that sets of key nodes for dynamics are determined from the topology of the network (Fiedler et al., 2013; Mochizuki, 2008; Mochizuki et al., 2013) as FVSs. In graph theory, an FVS is defined as a subset of vertices in a directed graph whose removal leaves a graph without directed cycles (Akutsu et al., 1998). In the above hypothetical network, gene *A* constitutes the minimum FVS (Figure 1A). Two additional examples are shown in Figure 1C. Here, we give an intuitive explanation of our theory using illustrative examples (see Methods for details). (1) In a simple regulatory system including two regulator nodes and a regulated node (Figure 1D), if the dynamics of the regulator nodes is given, the dynamics of the regulated node is determined uniquely; that is, for any initial state, the dynamics of the node converges to a single trajectory for a long time. (2) In a GRN without a cycle (Figure 1E), the dynamics of "top" nodes, which receive no regulatory input, converges on the unique equilibrium. By determining the dynamics of each node downward through the network, the dynamics of a system without a cycle should converge on a unique equilibrium, which is globally stable. (3) Then, consider a GRN including two cycles as shown in Figure 1F, the corresponding dynamics of which can have multiple solutions. Say it has two steady states $S_1$ and $S_2$. Node *a* is a single element of FVS because the graph without node *a*

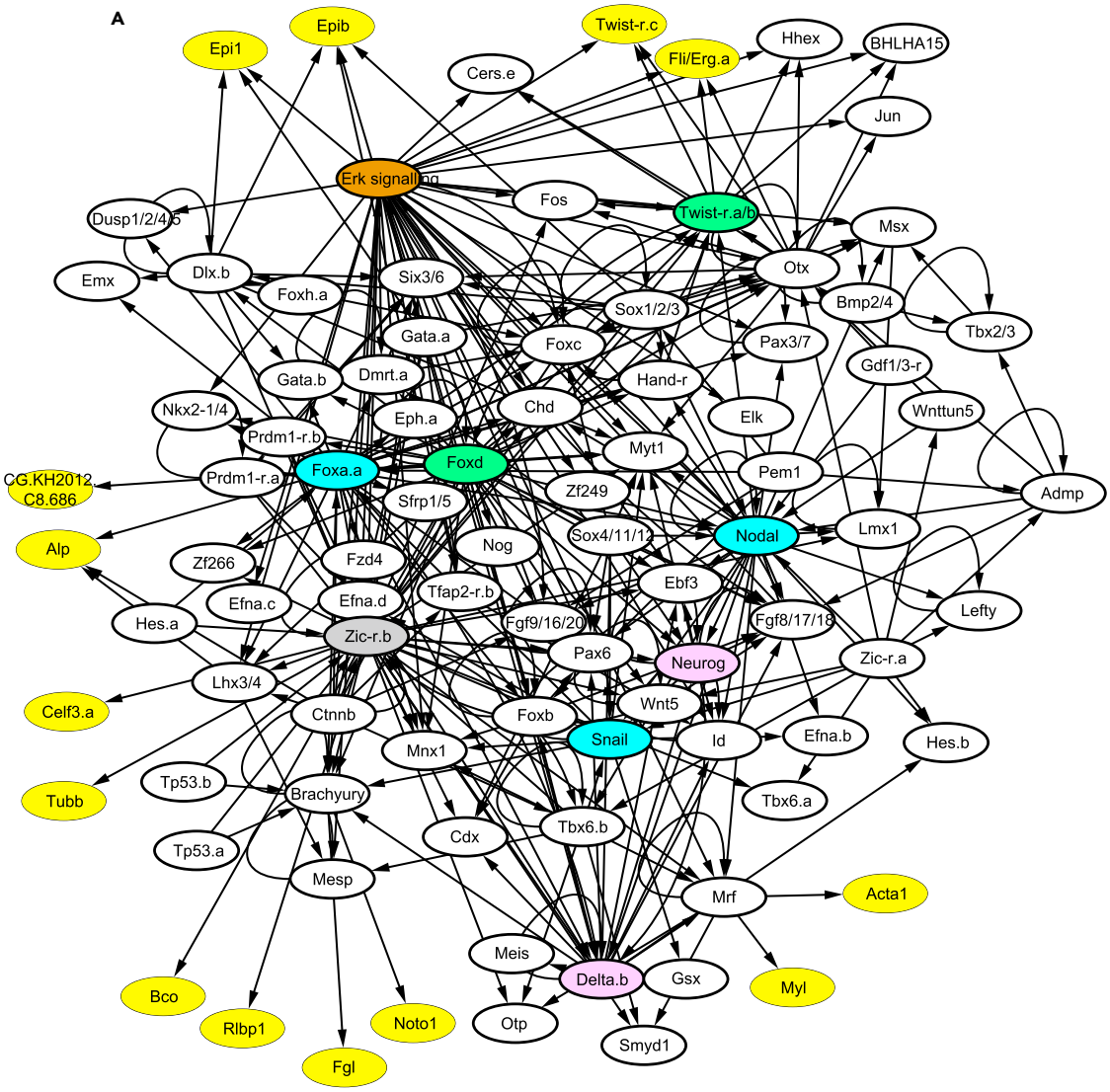

A
Epi1
Epib
Twist-r.c
Hhex
BHLHA15
Fli/Erg.a
Cers.e
Jun
Erk signalling
Fos
Twist-r.a/b
Msx
Dusp1/2/4/5
Emx
Dlx.b
Six3/6
Otx
Foxh.a
Sox1/2/3
Bmp2/4
Tbx2/3
Gata.a
Foxc
Pax3/7
Gdf1/3-r
Gata.b
Dmrt.a
Hand-r
Nkx2-1/4
Chd
Elk
Wnttun5
Eph.a
Prdm1-r.b
Myt1
Foxa.a
Foxd
Pem1
CG.KH2012
C8.686
Prdm1-r.a
Zf249
Admp
Sfrp1/5
Sox4/11/12
Nodal
Lmx1
Alp
Nog
Zf266
Fzd4
Ebf3
Tfap2-r.b
Efna.c
Efna.d
Lefty
Hes.a
Fgf9/16/20
Fgf8/17/18
Zic-r.b
Pax6
Neurog
Zic-r.a
Lhx3/4
Celf3.a
Wnt5
Ctnnb
Foxb
Snail
Id
Efna.b
Hes.b
Mnx1
Tp53.b
Tbx6.a
Tubb
Brachyury
Tbx6.b
Cdx
Tp53.a
Mesp
Mrf
Acta1
Meis
Delta.b
Gsx
Myl
Bco
Rlbp1
Noto1
Otp
Smyd1
Fgl
B
Ciona embryonic GRN
consisting of
92 nodes and 328 edges
identification of minimum
feedback vertex sets (mFVSs)
on the basis of the linkage logic theory
Identified 12 mFVSs
1. Foxa.a & Foxd & Neurog & Zic-r.b & Erk signalling
2. Foxa.a & Foxd & Delta.b & Zic-r.b & Erk signalling
3. Foxa.a & Twist-r.a/b & Neurog & Zic-r.b & Erk signalling
4. Foxa.a & Twist-r.a/b & Delta.b & Zic-r.b & Erk signalling
5. Nodal & Foxd & Neurog & Zic-r.b & Erk signalling
6. Nodal & Foxd & Delta.b & Zic-r.b & Erk signalling
7. Nodal & Twist-r.a/b & Neurog & Zic-r.b & Erk signalling
8. Nodal & Twist-r.a/b & Delta.b & Zic-r.b & Erk signalling
9. Snail & Foxd & Neurog & Zic-r.b & Erk signalling
10. Snail & Foxd & Delta.b & Zic-r.b & Erk signalling
11. Snail & Twist-r.a/b & Neurog & Zic-r.b & Erk signalling
12. Snail & Twist-r.a/b & Delta.b & Zic-r.b & Erk signalling

**Figure 2. Gene Regulatory Network for Cell Specification in *Ciona***

(A) The GRN consists of 92 factors (nodes) and 328 regulatory interactions (edges). The network possessing minimum FVS consists of five nodes. The 12 choices of the node sets are given by choosing a single node from each of five node sets colored light blue, green, pink, gray, and orange. Nodes filled in yellow are the marker genes, for which we performed observations of activities. See also Figure S1 and S2.

(B) List of nodes in the 12 minimum FVSs.

includes no cycle. Suppose we fix the activity level of node *a* to be equal to the value at steady state $S_1$ by experimental manipulation. Then, the remaining nodes constitute a graph without a cycle, and the dynamics of these nodes spontaneously converges on the unique steady state, which must be the same as the steady state $S_1$ in the original system. If instead we fix the activity level of the node equal to the value of steady state $S_2$, then the dynamics of other genes converges on the unique equilibrium that is equal to steady state $S_2$. Note that the dynamics of the top nodes, which receive no regulatory input, converges on the unique equilibrium. Hence, by experimentally manipulating a single node, we can make the system converge on steady state $S_1$ or steady state $S_2$ as required. When FVSs include multiple nodes, all of them must be fixed simultaneously to control the dynamics to converge on desired steady states.

The controllability by FVS has a broader meaning than switching between solutions that can be observed in natural conditions. For any fixed value of nodes in an FVS, the dynamics of other nodes, which are not included in the FVS, converges on a unique steady state, even if the given value is not chosen from known natural steady states. This implies that an exhaustive search of steady states is possible under an assumption of discreteness. If we assume that all possible steady states in a given system have binary values, that is, 0 or 1, on an FVS, we can examine all combinations of manipulations of nodes in the FVS. The obtained set of the states should include all the natural steady states of the system.

## FVSs of the Gene Regulatory Network for Fate Specification in Ascidian Embryos

We tested the FVS controllability using the GRN to specify cell fates in an ascidian embryo (Imai et al., 2006; Satou and Imai, 2015). Before the late gastrula stage of *Ciona* embryos, the cell fate of each blastomere is restricted to one of seven tissues, epidermis, brain, nerve cord, endoderm, notochord, mesenchyme, or muscle, which exhibit specific gene expression patterns in their descendants at later stages. Zygotic expression starts between the 8- and 16-cell stages, and the dynamics of gene expression until the late gastrula stage specifies the developmental fates of the above-mentioned seven tissues. The GRN responsible for specification of these cell fates includes 92 genes and 328 regulatory linkages (Figures 2A and S1). From an analysis of the structure of the network (see Methods), we identified 12 minimum FVSs, each of which contained five genes (color outlined in Figure 2). The minimum FVSs are {*Foxa.a*|*Nodal*|*Snail*, *Foxd*|*Twist-r.a/b*, *Neurog*|*Delta.b*, *Zic-r.b*, Erk signaling}, where "|" indicates an alternative choice ($3 \times 2 \times 2 \times 1 \times 1 = 12$ sets; see Figure 2B). The existence of FVSs indicates that the GRN potentially possesses multiple steady states. If the activity of the FVS factors is assumed to be binary, that is, active or inactive, all steady states will be obtained by up- and down-regulation of the activities of molecules in an FVS, as discussed earlier.

In multicellular embryos, GRNs encoded in individual cells are mutually connected through intercellular interactions and function as subnetworks to constitute a larger GRN. In addition, such interactions are affected by three-dimensional structures unique to various stages of embryos. To avoid such possible effects, we developed an experimental system of single-cell development by treating fertilized eggs with cytochalasin B (CytB) (Figure 3A). Although cells in CytB-treated embryos never divide, nuclear divisions continue, and specification dynamics is considered to proceed (Hudson et al., 2003; Hudson and Yasuo, 2006; Jeffery et al., 2008; Kodama et al., 2016; Meedel et al., 2007; Oda-Ishii and Di Gregorio, 2007; Satoh, 1979; Shi and Levine, 2008; Tokuoka et al., 2004; Whittaker, 1973; Yasuo and Hudson, 2007). We did not exclude signaling molecules from our analysis, because signaling molecules could work in an autocrine manner. To identify cell fates by reverse transcription-quantitative PCR (RT-qPCR) and *in situ* hybridization, we chose the following genes as markers: *Epi1* and *Epib* for the epidermis, *Bco* and *Rlbp1* for the brain, *Celf3.a* and *Tubb* for the entire neural system, *Alp* and *CG.KH2012.C8.686* for the endoderm, *Noto1* and *Fgl* for the notochord, *Fli/Erg.a* and *Twist-r.c* for the mesenchyme, and *Myl* and *Acta1* for the muscle (Chiba et al., 1998; Hotta et al., 1999; Imai et al., 2000, 2003, 2004; Kusakabe et al., 2002; Satou et al., 2001; Takahashi et al., 1999; Ueki et al., 1994; Yagi and Makabe, 2001). We confirmed that these marker genes were indeed regulated by *Dlx.b*, *Zic-r.b*, *Foxa.a*, *Brachyury*, *Twist-r.a/b*, or *Mrf*, which were included in the above-mentioned GRN (Figure S2).

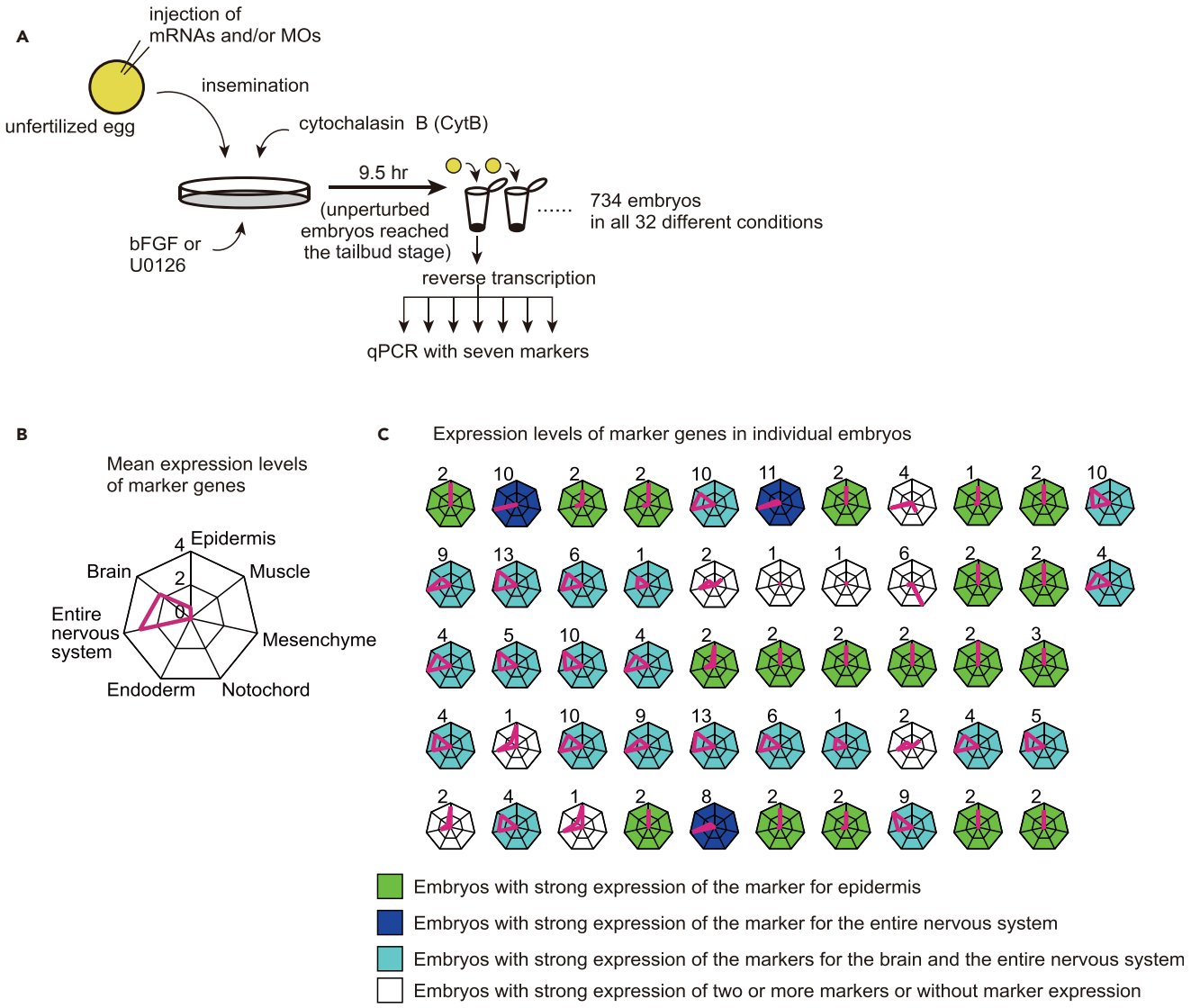



**Figure 3. The Experimental System Using Syncytium Embryos**

(A) Depiction of the experimental design for testing the prediction of the linkage logic theory.

(B and C) Syncytium embryos treated with CytB expressed epidermal and neural markers. Expression levels of marker genes are shown relative to the corresponding values in normal 9.5-hr (tailbud-stage) embryos. (B) Mean values and (C) values of individual embryos are shown. In (B), in these experimental embryos, two markers for the brain and the entire nervous system were predominantly expressed, and an epidermal marker was also expressed. Other markers were rarely expressed. The axes in (C) show expression of the same tissue marker genes as in (B). Embryos that predominantly expressed the epidermal marker, a set of the pan-neural and brain markers, and pan-neural marker are shown in green, light blue, and blue, respectively.

To confirm that specification dynamics proceeds in CytB-treated embryos and to examine whether the cell fate specification in CytB-treated embryos is deterministic, we used RT-qPCR and measured the expression of seven marker genes, *Epi1*, *Bco*, *Celf3.a*, *Alp*, *Noto1*, *Fli/Erg.a*, and *Myl*, for seven different tissues in 52 embryos at 9.5 hr after fertilization, which corresponded to the tailbud stage in normal embryos (Figures 3B and 3C; Table 1). Expression levels of marker genes were measured relative to the corresponding values in normal 9.5 hr embryos. Among the 52 embryos, 19, 21, and 3 strongly expressed *Epi1*, a set of *Bco* and *Celf3.a*, and *Celf3.a*, respectively. Namely, the marker gene expression patterns in these embryos resembled those in epidermal, brain, and nerve cord cells. Seven embryos expressed marker genes for multiple tissues, and the remaining two embryos rarely expressed marker genes. Under the manipulation of CytB treatment, cell specification became nondeterministic, and the resultant diversity of marker gene expression was smaller than in normal embryos. In contrast, the observation that marker genes were expressed at

| | | Experimental Condition[a] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Unperturbed | adnze | adnZe | adNze | Adnze | aDnze | adnZE | adnz | Z |
| Expression[b] | *Epi1* (epidermis) | 0.63 | 0.91 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.00 |
| | *Bco* (brain) | 2.30 | 0.08 | 2.20 | 0.06 | 0.15 | 0.14 | 0.43 | 0.18 | 2.16 |
| | *Celf3.a* (pan-neural) | 3.03 | 0.15 | 3.61 | 9.59 | 0.19 | 0.17 | 0.18 | 0.34 | 2.60 |
| | *Alp* (endoderm) | 0.01 | 0.02 | 0.03 | 0.13 | 1.94 | 0.83 | 0.01 | 0.04 | 0.03 |
| | *Noto1* (notochord) | 0.15 | 0.01 | 0.01 | 0.00 | 0.31 | 2.07 | 0.00 | 0.01 | 0.01 |
| | *Fli/Erg.a* (mesenchyme) | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 6.21 | 0.00 | 4.76 |
| | *Myl* (muscle) | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.08 |

**Table 1. Mean Expression Levels of Marker Genes in Nine Representative Conditions**

[a]Each of the experimental conditions is represented by a five-letter code in which up- and down-regulation of *Foxa.a*, *Foxd*, *Neurog*, *Zic-r.b*, and Erk signaling are represented by A/a, D/d, N/n, Z/z, and E/e, respectively. See also Table S1.

[b]Expression levels of marker genes are shown relative to the corresponding values in normal 9.5-hr (tailbud-stage) embryos.

9.5 hr after fertilization suggested that the dynamics of the GRN for cell fate specification proceeded in CytB-treated embryos.

## Dynamics of the Network for Cell Fate Specification Was Controllable by Manipulating the Activities of the FVS Factors

Among the 12 minimum FVSs, we chose an FVS consisting of *Foxa.a*, *Foxd*, *Neurog*, *Zic-r.b*, and Erk signaling because we have morpholino antisense oligonucleotides that are effective for the knockdown of *Foxa.a*, *Foxd*, *Neurog*, and *Zic-r.b* (Hudson et al., 2016; Imai et al., 2006). In addition, we used synthetic mRNAs for up-regulation of the activities of these genes. For the up- and down-regulation of Erk signaling, we added a recombinant FGF protein and an MEK inhibitor to seawater. Using these experimental tools, we performed exhaustive manipulation in a binary manner (i.e., up- or down-regulation of these FVS factors; $2^5 = 32$ combinations) to identify all possible steady states that the system reaches.

We examined a total of 734 embryos by RT-qPCR for seven marker genes including at least 12 embryos of a single batch for each of the 32 conditions (Figures S3–S6; Table S1). The expression of marker genes was deterministic in most cases under the manipulation of FVS factors. Namely, the embryos under the same manipulating condition exhibit almost the same pattern of expression of the marker genes. We applied sign tests to examine whether a single tissue marker or a set of *Bco* and *Celf3.a* was predominantly expressed in each of the 32 conditions (Figure S7). The expression of marker genes is summarized in Figure 4. Each of the experimental conditions is represented by a five-letter code in which up- and down-regulation of *Foxa.a*, *Foxd*, *Neurog*, *Zic-r.b*, and Erk signaling are represented by A/a, D/d, N/n, Z/z, and E/e, respectively; for example, embryos exhibiting up-regulation of *Foxa.a* and down-regulation of the other factors are referred to as Adnze. As shown in Figure 4, in 22 conditions, *Epi1*, a set of *Bco* and *Celf3.a*, *Celf3.a*, *Alp*, *Noto1*, or *Fli/Erg.a* was expressed (Figure 4A). Such gene expression patterns were similar to those in cells of the epidermis, brain, nerve cord, endoderm, notochord, and mesenchyme, respectively. In contrast, we rarely observed simultaneous expression of markers for multiple tissues in a single embryo. Figure 4B shows the average relative gene expression in six representative conditions (see also Table 1). In these six conditions, we also performed *in situ* hybridization with the same set of markers and an additional set of markers, *Epib*, *Rlbp1*, *Tubb*, *CG.KH2012.C8.686*, *Fgl*, and *Twist-r.c* (Figures 4C, 4D, and S8). The results were consistent with those of the aforementioned RT-qPCR. The observation that gene expression patterns did not basically differ among individual embryos in each condition indicated that manipulation of the activities of the FVS factors was deterministic.

## Dynamics of the Network for Cell Fate Specification Was Not Controllable by Manipulating the Activities of a Subset of the FVS Factors

In contrast to the findings described earlier, the manipulation of activities of a subset of the FVS factors (*Foxa.a*, *Foxd*, *Neurog* and *Zic-r.b*, but not Erk signaling) did not drive the GRN dynamics deterministically into a single steady state. Namely, gene expression patterns differed among the individual embryos that

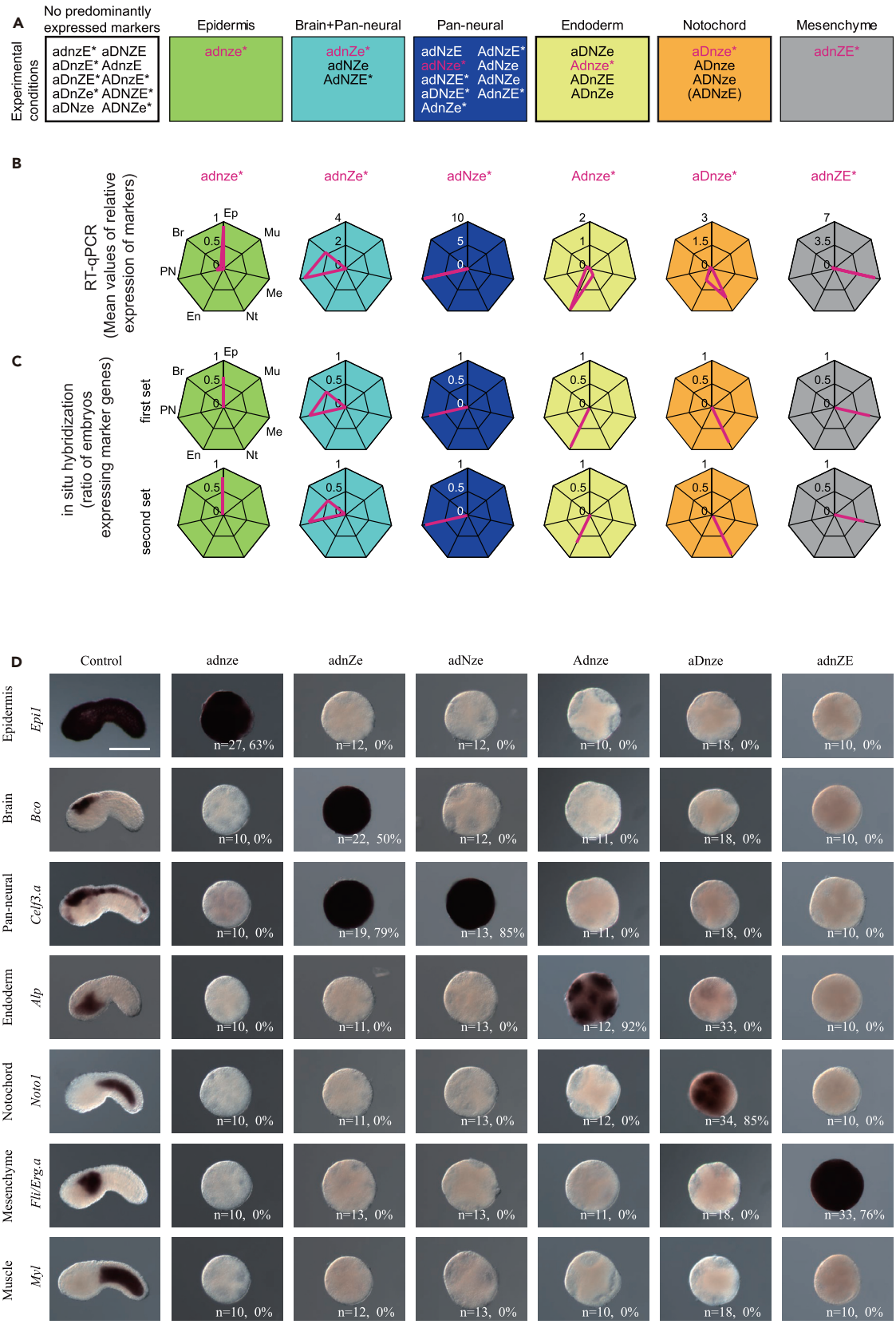

A
No predominantly expressed markers
Epidermis
Brain+Pan-neural
Pan-neural
Endoderm
Notochord
Mesenchyme
Experimental conditions
adnzE* aDNZE
aDnzE* AdnzE
aDnZE* ADnzE*
aDnZe* ADNZE*
aDNze ADNZe*
adnze*
adnZe*
adNZe
AdNZE*
adNzE AdNzE*
adNze* AdNze
adNZE* AdNZe
aDNzE* AdnZE*
AdnZe*
aDNZe
Adnze*
ADnZE
ADnZe
aDnze*
ADnze
ADNze
(ADNzE)
adnZE*
B
RT-qPCR
(Mean values of relative expression of markers)
adnze*
adnZe*
adNze*
Adnze*
aDnze*
adnZE*
Ep
Br
Mu
PN
Me
En
Nt
C
in situ hybridization
(ratio of embryos expressing marker genes)
first set
second set
D
Control
adnze
adnZe
adNze
Adnze
aDnze
adnZE
Epidermis
Epi1
Brain
Bco
Pan-neural
Celf3.a
Endoderm
Alp
Notochord
Noto1
Mesenchyme
Fli/Erg.a
Muscle
Myl
n=27, 63%
n=12, 0%
n=12, 0%
n=10, 0%
n=18, 0%
n=10, 0%
n=10, 0%
n=22, 50%
n=12, 0%
n=11, 0%
n=18, 0%
n=10, 0%
n=10, 0%
n=19, 79%
n=13, 85%
n=11, 0%
n=18, 0%
n=10, 0%
n=10, 0%
n=11, 0%
n=13, 0%
n=12, 92%
n=33, 0%
n=10, 0%
n=10, 0%
n=11, 0%
n=13, 0%
n=12, 0%
n=34, 85%
n=10, 0%
n=10, 0%
n=13, 0%
n=13, 0%
n=11, 0%
n=18, 0%
n=33, 76%
n=10, 0%
n=12, 0%
n=13, 0%
n=10, 0%
n=18, 0%
n=10, 0%

**Figure 4. Expression of Marker Genes by Manipulation of the Activities of the FVS Factors**

(A–C) Marker expression in experimental embryos. Conditions with asterisks were examined in multiple batches. Each of the experimental conditions is represented by a five-letter code in which up- and down-regulation of *Foxa.a*, *Foxd*, *Neurog*, *Zic-r.b*, and Erk signaling are represented by A/a, D/d, N/n, Z/z, and E/e, respectively. (A) Markers predominantly expressed in 32 experimental conditions. See also Figures S3–S7.

(B and C) The RT-qPCR results for six representative conditions shown in magenta are presented in (B) and were further examined by *in situ* hybridization as shown in (C). The axes of the first graph are labeled: Ep, epidermal marker; Br, brain marker; PN, pan-neural marker; En, endodermal marker; Nt, notochord marker; Me, mesenchyme marker; Mu, muscle marker. This configuration is applied to the other graphs. In (C), the results for the original set (upper; photographs are shown in D) and an additional set of markers (lower; photographs are shown in Figure S8) are shown.

(D) *In situ* hybridization of the first set of marker genes shown in (C). Gene names are shown on the left. The numbers of embryos we examined and the percentage of embryos that expressed the markers are shown in each panel. Scale bar, 100 μm. See also Figure S8.

expressed *Epi1*, a set of *Bco/Celf3.a*, or *Celf3.a* (Figure 5A; Table 1), as gene expression patterns differed in embryos without manipulation of the activities of the FVS factors (Figures 3B and 3C). Similarly, overexpression of *Zic-r.b* alone did not determine cell fate uniquely, either; namely, such embryos expressed *Fli/Erg.a*, a set of *Bco/Celf3.a*, or both (Figure 5B; Table 1). These observations were consistent with a proposition of the linkage logic theory, namely, that manipulation of the whole FVS is necessary to fully control network dynamics.

## Gene Expression Profiles of Induced Notochord and Mesenchyme

Finally, we compared the genome-wide expression profiles of embryos in two conditions (aDnze and adnZE), in which notochord and mesenchyme markers were predominantly expressed, with those of notochord and mesenchyme cells (Figure 6). For this purpose, we isolated two pairs of presumptive notochord cells and two pairs of presumptive mesenchyme cells because these isolated blastomeres differentiate into notochord and mesenchyme autonomously as partial embryos (Kim and Nishida, 1999; Nakatani and Nishida, 1994). Gene expression profiles for these four types of embryo were analyzed by RNA sequencing (RNA-seq) (Figure 6A). There were 929 genes with significantly different expression levels between notochord and mesenchyme partial embryos (NOIseq p value adjusted for multiple testing <0.001). Among them, the expression levels of 280 genes were higher in notochord than in mesenchyme (notochord partial embryo [N-PE]-enriched genes), and those of 649 genes were higher in mesenchyme (mesenchyme partial embryo [M-PE]-enriched genes). We also compared the gene expression profiles between aDnze and adnZE embryos and similarly identified aDnze-enriched genes and adnZE-enriched genes. Among 280 notochord-enriched genes, 71 genes were commonly found in aDnze-enriched genes, whereas only 1 gene was commonly found in adnZE-enriched genes. On the other hand, among 649 mesenchyme-enriched genes, 17 genes were found in aDnze-enriched genes and 163 genes were found in adnZE-enriched genes (Figure 6B). Namely, sets of genes expressed in notochord and mesenchyme partial embryos are more similar to those in aDnze and adnZE embryos, respectively.

Figure 6C indicates that the 71 genes commonly enriched in aDnze and notochord partial embryos show quantitatively similar expression levels, as do the 163 genes commonly enriched in adnZE and mesenchyme partial embryos. Indeed, the correlation coefficients were 0.652 and 0.958 upon excluding one outlier in the notochord genes (see Discussion). Thus, the expression levels of the above-mentioned specific genes were also highly reproduced in adnZE embryos and moderately in aDnze embryos.

## DISCUSSION

We developed a method to control nonlinear dynamic systems based on FVS, which are identified from the structure of networks. We confirmed that the dynamics of the GRN for fate specification in early *Ciona* embryo is controllable by manipulating the activities of FVS factors. The expression patterns that represent six of seven cell states observed in the embryo were actually induced in a deterministic manner. The results are consistent with the expected dynamic behavior of the multipotency of the system.

We performed our experiments to examine all possible manipulations in a binary control. The obtained marker gene expression patterns strongly suggest that six tissues were differentiated under at least one condition of the binary manipulations. This may indicate that qualitative regulation but not quantitative regulation is sufficient for fate specification of these six tissues in ascidian embryos. The RNA-seq experiments showed that adnZE embryos and mesenchyme partial embryos express specific genes at quantitatively similar levels. In contrast, although aDnze embryos and notochord partial embryos commonly expressed 71 genes specifically, their expression levels were not so well reproduced. This

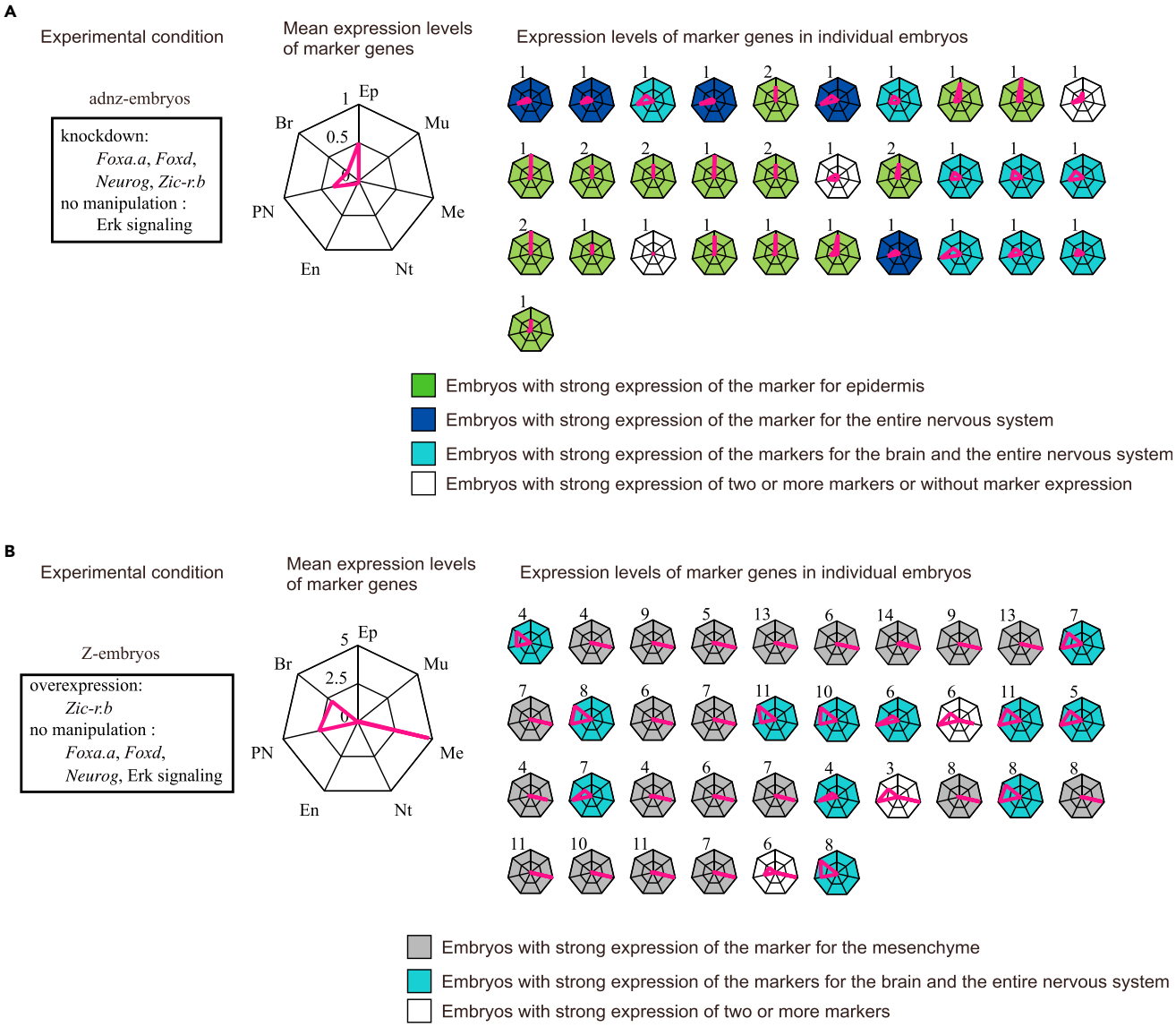



**Figure 5. Dynamics of the Network for Cell Fate Specification Is Uncontrollable by Manipulating the Activities of a Subset of the FVS Factors**

(A and B) Marker gene expression determined by RT-qPCR (A) in adnz-embryos and (B) in Z-embryos. Mean values (left large graphs) and all values for individual embryos (right small graphs) are shown. The axes of the first graph are labeled: Ep, epidermal marker; Br, brain marker; PN, pan-neural marker; En, endodermal marker; Nt, notochord marker; Me, mesenchyme marker; Mu, muscle marker. This configuration is applied to the other graphs. Different colors in the small graphs indicate that different tissue markers are predominantly expressed.

might be explained by the difference in conditions between natural dynamics and the artificial fixation of FVSs. Under the continuous fixation of an FVS, the expression of some genes may differ from natural conditions. Indeed, the expression level of *Brachyury*, which is a key regulatory gene for notochord differentiation (Takahashi et al., 1999) and an outlier shown in Figure 6C, was markedly higher in aDnze (reads per kilobase of transcript per million mapped reads [RPKM] = 11,897) than in notochord partial embryos (RPKM = 898).

We could not induce marker gene expression corresponding to muscle in any of the conditions that we examined. Quantitative manipulation of the FVSs may be required to induce the expression of muscle markers. Another possibility is that the GRN that we used in this study did not include factors (nodes or edges) that are responsible for the specification of muscle fate. Indeed, a previous study (Nishida and Sawada, 2001) showed that a localized maternal factor plays an important role in the specification of muscle

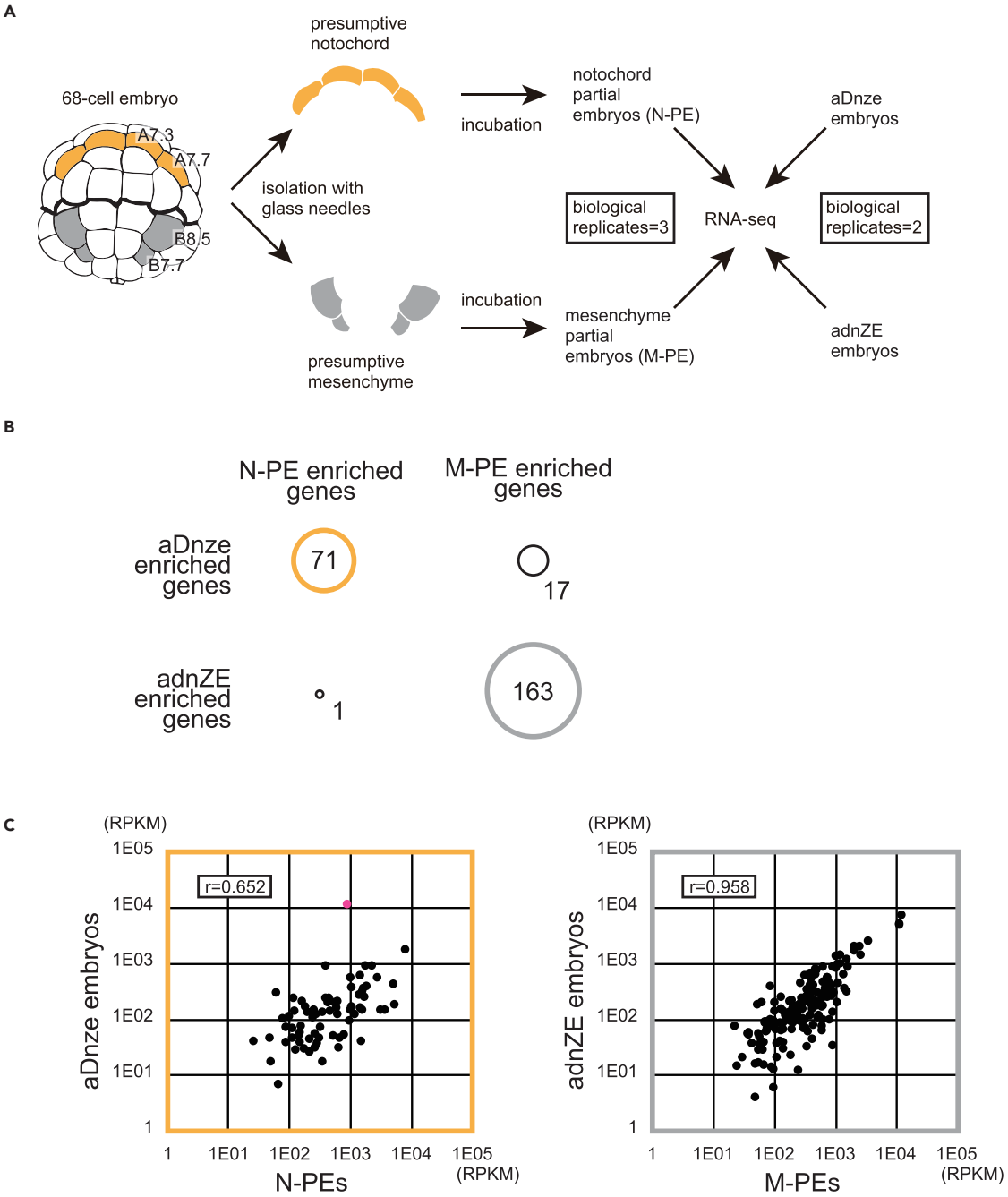



**Figure 6. Analysis of Expression Profiles for aDnze and adnZE Embryos**

(A) The experimental design for RNA sequencing of partial embryos of notochord and mesenchyme (N-PE and M-PE) and of aDnze and adnZE embryos.

(B) Comparisons between gene fractions enriched in N-PE and M-PE and those enriched in aDnze and adnZE embryos.

(C) Scatterplots showing expression levels of the 71 genes found commonly between gene fractions enriched in N-PE and aDnze embryos and those of the 163 genes found commonly between gene fractions enriched in M-PE and adnZE embryos.

fate. If such factors take more dominant roles than the GRN, controlling GRN alone would not be sufficient to induce muscle fate.

Differentiated tissues are generally thought to be at steady states of dynamics of gene activities, and they may be established at the tailbud stage in *Ciona*. However, the GRN analyzed in this study includes genes that are

not expressed in such a late stage. Although we analyzed possible steady states of the GRN up to the late gastrula stage by fixing activities of the FVS factors, it might be difficult to observe these steady states in the actual development of *Ciona*. Indeed, expression of the FVS factors except *Foxa.a* is transient in *Ciona* embryos. However, the results of our analysis imply that artificially induced steady states of the GRN are sufficient to specify cell fates at a later stage. One possible explanation for this is as follows. In normal development, the dynamics of gene activities may fall into a steady state of the GRN by the late gastrula stage and thereafter genes downstream of the GRN may suppress the expression of the FVS factors. If this is the case, steady states for fate decision, which possibly exist in the GRN, are transient and therefore become undetectable by integrating regulation at a later stage. Our result suggests that it is practical to decompose a GRN into subnetworks and to study steady states of the subnetworks to understand cell specification processes.

In normal embryos, the GRN governs specific gene expression temporally and spatially. The important function of the GRN for specification of cell fates may be to create specific expression patterns of the FVS genes, which activate cell-type-specific downstream pathways, because a specific combination of the activities of the FVS factors determined a specific cell fate (Figure 4). Gene expression patterns of *Foxa.a*, *Foxd*, *Neurog*, and *Zic-r.b* and temporal and spatial patterns of the activity of the Erk pathway are mostly consistent with the above-mentioned speculation (Haupaix et al., 2013; Hudson et al., 2003; Imai et al., 2002a, 2002b, 2004; Ohta and Satou, 2013; Picco et al., 2007; Shi and Levine, 2008; Shimauchi et al., 2001). The epidermal markers are expressed in adnze embryos. In normal embryos, *Foxd*, *Neurog*, and *Zic-r.b* are not expressed in the epidermal lineage, and *Foxa.a* is expressed transiently only in early embryos. In addition, the Erk pathway is not turned on in this lineage before gastrulation. Mesenchyme markers were expressed in adnZE embryos. In the mesenchyme lineage of normal embryos, *Zic-r.b* is strongly expressed and the Erk pathway is activated, whereas *Foxa.a* and *Foxd* are expressed transiently only in early embryos. The notochord maker was expressed in aDnze embryos. In the notochord lineage, *Foxd, Foxa.a*, and *Zic-r.b* are expressed and the Erk pathway is activated. Although this pattern does not fully support the aforementioned hypothesis, expression patterns of Foxd, Foxa.a, and Zic-r.b proteins are not known; if Foxd is not degraded for a long time, it is possible that conditions that lead to notochord differentiation appear in the notochord lineage of normal embryos.

The structural theory provides strong predictions directly from the structure of the network without assuming other quantitative details of dynamics. Although another theory has been proposed that gives criteria to choose driver nodes in linear systems structurally (Lin, 1974; Liu et al., 2011), linkage logic is the first theory to determine key nodes for controlling nonlinear systems only from the structure of networks. Theoretically, the strategy is applicable to any nonlinear dynamic system, which includes networks other than GRNs (Fiedler et al., 2013; Mochizuki et al., 2013; Zanudo et al., 2017), and is particularly useful for controlling or engineering complex biological networks.

## METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, eight figures, and five tables and can be found with this article online at https://doi.org/10.1016/j.isci.2018.05.004.

## ACKNOWLEDGMENTS

We appreciate the critical reading of our manuscript by Hiroshi Hamada (RIKEN Center for Developmental Biology) and Yoh Iwasa (Kyushu University). We thank individuals working in Satou's laboratory and Misaki marine station of the University of Tokyo under the National Bio-Resource Project for providing experimental animals. This research was supported by the CREST program (grant no. JPMJCR13W6) of the Japan Science and Technology Agency (JST) (http://www.jst.go.jp/EN/index.html), the RIKEN iTHES Project, and RIKEN iTHEMS Program. We thank Zenis Co., Ltd. (https://www.zenis.co.jp/) and Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

## AUTHOR CONTRIBUTIONS

A.M. and Y.S. conceived and supervised the research and wrote the paper. K.M. performed mathematical analyses. K.K. and M.T. performed experiments.

## DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: March 3, 2018
Revised: April 19, 2018
Accepted: May 7, 2018
Published: June 07, 2018

## REFERENCES

Akutsu, T., Kuhara, S., Maruyama, O., and Miyano, S. (1998). A system for identifying genetic networks from gene expression patterns produced by gene disruptions and overexpressions. Genome Inform. Ser Workshop Genome Inform. *9*, 151–160.

Alon, U. (2007). Network motifs: theory and experimental approaches. Nat. Rev. Genet. *8*, 450–461.

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

Fiedler, B., Mochizuki, A., Kurosawa, G., and Saito, D. (2013). Dynamics and control at feedback vertex sets. I: informative and determining nodes in regulatory networks. J. Dynam. Differ. Equat. *25*, 563–604.

Haupaix, N., Stolfi, A., Sirour, C., Picco, V., Levine, M., Christiaen, L., and Yasuo, H. (2013). p120RasGAP mediates ephrin/Eph-dependent attenuation of FGF/ERK signals during cell fate specification in ascidian embryos. Development *140*, 4347–4352.

Hotta, K., Takahashi, H., Erives, A., Levine, M., and Satoh, N. (1999). Temporal expression patterns of 39 Brachyury-downstream genes associated with notochord formation in the *Ciona intestinalis* embryo. Dev. Growth Differ. *41*, 657–664.

Hudson, C., Darras, S., Caillol, D., Yasuo, H., and Lemaire, P. (2003). A conserved role for the MEK signalling pathway in neural tissue specification and posteriorisation in the invertebrate chordate, the ascidian *Ciona intestinalis*. Development *130*, 147–159.

Hudson, C., Sirour, C., and Yasuo, H. (2016). Co-expression of Foxa.a, Foxd and Fgf9/16/20 defines a transient mesendoderm regulatory state in ascidian embryos. Elife *5*, e14692.

Hudson, C., and Yasuo, H. (2006). A signalling relay involving Nodal and Delta ligands acts during secondary notochord induction in *Ciona* embryos. Development *133*, 2855–2864.

Imai, K., Satoh, N., and Satou, Y. (2003). A twist-like bHLH gene is a downstream factor of an endogenous FGF and determines mesenchymal fate in the ascidian embryos. Development *130*, 4461–4472.

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

Imai, K.S., Hino, K., Yagi, K., Satoh, N., and Satou, Y. (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. Development *131*, 4047–4058.

Imai, K.S., Levine, M., Satoh, N., and Satou, Y. (2006). Regulatory blueprint for a chordate embryo. Science *312*, 1183–1187.

Imai, K.S., Satoh, N., and Satou, Y. (2002a). An essential role of a FoxD gene in notochord induction in *Ciona* embryos. Development *129*, 3441–3453.

Imai, K.S., Satou, Y., and Satoh, N. (2002b). Multiple functions of a Zic-like gene in the differentiation of notochord, central nervous system and muscle in *Ciona savignyi* embryos. Development *129*, 2723–2732.

Jeffery, W.R., Chiba, T., Krajka, F.R., Deyts, C., Satoh, N., and Joly, J.S. (2008). Trunk lateral cells are neural crest-like cells in the ascidian *Ciona intestinalis*: insights into the ancestry and evolution of the neural crest. Dev. Biol. *324*, 152–160.

Kim, G.J., and Nishida, H. (1999). Suppression of muscle fate by cellular interaction is required for mesenchyme formation during ascidian embryogenesis. Dev. Biol. *214*, 9–22.

Kodama, H., Miyata, Y., Kuwajima, M., Izuchi, R., Kobayashi, A., Gyoja, F., Onuma, T.A., Kumano, G., and Nishida, H. (2016). Redundant mechanisms are involved in suppression of default cell fates during embryonic mesenchyme and notochord induction in ascidians. Dev. Biol. *416*, 162–172.

Kusakabe, T., Yoshida, R., Kawakami, I., Kusakabe, R., Mochizuki, Y., Yamada, L., Shin-i, T., Kohara, Y., Satoh, N., Tsuda, M., et al. (2002). Gene expression profiles in tadpole larvae of *Ciona intestinalis*. Dev. Biol. *242*, 188–203.

Lin, C.T. (1974). Structural controllability. IEEE Trans. Automat. Contr. *Ac19*, 201–208.

Liu, Y.Y., Slotine, J.J., and Barabasi, A.L. (2011). Controllability of complex networks. Nature *473*, 167–173.

Meedel, T.H., Chang, P., and Yasuo, H. (2007). Muscle development in *Ciona intestinalis* requires the b-HLH myogenic regulatory factor gene Ci-MRF. Dev. Biol. *302*, 333–344.

Mochizuki, A. (2008). Structure of regulatory networks and diversity of gene expression patterns. J. Theor. Biol. *250*, 307–321.

Mochizuki, A., Fiedler, B., Kurosawa, G., and Saito, D. (2013). Dynamics and control at feedback vertex sets. II: a faithful monitor to determine the diversity of molecular activities in regulatory networks. J. Theor. Biol. *335*, 130–146.

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

Nishida, H., and Sawada, K. (2001). macho-1 encodes a localized mRNA in ascidian eggs that specifies muscle fate during embryogenesis. Nature *409*, 724–729.

Oda-Ishii, I., and Di Gregorio, A. (2007). Lineage-independent mosaic expression and regulation of the *Ciona multidom* gene in the ancestral notochord. Dev. Dyn. *236*, 1806–1819.

Oda, K., Matsuoka, Y., Funahashi, A., and Kitano, H. (2005). A comprehensive pathway map of epidermal growth factor receptor signaling. Mol. Syst. Biol. *1*, 2005.0010.

Ohta, N., and Satou, Y. (2013). Multiple signaling pathways coordinate to induce a threshold response in a chordate embryo. PLoS Genet. *9*, e1003818.

Peter, I.S., and Davidson, E.H. (2016). Implications of developmental gene regulatory networks inside and outside developmental biology. Curr. Top. Dev. Biol. *117*, 237–251.

Picco, V., Hudson, C., and Yasuo, H. (2007). Ephrin-Eph signalling drives the asymmetric division of notochord/neural precursors in *Ciona* embryos. Development *134*, 1491–1497.

Satoh, N. (1979). On the 'clock' mechanism determining the time of tissue-specific enzyme development during ascidian embryogenesis. I. Acetylcholinesterase development in cleavage-arrested embryos. J. Embryol. Exp. Morphol. *54*, 131–139.

Satou, Y., and Imai, K.S. (2015). Gene regulatory systems that control gene expression in the *Ciona* embryo. Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. *91*, 33–51.

Satou, Y., Takatori, N., Yamada, L., Mochizuki, Y., Hamaguchi, M., Ishikawa, H., Chiba, S., Imai, K., Kano, S., Murakami, S.D., et al. (2001). Gene expression profiles in *Ciona* intestinalis tailbud embryos. Development *128*, 2893–2904.

Shi, W., and Levine, M. (2008). Ephrin signaling establishes asymmetric cell fates in an endomesoderm lineage of the *Ciona* embryo. Development *135*, 931–940.

Shimauchi, Y., Chiba, S., and Satoh, N. (2001). Synergistic action of HNF-3 and Brachyury in the notochord differentiation of ascidian embryos. Int. J. Dev. Biol. *45*, 643–652.

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

Tokuoka, M., Imai, K., Satou, Y., and Satoh, N. (2004). Three distinct lineages of mesenchymal cells in *Ciona intestinalis* embryos demonstrated by specific gene expression. Dev. Biol. *274*, 211–224.

Ueki, T., Yoshida, S., Marikawa, Y., and Satoh, N. (1994). Autonomy of expression of epidermis-specific genes in the ascidian embryo. Dev. Biol. *164*, 207–218.

Whittaker, J.R. (1973). Segregation during ascidian embryogenesis of egg cytoplasmic information for tissue-specific enzyme development. Proc. Natl. Acad. Sci. USA *70*, 2096–2100.

Yagi, K., and Makabe, K.W. (2001). Isolation of an early neural maker gene abundantly expressed in the nervous system of the ascidian, *Halocynthia roretzi*. Dev. Genes Evol. *211*, 49–53.

Yasuo, H., and Hudson, C. (2007). FGF8/17/18 functions together with FGF9/16/20 during formation. of the notochord in *Ciona* embryos. Dev. Biol. *302*, 92–103.

Zanudo, J.G.T., Yang, G., and Albert, R. (2017). Structure-based control of complex networks with nonlinear dynamics. Proc. Natl. Acad. Sci. USA *114*, 7234–7239.

**ISCI, Volume *4***

## Supplemental Information

# Controlling Cell Fate Specification System by Key Genes Determined from Network Structure

**Kenji Kobayashi, Kazuki Maeda, Miki Tokuoka, Atsushi Mochizuki, and Yutaka Satou**

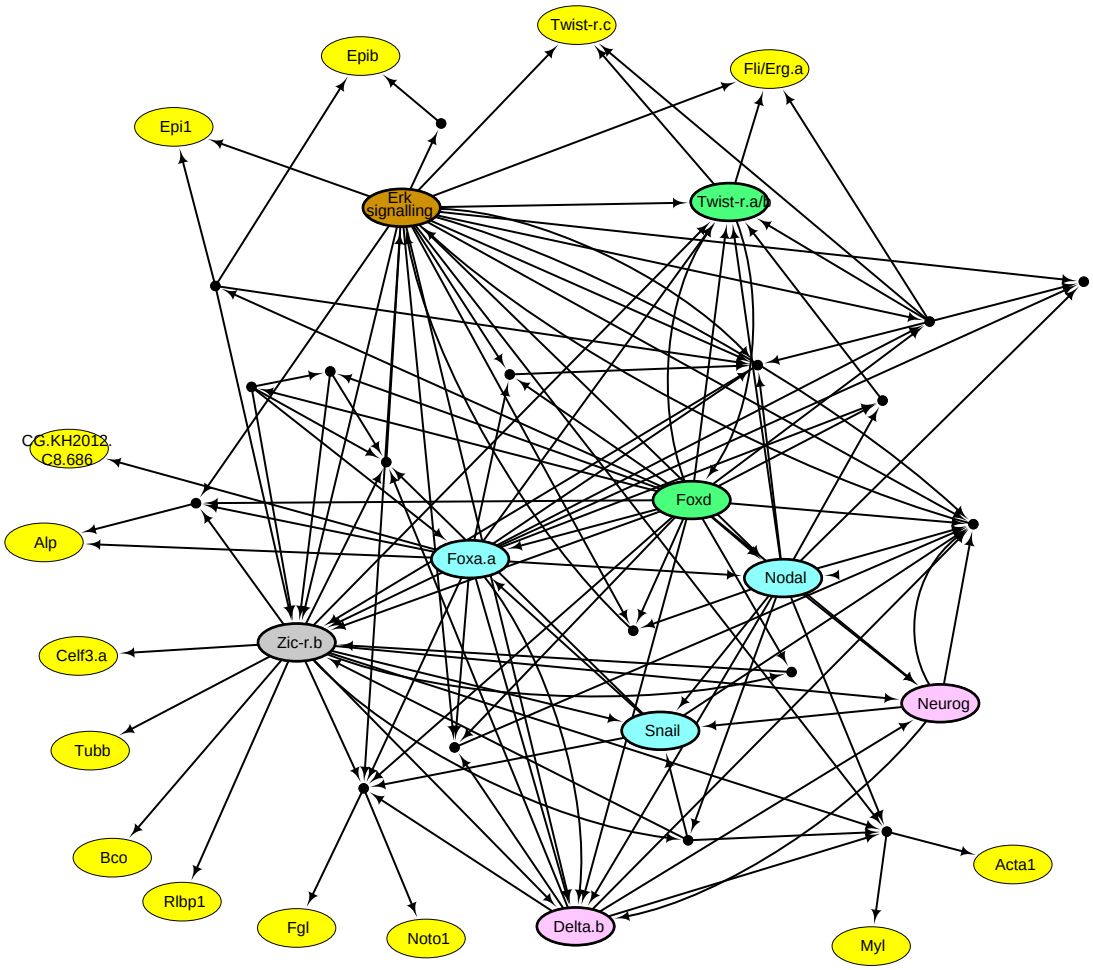



**Figure S1. A simplified version of the gene regulatory network for cell specification in *Ciona*. Related to Figure 2.** Only factors that are involved in directed cycles or that are directly involved in regulation of markers are shown. The FVS factors are shown by ovals colored by light blue, green, pink, grey and orange. Nodes filled in yellow are the marker genes, for which we performed observations of activities. The remaining factors are shown by black dots.

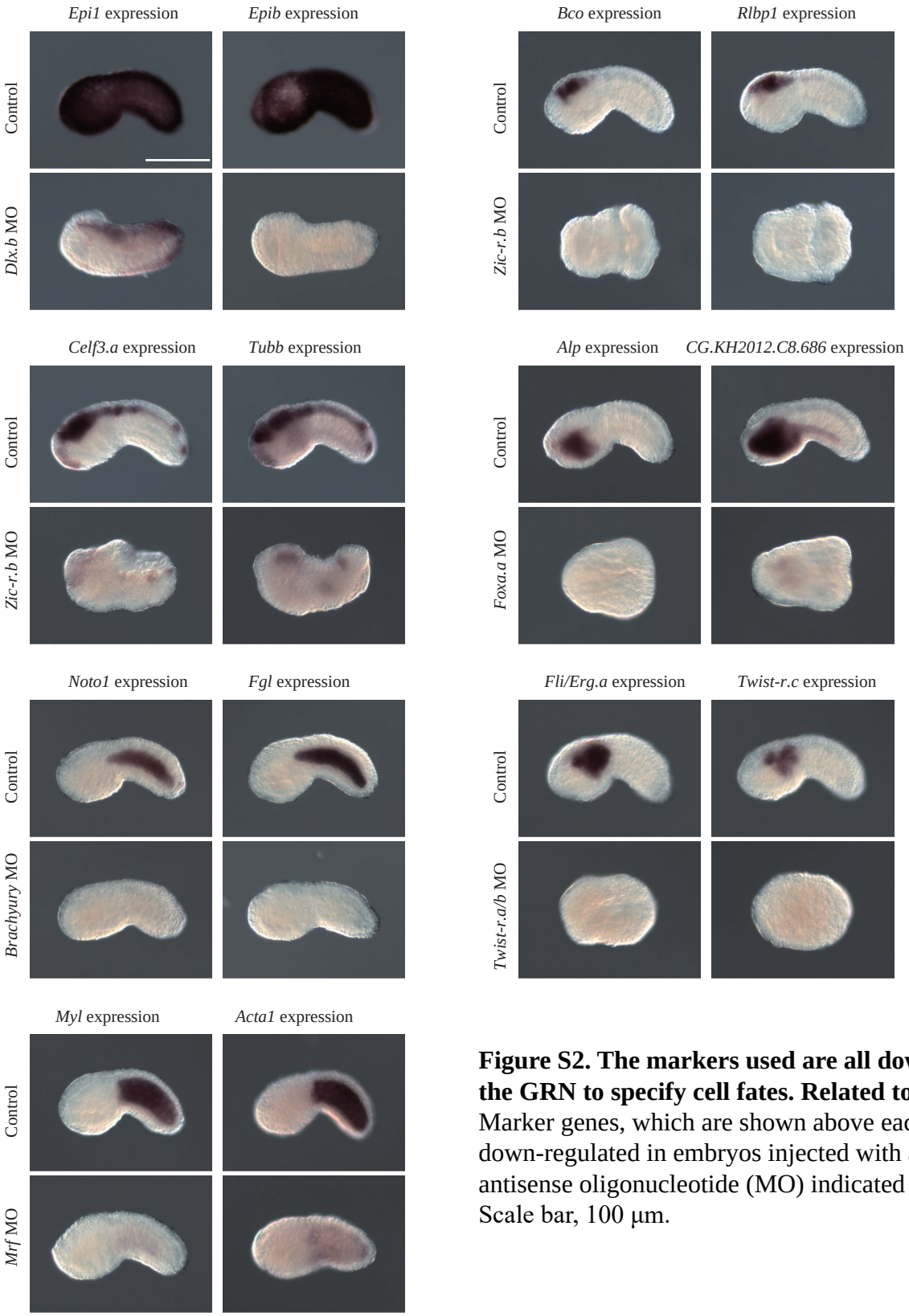



**Figure S2. The markers used are all downstream of the GRN to specify cell fates. Related to Figure 2.** Marker genes, which are shown above each panel, are down-regulated in embryos injected with a morpholino antisense oligonucleotide (MO) indicated on the left. Scale bar, 100 μm.

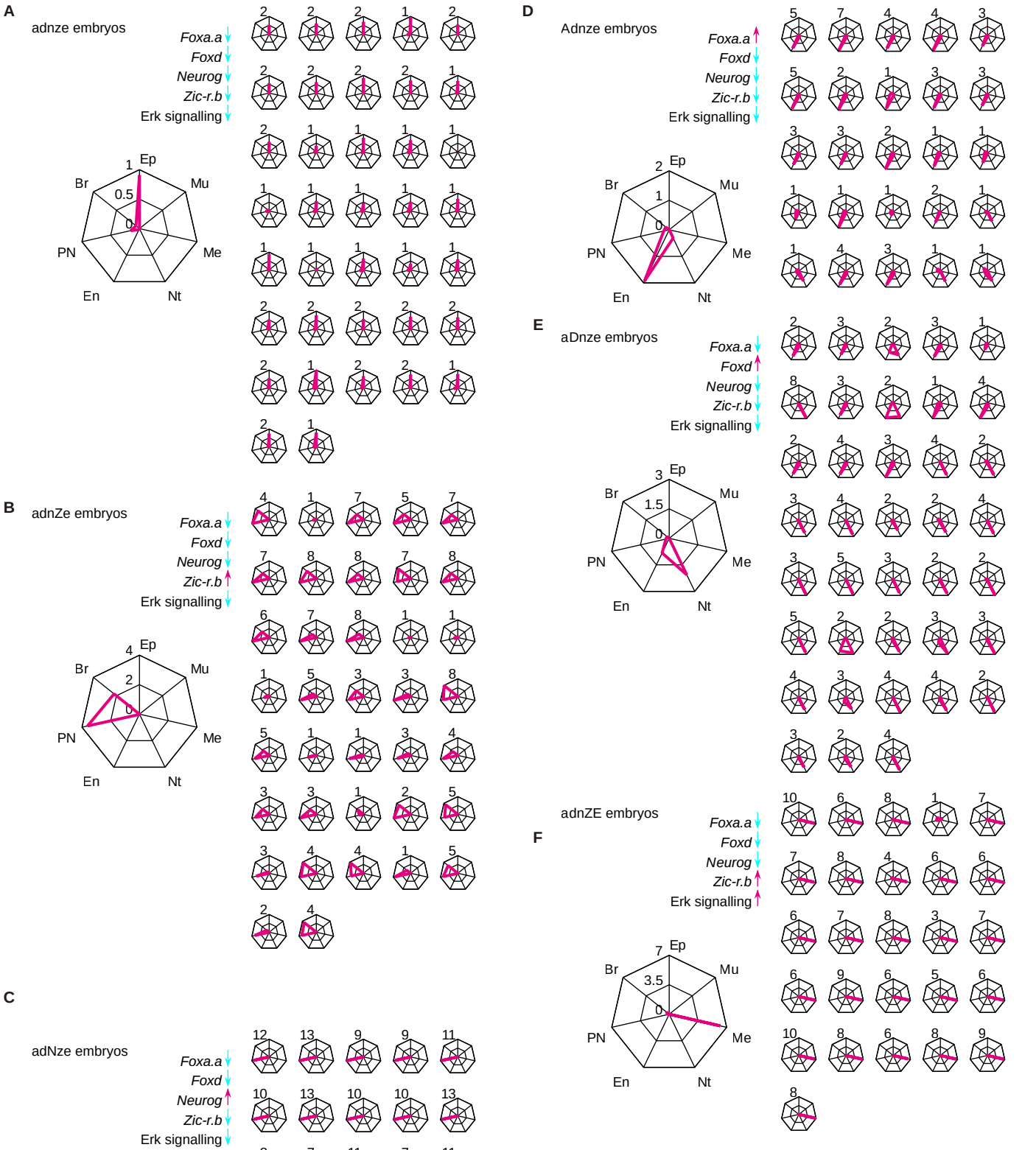
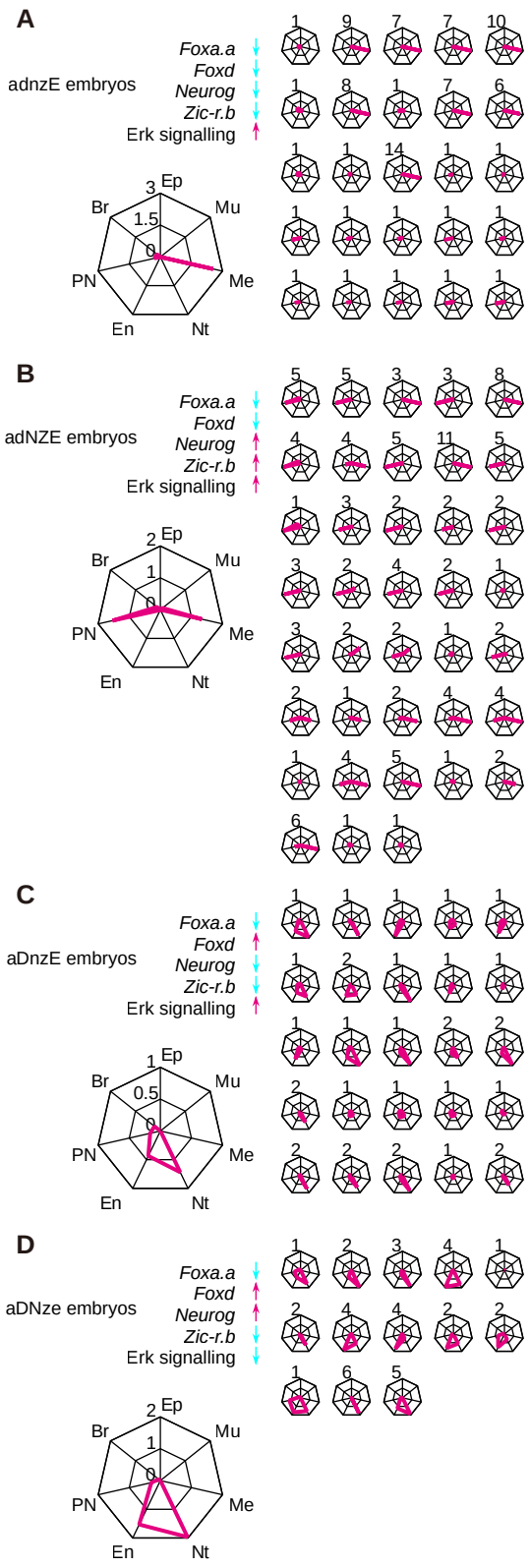



**Figure S3. Marker gene expression in adnze, adnZe, adNze, Adnze, aDnze and adnZE embryos. Related to Figure 4.** The large graphs on the left of each panel show the average expression levels of tissue markers, as determined by RT-qPCR. Small graphs on the right of each panel show expression levels in individual embryos. Expression levels of marker genes are shown relative to the average expression levels in normal 9.5 hr (tailbud-stage) embryos.



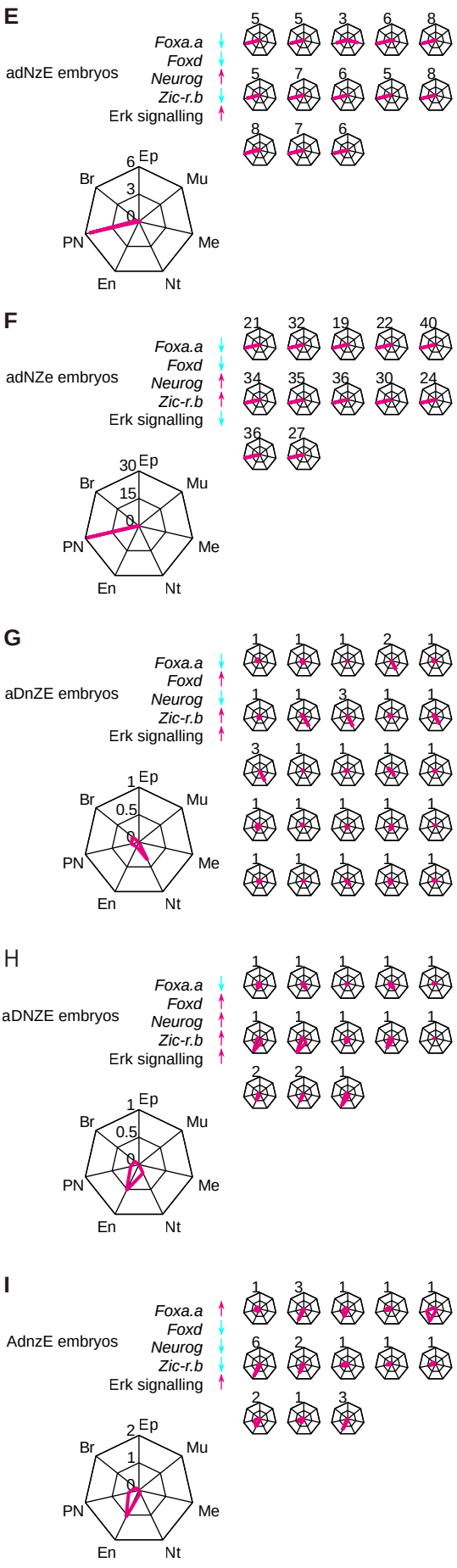
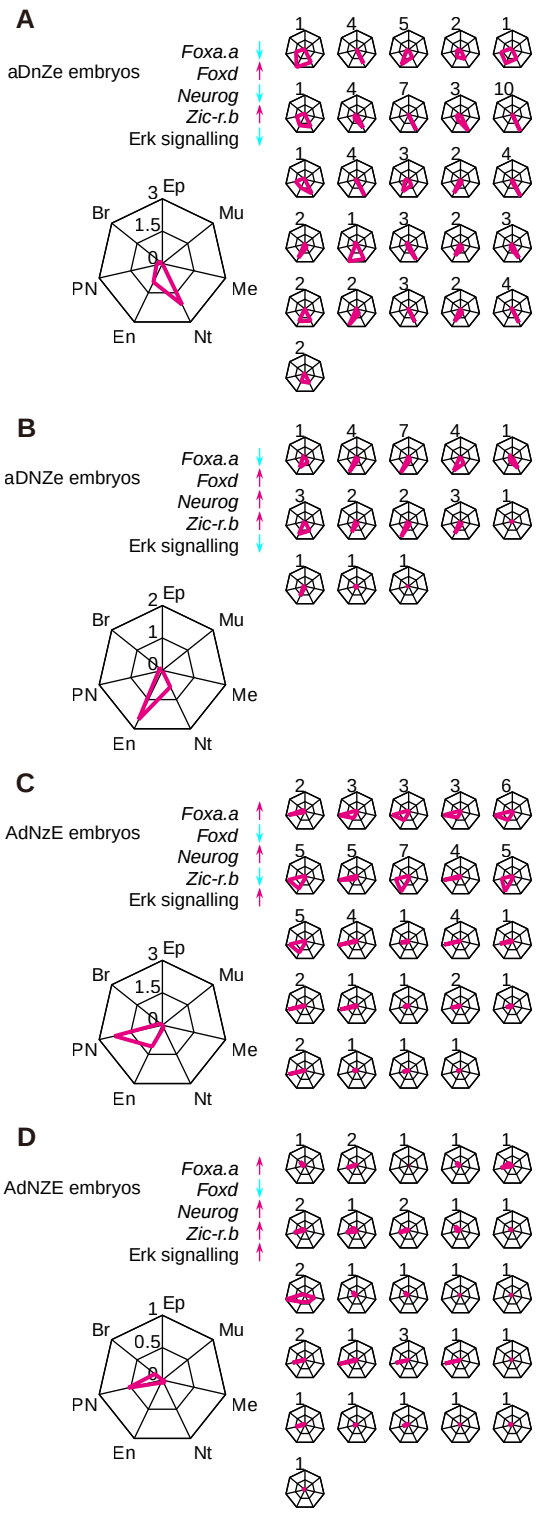




**Figure S4. Marker gene expression in adnzE, adNZE, aDnzE, aDNze, adNzE, adNZe, aDnZE, aDNZE and AdnzE embryos. Related to Figure 4.** The large graphs on the left of each panel show the average expression levels of tissue markers, as determined by RT-qPCR. Small graphs on the right of each panel show the expression levels in individual embryos. Expression levels of marker genes are shown relative to the average expression levels in normal 9.5 hr (tailbud-stage) embryos.



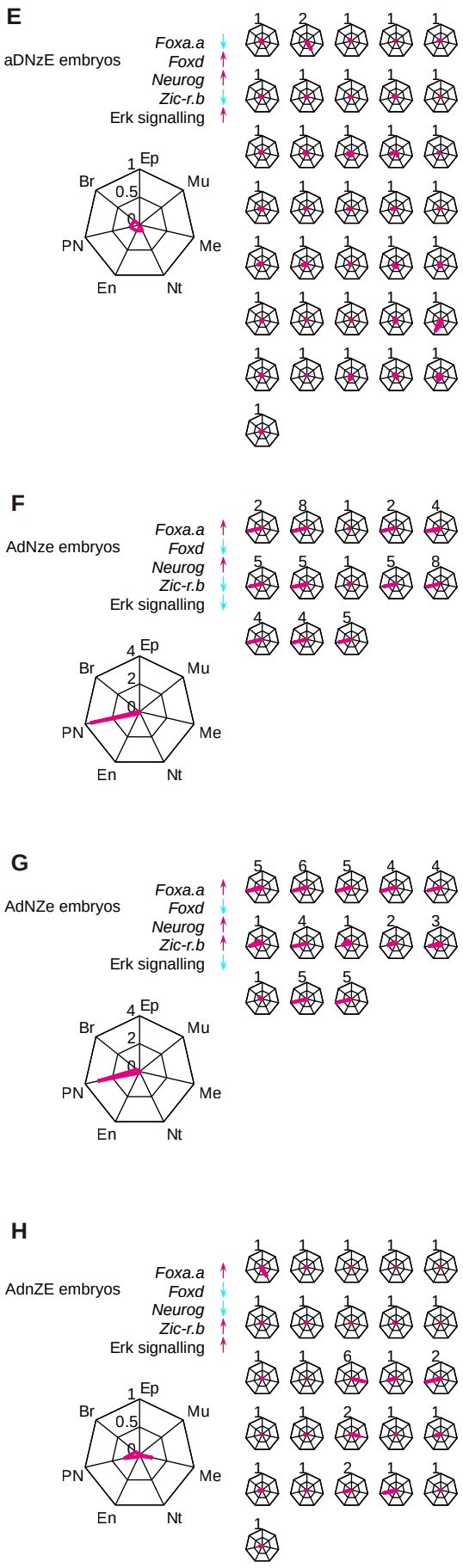




**Figure S5. Marker gene expression in aDnZe, aDNZe, AdNzE, AdNZE, aDNzE, AdNze, AdNZe and AdnZE embryos. Related to Figure 4.** The large graphs on the left of each panel show the average expression levels of tissue markers, as determined by RT-qPCR. Small graphs on the right of each panel show expression levels in individual embryos. Expression levels of marker genes are shown relative to the average expression levels in normal 9.5 hr (tailbud-stage) embryos.

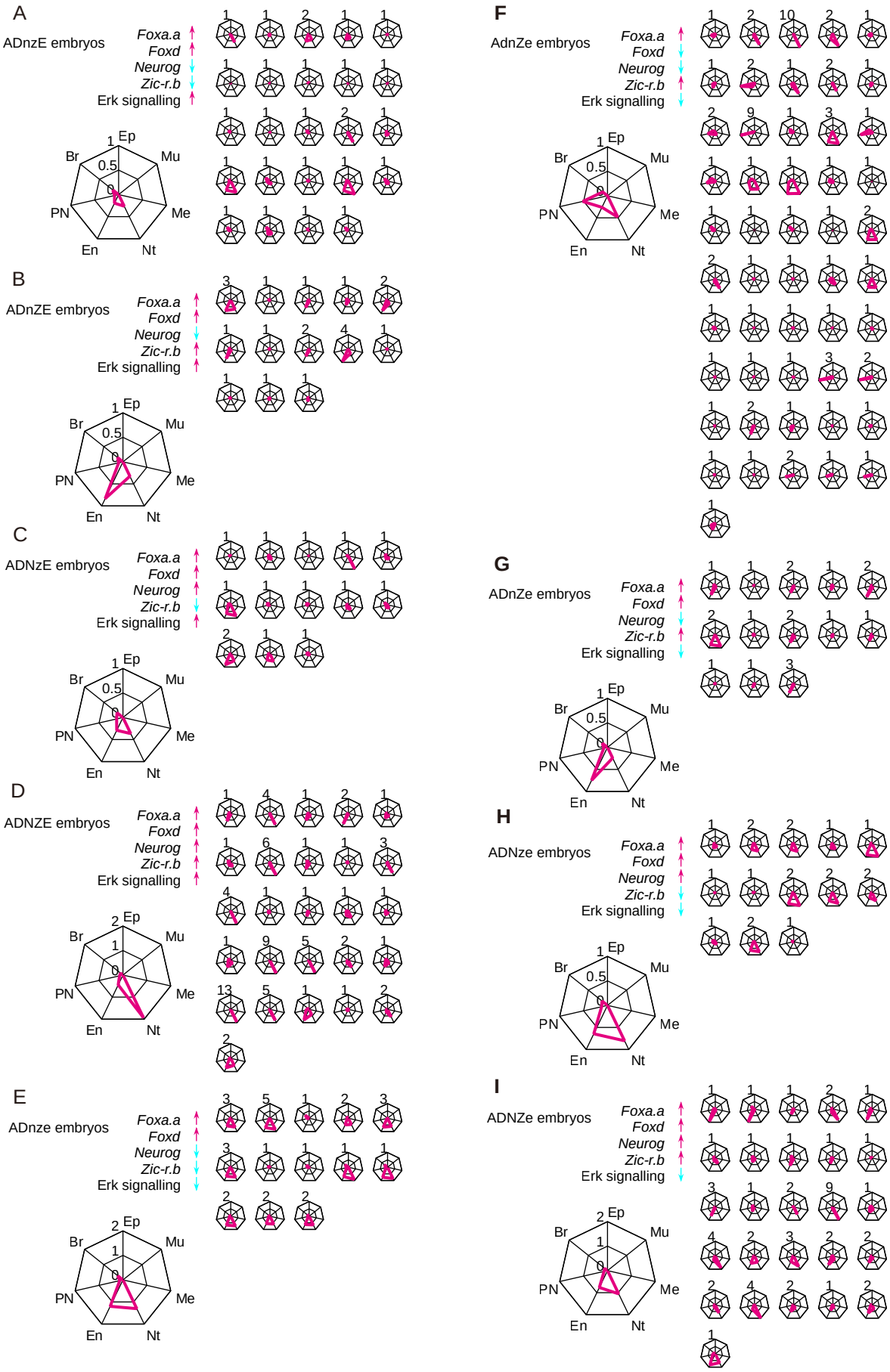



**Figure S6. Marker gene expression in ADnzE, ADnZE, ADNzE, ADNZE, ADnze, AdnZe, ADnZe, ADNze and ADNZe embryos. Related to Figure 4.** The large graphs on the left of each panel show the average expression levels of tissue markers, as determined by RT-qPCR. Small graphs on the right of each panel show expression levels in individual embryos. Expression levels of marker genes are shown relative to the average expression levels in normal 9.5 hr (tailbud-stage) embryos.

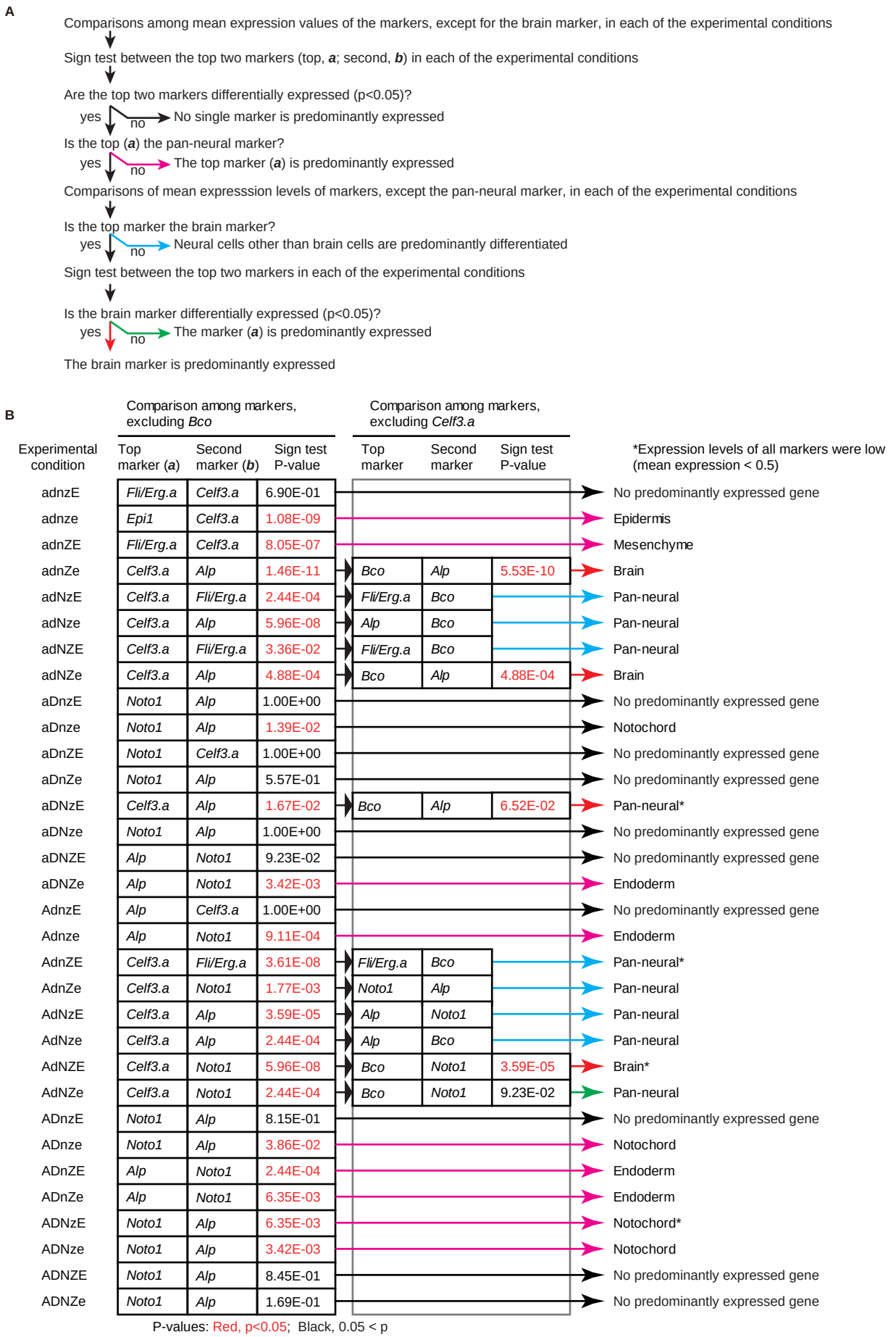



**Figure S7. Markers predominantly expressed in the 32 experimental conditions. Related to Figure 4.** **(A)** Flowchart for definition of the predominantly expressed gene in each condition. **(B)** Results of sign tests. Colours of arrows correspond to those in (A).

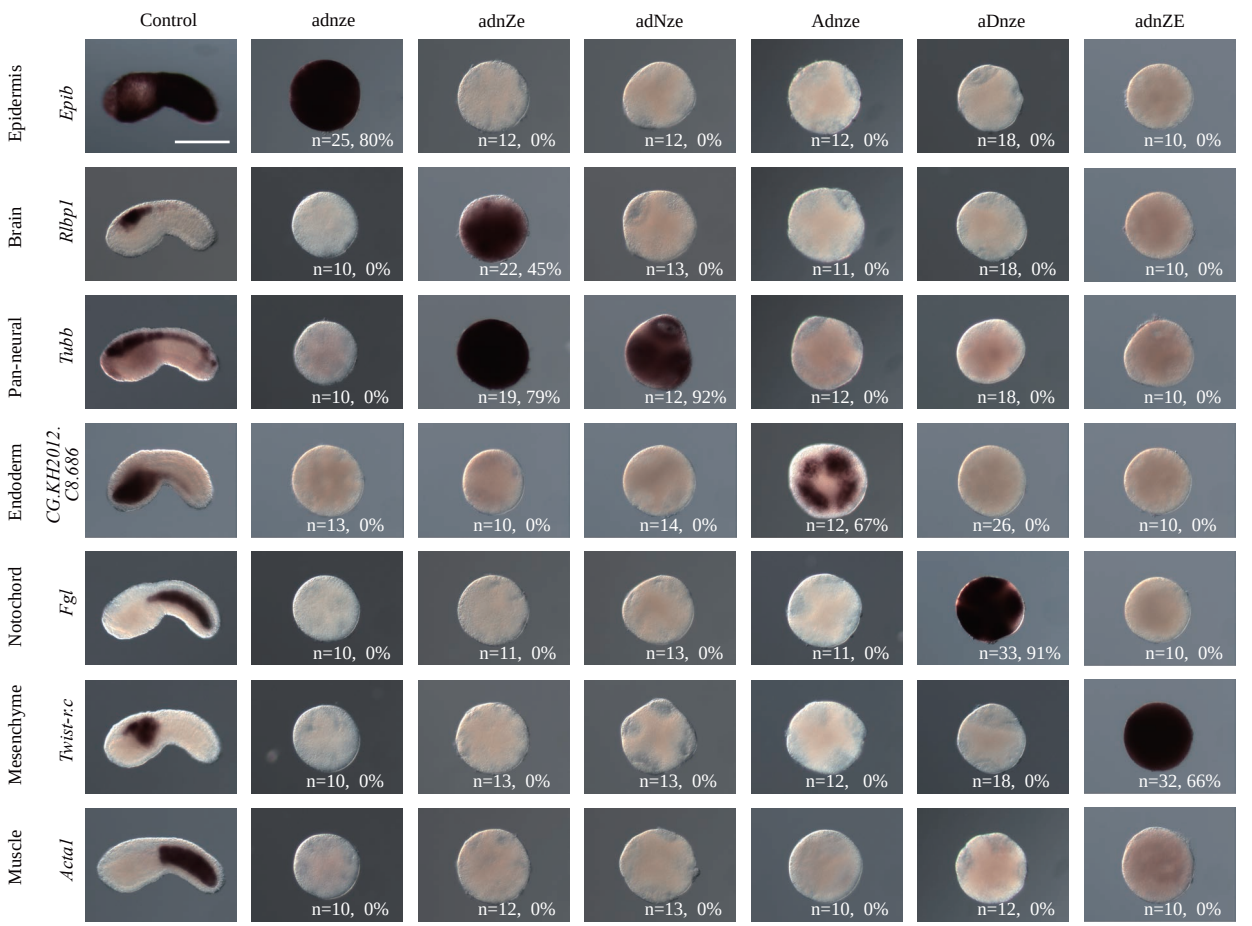



**Figure S8. Expression of marker genes in adnze, adnZE, adnZe, adNze, aDnze and Adnze embryos revealed by *in situ* hybridization. Related to Figure 4.** Photographs show the expression of the second set of tissue markers (shown on the left of each row). Numbers of embryos examined and percentages of embryos that expressed the specified marker are shown within each photograph. Scale bar, 100 μm.

**Table S1. Mean expression levels of marker genes. Related to Table 1.**

| | | Experimental Condition [a] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | unperturbed [b] | adnze [b] | adnZe [b] | adNze [b] | Adnze [b] | aDnze [b] | adnZE [b] | adnz [b] | Z [b] |
| Expression [c] | *Epi1* (epidermis) | 0.63 | 0.91 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.00 |
| | *Bco* (brain) | 2.30 | 0.08 | 2.20 | 0.06 | 0.15 | 0.14 | 0.43 | 0.18 | 2.16 |
| | *Celf3.a* (pan-neural) | 3.03 | 0.15 | 3.61 | 9.59 | 0.19 | 0.17 | 0.18 | 0.34 | 2.60 |
| | *Alp* (endoderm) | 0.01 | 0.02 | 0.03 | 0.13 | 1.94 | 0.83 | 0.01 | 0.04 | 0.03 |
| | *Noto1* (notochord) | 0.15 | 0.01 | 0.01 | 0.00 | 0.31 | 2.07 | 0.00 | 0.01 | 0.01 |
| | *Fli/Erg.a* (mesenchyme) | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 6.21 | 0.00 | 4.76 |
| | *Myl* (muscle) | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.08 |

| | | Experimental Condition [a] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | adnzE | adNZE | aDnzE | aDNze | adNzE | adNZe | aDnZE | aDNZE | AdnzE |
| Expression [c] | *Epi1* (epidermis) | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 |
| | *Bco* (brain) | 0.25 | 0.14 | 0.12 | 0.07 | 0.23 | 0.42 | 0.16 | 0.10 | 0.15 |
| | *Celf3.a* (pan-neural) | 0.25 | 1.54 | 0.16 | 0.29 | 5.55 | 29.05 | 0.14 | 0.15 | 0.38 |
| | *Alp* (endoderm) | 0.02 | 0.01 | 0.44 | 1.53 | 0.01 | 0.05 | 0.09 | 0.52 | 1.08 |
| | *Noto1* (notochord) | 0.00 | 0.00 | 0.72 | 1.97 | 0.00 | 0.00 | 0.36 | 0.17 | 0.11 |
| | *Fli/Erg.a* (mesenchyme) | 2.88 | 1.34 | 0.00 | 0.00 | 0.24 | 0.00 | 0.00 | 0.00 | 0.00 |
| | *Myl* (muscle) | 0.00 | 0.09 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 |

| | | Experimental Condition [a] | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | aDnZe | aDNZe | AdNzE | AdNZE | aDNzE | AdNze | AdNZe | AdnZE |
| Expression [c] | *Epi1* (epidermis) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | *Bco* (brain) | 0.16 | 0.11 | 0.12 | 0.15 | 0.12 | 0.03 | 0.25 | 0.09 |
| | *Celf3.a* (pan-neural) | 0.28 | 0.12 | 2.27 | 0.54 | 0.18 | 3.63 | 3.08 | 0.29 |
| | *Alp* (endoderm) | 0.95 | 1.69 | 1.10 | 0.05 | 0.12 | 0.15 | 0.11 | 0.04 |
| | *Noto1* (notochord) | 2.11 | 0.57 | 0.12 | 0.05 | 0.12 | 0.01 | 0.12 | 0.03 |
| | *Fli/Erg.a* (mesenchyme) | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.24 |
| | *Myl* (muscle) | 0.03 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 |

| | | Experimental Condition [a] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | ADnzE | ADnZE | ADNzE | ADNZE | ADnze | AdnZe | ADnZe | ADNze | ADNZe |
| Expression [c] | *Epi1* (epidermis) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | *Bco* (brain) | 0.15 | 0.14 | 0.15 | 0.12 | 0.27 | 0.12 | 0.13 | 0.14 | 0.15 |
| | *Celf3.a* (pan-neural) | 0.09 | 0.09 | 0.13 | 0.14 | 0.13 | 0.52 | 0.08 | 0.12 | 0.14 |
| | *Alp* (endoderm) | 0.19 | 0.83 | 0.29 | 0.45 | 1.20 | 0.26 | 0.75 | 0.64 | 0.77 |
| | *Noto1* (notochord) | 0.26 | 0.33 | 0.36 | 1.93 | 1.30 | 0.48 | 0.24 | 0.80 | 1.05 |
| | *Fli/Erg.a* (mesenchyme) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | *Myl* (muscle) | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 |

[a.] Each of the experimental conditions is represented by a five-letter code in which up- and down-regulation of *Foxa.a*, *Foxd*, *Neurog*, *Zic-r.b*, and Erk signalling are represented by A/a, D/d, N/n, Z/z, and E/e, respectively.

[b.] These results are also included in Table 1.

[c.] Expression levels of marker genes are shown relative to the corresponding values in normal 9.5 hr (tailbud-stage) embryos.

**Table S2. Regulatory interactions in the developmental gene regulatory network in *Ciona* embryos up to the late gastrula stage. Related to Figure 2; Transparent Methods.**

**Table S3. Cycles in the network of Fig. 2 determine FVS. Related to Figure 2; Transparent Methods.** $C_i$ are minimum cycles identified by sets of nodes. An FVS $I$ is given as a set of nodes that includes at least one member of all $C_i$, that is, $I \cap C_i \neq \emptyset, \quad \forall i$. The red nodes indicate our choice of FVS.

| Node | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ | $C_6$ | $C_7$ | $C_8$ | $C_9$ | $C_{10}$ | $C_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Fgf8/17/18* | * | | | | | | | | | | |
| Erk signalling | * | * | | | | | * | * | | | * |
| *Fgf9/16/20* | | * | | | | | | | | | |
| *Foxd* | | | * | | | | | | | | |
| *Twist-r.a/b* | | | * | | | | | | | | |
| *Tbx6.b* | | | | * | | | | | | | |
| *Zic-r.b* | | | | * | * | | * | | * | | |
| *Wnt5* | | | | | * | | | | | | |
| *Delta.b* | | | | | | * | | * | | | |
| *Neurog* | | | | | | * | | | | | |
| *Efna.d* | | | | | | | * | * | | | * |
| *Foxa.a* | | | | | | | | | * | * | |
| *Snail* | | | | | | | | | * | * | * |
| *Nodal* | | | | | | | | | | * | * |

**Table S4. Gene identifiers for genes used in the present study. Related to Related to Figure 2; Transparent Methods.**

| Gene | Identifier (CG.KH2012) |
|---|---|
| *Acta1* | C1.570 |
| *Admp* | C2.421 |
| *Alp* | L153.31 |
| *Bco* | C9.224 |
| *BHLHA15* | C3.308 |
| *Bmp2/4* | C4.125 |
| *Brachyury* | S1404.1 |
| *Cdx* | C14.408 |
| *Celf3.a* | C6.128 |
| *Cers.e* | C3.255 |
| *CG.KH2012.C8.686* | C8.686 |
| *Chd* (*Chordin*) | C6.145 |
| *Ctnnb* (*β-catenin*) | C9.53 |
| *Delta.b* | L50.6 |
| *Dlx.b* | L57.25 |
| *Dmrt.a* | S544.3 |
| *Dusp1/2/4/5* | C1.1079 |
| *Ebf3* | L24.10 |
| *Efna.b* | C3.202 |
| *Efna.c* | C3.52 |
| *Efna.d* | C3.716 |
| *Elk* | C8.247 |
| *Emx* | L142.14 |
| *Eph.a* | C1.404 |
| *Epi1* | C1.188 |
| *Epib* | C7.154 |
| *Fgf8/17/18* | C5.5 |
| *Fgf9/16/20* | C2.125 |

| | |
|---|---|
| *Fgl* | C1.832 |
| *Fli/Erg.a* | C4.539 |
| *Fos* | C11.314 |
| *Foxa.a* | C11.313 |
| *Foxb* | C4.341 |
| *Foxc* | L57.25 |
| *Foxd* | C8.890/C8.396 |
| *Foxh.a* | C9.717 |
| *Fzd4* | C6.162 |
| *Gata.a* | L20.1 |
| *Gata.b* | S696.1 |
| *Gdf1/3-r* | C4.547 |
| *Gsx* | C2.917 |
| *Hand-r* | C1.1116 |
| *Hes.a* | C1.159 |
| *Hes.b* | C3.312 |
| *Hhex* | L171.10 |
| *Id* | C7.692/C7.157 |
| *Jun* | C5.610 |
| *Lefty* | C3.411 |
| *Lhx3/4* | S215.4 |
| *Lmx1* | C9.616 |
| *Meis* | C10.174 |
| *Mesp* | C3.100 |
| *Mnx1* | L128.12 |
| *Mrf* | C14.307 |
| *Msx* | C2.957 |
| *Myl* | C1.1186/C1.20 |
| *Myt1* | C1.274 |
| *Neurog* | C6.129 |
| *Nkx2-1/4* | C10.338 |
| *Nodal* | C1.99 |
| *Nog* (*Noggin*) | C12.562 |

| | |
|---|---|
| *Noto1* | L20.18 |
| *Otp* | C14.377 |
| *Otx* | C4.84 |
| *Pax3/7* | C10.150 |
| *Pax6* | C9.68 |
| *Pem1* | C1.755 |
| *Pou4* | C2.42 |
| *Prdm1-r.a* | C12.493 |
| *Prdm1-r.b* | C12.105 |
| *Rlbp1* | C11.439 |
| *Sfrp1/5* | L171.5 |
| *Six3/6* | C10.367 |
| *Smyd1* | S423.6 |
| *Snail* | C3.751 |
| *Sox1/2/3* | C1.99 |
| *Sox4/11/12* | C7.523 |
| *Tbx2/3* | L96.87 |
| *Tbx6.a* | L8.11 |
| *Tbx6.b* | S654.3 |
| *Tfap2-r.b* | C7.43 |
| *Tp53.a* | C1.573 |
| *Tp53.b* | C3.713 |
| *Tubb* | L116.85 |
| *Twist-r.a/b* | C5.416/C5.554 |
| *Twist-r.c* | C5.202 |
| *Wnt5* | L152.45 |
| *Wnttun5* | C9.257 |
| *Zf249* | C4.182 |
| *Zf266* | C1.777 |
| *Zic-r.a* | C1.727 |
| *Zic-r.b* | L59.12/L59.1/S816.1/S816.4 |

**Table S5. Probes and primers used for quantitative PCR. Related to Related to Figure 3; Figure 4; Transparent Methods.**

| Gene | Fluorescent Probe (5' to 3') | Forward primer (5' to 3') | Reverse primer (5' to 3') |
|---|---|---|---|
| *Epi1* | (FAM)-ATCCTCGATATGAATGCGGTTTCCCC-(TAMRA) | CCAGACAATGGTGTTGGAAGAC | AACGCAGTGGAATTGAGTCACA |
| *Bco* | (VIC)-TCAGATCGATCCGGTGACCCTTGATACA-(TAMRA) | TCGCCATCACTGAAAGCAACT | GTGTTTCGCAAGATCAACCTTGT |
| *Celf3.a* | (FAM)-CTCGCCAGTAGCACGAACGCCC-(TAMRA) | GGCAAACCAACTGCAAACAA | CAACCATCAGGCCCTTCTTTT |
| *Alp* | (FAM)-AATCCTATTTTCGGCGCCGCTCC-(TAMRA) | CGGATCACAGCCATGTTTTTAC | CGACGAGCTTTGGATTATTAACGT |
| *Noto1* | (VIC)-CGTTCATGTACGGGTTTCTTGCAACCA-(TAMRA) | GGCTTGCCTGCGAATGG | GAGCACACGACTGCATCGTAA |
| *Fli/Erg.a* | (FAM)-ACGAGAAGGCGACCACCAATACACGA-(TAMRA) | TCCTACTACAGGGCAGGAAGCT | ACCCAAAGTATGCAACGTGTTTT |
| *Myl* | (VIC)-CGAGCCATTAACCTTAACCCAACCATTGAA-(TAMRA) | TGGATTCGATCAAGTAGGAGATGTT | CAATTTTTTGGCAGCCATATCTT |

## Transparent Methods

### Linkage logic theory (LLT)

*Formulation*

Consider a directed graph $\Gamma = (V, E)$ consisting of a node set $V$ and edge set $E$, and dynamics on the graph $\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x})$ $(\boldsymbol{x}, \boldsymbol{F} \in \mathbb{R}^{|V|})$ (Fiedler et al., 2013; Mochizuki, 2008; Mochizuki et al., 2013). We assume (i) continuous differentiability of $F_n$, that is, $F_n \in C^1$, and (ii) dissipativity, that is, for any initial condition $x(0)$ and for a finite time $t \geq 0$, the dynamical state $x(t)$ is bounded by a positive constant $C$: $|x_n(t)| \leq C$. Suppose that the dynamics of activity $x_n$ of biomolecule $n \in V$ is written in the form:

$$\begin{aligned} \dot{x}_n &= F_n(\boldsymbol{x}) \\ &= F_n\left(x_n, \boldsymbol{x}_{I_n}\right) \end{aligned} \tag{1}$$

with the third assumption (iii) decay condition:

$$\partial_1 F_n\left(x_n, \boldsymbol{x}_{I_n}\right) < 0. \tag{2}$$

In the expression, the bold face notation $\boldsymbol{x}_I$ with subset $I \subseteq V$ denotes the vector of components $x_i$ with $i \in I$. We explicitly specify self-regulation ($n \in I_n$) and self-loop

on the graph $\Gamma$, if and only if $\partial F_n/\partial x_n$ is 'not always negative'. Note that we omit the self-loop from $I_n$ and $\Gamma$, if the self-regulatory influence from $n$ to $n$ is negative (i.e. self-repression or decay) and representable by the decay condition.

Note that the decay condition $\partial_1 F_n < 0$ does not always imply $\partial F_n/\partial x_n < 0$. If total partial derivative $\partial F_n/\partial x_n$ is not negative, we can redefine $\tilde{F}_n$ by including a self-loop via $\tilde{I}_n \equiv I_n \cup \{n\}$ as:

$$\tilde{F}_n(x_n, \boldsymbol{x}_{\tilde{I}_n}) \equiv F_n(\boldsymbol{x}_{I_n}) + x_n - x_n.$$

Therefore, $\tilde{F}_n$ instead of $F_n$ with $\tilde{I}_n$ instead of $I_n$ always satisfies the decay condition even if $F_n$ itself does not. Thus, the decay condition does not limit the use of the formula of an ordinary differential equation.

*Theorem and proof*

Under formulations (1) and (2), we proved that a set of key nodes for dynamics is determined from the topology of the network (Fiedler et al., 2013; Mochizuki, 2008; Mochizuki et al., 2013).

Definition 1: In a directional graph $\Gamma = (V, E)$, a subset $I \subseteq V$ of nodes is called a feedback vertex set (FVS), if and only if a removal of the set $\Gamma \setminus I$ leaves a graph without directed cycles.

Definition 2: In a dynamic system, a subset $J \subseteq V$ of variables is called a set of determining nodes, if and only if two solutions satisfy $\tilde{\boldsymbol{x}}(t) - \boldsymbol{x}(t) \to 0 \ \ (t \to +\infty)$, whenever $\tilde{x}_n(t) - x_n(t) \to 0 \ \ (t \to +\infty)$ for all components $n \in J \subseteq V$(Fiedler et al., 2013; Foias and Temam, 1984).

We proved that these two different concepts are equivalent for the dynamics in a network (Fiedler et al., 2013; Mochizuki et al., 2013). In other words, observation of the long-term dynamics of the FVS $I$ is sufficient to identify all possible attractors of an entire system. Similarly, controlling the dynamics of the FVS ($\boldsymbol{x}_I^*(t) - \boldsymbol{x}_I(t) \to 0$) is sufficient to drive the dynamics $\boldsymbol{x}(t)$ of a whole system to converge on one of any attractors $\boldsymbol{x}^*(t)$.

We explain the theorem and sketch the proof in the following. Details of the proof have been reported by Fiedler *et al.* (Fiedler et al., 2013).

Theorem: In the dynamics on the directed graphs (1) and (2), an FVS of the graph is a set of determining nodes regardless of the choice of nonlinear function $F_n$. Conversely, if a subset of vertices of the graph is a set of determining nodes regardless of the choice of $F_n$, it is an FVS.

Proof: First, we show the if-part (FVS $\Rightarrow$ determining nodes). The first step is the rearrangement of non-FVS. From the definition of FVS, nodes in a complement $K(= V \setminus I) = \{1, \dots, |K|\}$ of FVS can be aligned so that a regulating (upper) node has a smaller number than a regulated (lower) node. In other words, $I_k \subseteq I \cup \{1, \dots, k-1\}$ $(\forall k \in K)$.

The second step is proof of the convergence of non-FVS under the convergence of FVS given. Let the difference of trajectories be $w_n(t) = \tilde{x}_n(t) - x_n(t)$. Suppose that $w_i(t) \to 0$ is given for all $i \in I$ included in FVS, we show $w_k(t) \to 0$ for all $k \in K$ in non-FVS via mathematical induction. For the dynamics of the difference of trajectories $\boldsymbol{w}(\mathrm{t})$, the following is induced by the mean value theorem.

$$\begin{aligned}\dot{\boldsymbol{w}}(t) &= \left[\boldsymbol{F}\big(\boldsymbol{x}(t)+\theta\boldsymbol{w}(t)\big)\right]_{\theta=0}^{1} \\ &= \int_0^1 \frac{d}{d\theta}\boldsymbol{F}\big(\boldsymbol{x}(t)+\theta\boldsymbol{w}(t)\big)d\theta \\ &= \int_0^1 \frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{F}\big(\boldsymbol{x}(t)+\theta\boldsymbol{w}(t)\big)\cdot\boldsymbol{w}(t)d\theta \\ &= A(t)\boldsymbol{w}(t)\ ,\end{aligned}$$

where

$$A(t) \coloneqq \left(\int_0^1 \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}\Big|_{\boldsymbol{x}(t)+\theta\boldsymbol{w}(t)}\, d\theta\right).$$

The $A$ is a matrix given by integration of each element of the Jacobian of $\boldsymbol{F}$ at $\boldsymbol{x}(t)+\theta\boldsymbol{w}(t)$ with respect to $\theta$. Note that $K$ does not have any self-regulatory nodes, and takes a linear nonautonomous dynamic equation for each $k \in K$ as $\dot{w}_k(t) = -a_k(t)w_k(t) + \boldsymbol{b}_k^T(t)\cdot\boldsymbol{w}_{I_k}(t)$. From the assumption of dissipativity, the nonautonomous coefficients $a_k(t) \in \mathbb{R}$, $\boldsymbol{b}_k(t) \in \mathbb{R}^{|I_k|}$ are bounded by constants $a_0$, $b_0$ as $0 < a_0 \leq a_k(t)$, $|\boldsymbol{b}_k(t)| \leq b_0$. Assuming that $w_n(t) \to 0$ is already given $\forall n \in \{1,\cdots,k-1\}$, we show it for $n = k$. By solving the nonautonomous dynamic equation, we have

$$
\begin{aligned}
w_k(t) &\le \exp\left(-\int_0^t a_k(s)ds\right)|w_k(0)| \\
&\quad + \sum_{j\in I_k}\int_0^t \exp\left(-\int_s^t a_k(\sigma)d\sigma\right)|\boldsymbol{b}_k(s)|\left|w_j(s)\right|ds \\
&\le \exp(-a_0 t)|w_k(0)| + \sum_{j\in I_k}\int_0^t \exp\big(-a_0(t-s)\big)b_0\left|w_j(s)\right|ds \\
&\to 0 \quad (t\to\infty)
\end{aligned}
$$

Here, the first term is shown by $\exp(-a_0 t)\to 0$, and the second term is shown by $w_j(s)\to 0\ \ (s\to\infty)\ \ \forall j\in I_k\subseteq I\cup\{1,\dots,k-1\}$. In the case of $k=1,\ I_k\subseteq I$, trivially $w_k(t)\to 0$ because $w_j(s)\to 0$ is given $\forall j\in I_k\subseteq I$. From the above, the first half of the theorem is proven.

Finally, the only-if-part (FVS $\Leftarrow$ determining nodes) is shown by taking the contrapositive, that is, a subset of vertices that is not an FVS is not a set of determining nodes. Suppose $I'$ is not an FVS; in other words, $\Gamma\setminus I'$ contains directed cycles. By appropriate selection of the function of nodes, dynamics can be constructed so that $I'$ is not a set of determining nodes. For example, all functions included in $I'$ are taken to be simple decay $F_n\big(x_n,\boldsymbol{x}_{I_n}\big):=-x_n$. From this, the behaviour of $\Gamma\setminus I'$ cannot be captured by $I'$. However, there is a cycle within $\Gamma\setminus I'$ and, by choosing these functions, it is

possible to create diversity in the solutions such as multiple stationary points. In other words, $I'$, which is not an FVS, is not always a set of determining nodes for arbitrary functions.

Our theory of controllability has a broader meaning than switching between solutions that can be observed in natural conditions. For any given trajectory $\boldsymbol{x}_I^*(t)$ of an FVS, dynamics $x_k^*(t)$ of other nodes, which are not included in the FVS ($k \notin I$), converges to a unique trajectory for a long time, even if the given trajectory $\boldsymbol{x}_I^*(t)$ is not chosen from known natural solutions $s \in S$.

**Identification of FVSs**

The GRN is shown in Figure 2 and also given as a list of linkages (edges) connecting genes (nodes) in Table S2. Note that our theory does not require a distinction between positive and negative regulation, except for self-regulation. We omitted linkages of self-repression from the decay condition (see treatment of $n$ in $I_n$). Multiple methods

and algorithms have been proposed to identify minimum FVS from network structures. Our algorithm is as follows.

We first identified nodes (genes) that are not regulated by others and nodes that do not regulate other nodes, and removed these nodes and connecting edges from the gene regulatory network repeatedly because removal of these nodes does not affect the identification of FVSs.

We next identified independent directed cycles in the graph by a 'depth-first search algorithm', starting from all nodes, examining paths through directed edges and recursively choosing all emanating edges at branching points. We repeated this process until it came back to the starting point (identification of a cycle) or reached a cycle that had already been identified. We chose a cycle with a smaller number of nodes if a pair of cycles exhibited an inclusion relationship. For each cycle, we identified a set of nodes $C_i = \{n_1^i, n_2^i, \dots, n_{m_i}^i\}$ ($i = 1,2, \dots, i_{\max}$) passed through by that cycle. The $C_i$ with $i_{\max} = 11$ for the gene regulatory network of *Ciona intestinalis* is shown in Table S3. Finally, we identified sets (Is) of nodes in which at least one member of all $C_i$ ($i = 1,2, \dots, i_{\max}$) was included: $I \cap C_i \neq \emptyset$ for all $i$. We found that 12 $I$s were the smallest sets, each of which contained five

genes. These $I$s were the minimum FVSs. The minimum FVSs are {*Foxa.a*|*Nodal*|*Snail*, *Foxd*|*Twist-r.a/b*, *Neurog*|*Delta.b*, *Zic-r.b*, Erk signalling}, where '|' indicates an alternative choice.

The computer code for identifying minimum FVSs is available on https://github.com/kmaed/searchfvs.

**Comparison with alternative method**

Another study has already provided a criterion to choose driver nodes based on network information. The formulation and proof are given for a linear or linearized system, based on standard control theory (Kalman, 1963; Liu et al., 2011).

$$\dot{\boldsymbol{x}}(\mathrm{t}) = \mathrm{A}\boldsymbol{x}(\mathrm{t}) + \mathrm{B}\boldsymbol{u}(\mathrm{t})$$

Here, $\boldsymbol{x}(t) \in \mathbb{R}^N$ is the state vector, $\boldsymbol{u}(t) \in \mathbb{R}^M$ is the input vector, $\mathrm{A}$ is an $N \times N$ matrix, and $\mathrm{B}$ is an $N \times M$ matrix. The network structure is reflected in the distribution of nonzero entries in the matrix $\mathrm{A}$. Kalman's controllability is defined as the full rankness of controllability matrix, namely:

$$\text{rank}[\text{B AB A}^2\text{B } \ldots \text{ A}^{N-1}\text{B}] = N.$$

Lin (Lin, 1974) gave proof of the controllability of a linear system by introducing the concepts ‘cactus’ and ‘spanned by a cactus’. Later, Liu *et al.* (Liu et al., 2011) adapted the method to many examples of regulatory networks by an algorithm using different terminology, ‘maximum matching’. The minimum number of inputs or driver nodes needed to maintain full control of the network is determined by the ‘maximum matching’ in the network, that is, the maximum set of links that do not share both start and end nodes. A node is considered to be ‘matched’ if a link in the maximum matching points at it; otherwise, it is unmatched. If there are directed paths from the input to all matched nodes, then the system is controllable by driving unmatched nodes.

There are multiple differences between our theory (FVS control) and that proposed by Liu *et al*. (i) Our theory is applicable for any nonlinear dynamic system, even if the nonlinear functions are unknown. (ii) We use the same FVS for both observation and control. The equivalence of sets for observability and controllability realizes ‘observation-base control’. We achieve this control by prescribing the behaviour $\boldsymbol{x}_I$ on FVS $I$ to their previously observed trajectory $\boldsymbol{x}_I^S$, $\boldsymbol{x}_I(t) - \boldsymbol{x}_I^S(t) \to 0$. The procedure is much closer to the

idea of ‘reprogramming’ regulatory networks in life sciences. (iii) Moreover, the aim and meaning of ‘control’ differ between the two methods. Liu *et al.* sought to steer the network state $x(t)$ from any initial state $\boldsymbol{x}_0$ to any target state $\boldsymbol{x}_T$ in a linear space. Instead, we seek to steer $\boldsymbol{x}(t)$ from any initial state $\boldsymbol{x}_0$ to any target solution $\boldsymbol{x}^*(t)$ of the original system (1), (2).

**Animals, whole-mount *in situ* hybridization, and gene identifiers**

*Ciona intestinalis* (type A; also called *Ciona robusta*) adults were obtained from the National Bio-Resource Project for *Ciona* in Japan. cDNA clones were obtained from our EST clone collection (Satou et al., 2005). Whole-mount *in situ* hybridization was performed as described previously (Satou et al., 1995). Gene identifiers according to the nomenclature rule (Satou et al., 2008; Stolfi et al., 2015) are shown in Table S4.

**Gene knockdown and overexpression**

All morpholino antisense oligonucleotides (MOs) (Gene Tools, LLC) used in the present study block translation. These MOs have been used previously and their specificity has

been evaluated (Imai et al., 2006). For synthetic mRNAs, coding sequences of *Foxa.a*, *Foxd*, *Neurogenin*, and *ZicL* were cloned into pBluescript RN3 (Lemaire et al., 1995), and synthetic mRNAs were transcribed using the mMESSAGE mMACHINE T3 Transcription Kit (Thermo Fisher Scientific). Each of the MOs was prepared at a concentration of 0.4 mM, and each mRNA was prepared at a concentration of 0.5 μg/μL. Mixtures of MOs and mRNAs were injected into eggs in volumes of 30 pL. Injection of a control MO against *E. coli lacZ* (5′-TACGCTTCTTCTTTGGAGCAGTCAT-3′) at a concentration of 1.6 mM or control *lacZ* mRNA at a concentration of 2 μg/μL yielded larvae with normal morphology. For the arrest of cell division, embryos were incubated in seawater containing 2.5 μg/mL cytochalasin B (Sigma). For up- and down-regulation of Erk signalling, we treated embryos with 10 ng/mL human recombinant basic FGF (Sigma, F0291) and 2 μM of the MEK inhibitor U0126 (Calbiochem). Reverse transcription followed by quantitative PCR was performed using the Cells-to-CT kit (Thermo Fisher Scientific). Each experimental embryo was placed into a single tube and reverse-transcription was performed in accordance with the manufacturer's instructions. Quantitative PCR was performed using the TaqMan method with primers and probes shown in Table S5.

**RNA sequencing (RNA-seq)**

For the RNA-seq experiments, notochord partial embryos, mesenchyme partial embryos, aDnze embryos, and adnZE embryos were collected. Two partial embryos were obtained from presumptive notochord (A7.3 and A7.7) and mesenchyme (B8.5 and B7.7) cells isolated using glass needles. RNA-seq experiments were performed as described previously

(Tokuhiro et al., 2017). NOISeq (Tarazona et al., 2011) was used to identify differentially expressed genes. We used adjusted *p*-values for multiple testing to identify differentially expressed genes, setting a threshold of 0.001.

## Data Availability

The RNA-seq data is available under the SRA accession number, DRA006310.

**Supplemental References**


Fiedler, B., Mochizuki, A., Kurosawa, G., and Saito, D. (2013). Dynamics and Control at Feedback Vertex Sets. I: Informative and Determining Nodes in Regulatory Networks. J Dynam Differential Equations 25, 563-604.

Foias, C., and Temam, R. (1984). Determination of the Solutions of the Navier-Stokes Equations by a Set of Nodal Values. Math Comput 43, 117-133.

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

Kalman, R.E. (1963). Mathmatical Description of Linear Dynamical Systems. J SIAM Control Ser A 5, 152-192.

Lemaire, P., Garrett, N., and Gurdon, J.B. (1995). Expression cloning of Siamois, a Xenopus homeobox gene expressed in dorsal-vegetal cells of blastulae and able to induce a complete secondary axis. Cell 81, 85-94.

Lin, C.T. (1974). Structural Controllability. Ieee T Automat Contr Ac19, 201-208.

Liu, Y.Y., Slotine, J.J., and Barabasi, A.L. (2011). Controllability of complex networks. Nature 473, 167-173.

Mochizuki, A. (2008). Structure of regulatory networks and diversity of gene expression patterns. Journal of theoretical biology 250, 307-321.

Mochizuki, A., Fiedler, B., Kurosawa, G., and Saito, D. (2013). Dynamics and control at feedback vertex sets. II: a faithful monitor to determine the diversity of molecular activities in regulatory networks. Journal of theoretical biology 335, 130-146.

Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A., and Satoh, N. (2005). An integrated database of the ascidian, Ciona intestinalis: Towards functional genomics. Zool Sci 22, 837-843.

Satou, Y., Kusakabe, T., Araki, S., and Satoh, N. (1995). Timing of Initiation of Muscle-Specific Gene-Expression in the Ascidian Embryo Precedes That of Developmental Fate Restriction in Lineage Cells. Dev Growth Differ 37, 319-327.

Satou, Y., Mineta, K., Ogasawara, M., Sasakura, Y., Shoguchi, E., Ueno, K., Yamada, L., Matsumoto, J., Wasserscheid, J., Dewar, K., et al. (2008). Improved genome assembly and evidence-based global gene model set for the chordate Ciona intestinalis: new insight into intron and operon populations. Genome Biol 9, R152.

Stolfi, A., Sasakura, Y., Chalopin, D., Satou, Y., Christiaen, L., Dantec, C., Endo, T., Naville, M., Nishida, H., Swalla, B.J., et al. (2015). Guidelines for the nomenclature of genetic elements in tunicate genomes. Genesis 53, 1-14.

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in RNA-seq: A matter of depth. Genome Res 21, 2213-2223.

Tokuhiro, S., Tokuoka, M., Kobayashi, K., Kubo, A., Oda-Ishii, I., and Satou, Y. (2017). Differential gene expression along the animal-vegetal axis in the ascidian embryo is maintained by a dual functional protein Foxd. PLoS genetics 13, e1006741.