DYNAMICS AND CONTROL AT FEEDBACK VERTEX SETS.

II: A FAITHFUL MONITOR TO DETERMINE THE DIVERSITY OF MOLECULAR ACTIVITIES IN REGULATORY NETWORKS.

Atsushi Mochizuki[1], Bernold Fiedler[2], Gen Kurosawa[1] and Daisuke Saito[1]

[1]*Theoretical Biology Laboratory, RIKEN, Wako 351-0198, Japan*

[2]*Institut für Mathematik, Freie Universität Berlin, Arnimallee 3, D-14195 Berlin, Germany*

Running Headline: feedback vertex set as determining nodes

Corresponding Author: Atsushi Mochizuki

Address: Theoretical Biology Laboratory, RIKEN,
Wako 351-0198, Japan

E-mail: mochi@riken.jp

Telephone: +81-48-467-8422

Facsimile: +81-48-462-1709

(Submitted to Journal of Theoretical Biology, April 4, 2013)

(Revised and Resubmitted to Journal of Theoretical Biology, May 27, 2013)

Abstract

Modern biology provides many networks describing regulations between many species of molecules. It is widely believed that the dynamics of molecular activities based on such regulatory networks are the origin of biological functions. However, we currently have a limited understanding of the relationship between the structure of a regulatory network and its dynamics. In this study we develop a new theory to provide an important aspect of dynamics from information of regulatory linkages alone. We show that the “feedback vertex set” (FVS) of a regulatory network is a set of "determining nodes" of the dynamics. The theory is powerful to study real biological systems in practice. It assures that i) any long-term dynamical behavior of the whole system, such as steady states, periodic oscillations or quasi-periodic oscillations, can be identified by measurements of a subset of molecules in the network, and that ii) the subset is determined from the regulatory linkage alone. For example, dynamical attractors possibly generated by a signal transduction network with 113 molecules can be identified by measurement of the activity of only 5 molecules, if the information on the network structure is correct. Our theory therefore provides a rational criterion to select key molecules to control a system. We also demonstrate that controlling the dynamics of the FVS is sufficient to switch the dynamics of the whole system from one attractor to others, distinct from the original.

Keywords: Regulatory networks; complex systems; feedback vertex set; determining nodes; informative nodes

1. Introduction

By the success of modern biology, we have many examples of large networks which describe regulations between a large number of species of molecules, such as genes, proteins or ions (e.g. Davidson *et al*., 2002; Oda *et al*., 2005). It is widely believed that the dynamics of molecular activities based on such regulatory networks are the origin of biological functions. For example, circadian rhythms observed in many species are produced by periodic oscillation of gene activities. The differences in characteristics of cells are produced by differences in gene expression patterns generated in the developmental process. Diversities of differentiated cells are considered to be caused by the diversity of steady states of gene expressions (Davidson *et al*., 2002). One of the major objectives in modern biology is to understand biological functions in terms of the dynamics of the activity of bio-molecules, based on experimentally determined regulatory networks.

However, a variety of obstacles still impede attempts to study the dynamics of biological systems based on the knowledge of regulatory networks systematically. One of the difficulties is the observation of dynamic processes. It is still difficult to observe the dynamics of the activity of bio-molecules with sufficient time resolution. Most of the data obtained by present experimental methods are snapshots of molecular activities rather than time tracks. The second problem is the reliability of the regulatory network itself. At present the regulatory networks are possibly incomplete in many studies of biological systems because of the complexity and working cost of experimental procedures to identify regulatory edges. The problem is fundamental because we can never exclude the possibility that unknown species of molecules or unknown regulations may take an important role in the focal phenomena.

The third and largest problem is that the information on the regulatory network alone is not sufficient to determine the resulting dynamics. The regulatory edges only provide qualitative information on dependencies between activities of bio-molecules in the system. They lack essential quantitative details like the regulatory functions, parameter values of reaction rates, and initial states. "In silico" numerical simulations therefore rely on many unverified assumptions as to the regulatory functions and their dozens or hundreds of unknown parameters. In general, numerical parameter identification does not seem to be a viable option.

Figure 1a is an example of an extremely simple regulatory network, which is sometimes called a "binary switch" because it is expected to generate two stable stationary states of gene expressions. We analyzed this system based on an ordinary differential equations (ODE) system of the form (1), which will be explained in detail later. We assumed two different types of regulatory functions to occur in the regulatory network. Figure 1b and 1d describe these two possible choices. The resulting dynamical behavior depends on the choices of regulatory functions. For the regulatory functions shown in Figure 1b, we observed bi-stability of gene expressions, which meets with the expectation of a "binary switch". However, if we use the functions in Figure 1d, which is another possible interpretation of Figure 1a, we rather observe three stable stationary points. This is just one example showing that the dynamics of molecular activity may depend on the form of the regulatory function. In general it is therefore not possible to infer the dynamics just from information on regulatory networks like Figure 1a.

There are some studies that discuss the general relation between regulatory networks and the dynamics of bio-molecules. Shen-Orr *et al*. (2002) introduced the

idea of a network motif, which is a small sub-structure, and they showed that some network motifs occur in actual biological networks with significantly higher probability than expected from randomly generated networks.

To overcome or circumvent some of these problems, we developed a mathematical theory to analyze the dynamical properties of complex biological systems based on information of the regulatory linkages alone (Mochizuki, 2008; Mochizuki and Saito, 2010; Fiedler *et al*., 2013). The theory ensures that:

i) all non-transient dynamical behavior of the whole system can be identified faithfully by the measurement of only a subset of variables in the system, and

ii) the subset is determined from the regulatory linkage alone as a "feedback vertex set" (FVS) of the network graph, as will be explained below.

In the companion paper (Fiedler *et al*., 2013), we provide mathematical proofs of our theory. In the present paper we mainly explore some applications of our theory to analyze regulatory networks in biology. We discuss three prominent regulatory networks from the biology literature for cell-differentiation, signal transduction, and circadian rhythms. In the first two examples, we determine very small feedback vertex sets from rather large regulatory networks. Comparing with experimental data on real molecular activities, we discuss the possible inadequacy of the proposed networks to generate the observed biological phenomena. This implies that unknown edges for unknown molecules may exist which are responsible for the observed phenomena.

To explore the control aspect of our theory, we demonstrate that a system can be controlled by just prescribing the dynamics on the feedback vertex set. Using a gene regulatory network of circadian rhythms, we numerically demonstrate how to control a system by a minimal and sufficient number of key variables related to the

feedback vertex set.

We will show that our theory is powerful to understand dynamics of biological systems in practice, based on experimentally determined networks. By our theory, we conclude that some aspects of the dynamical properties of the system can be derived from information on the regulatory linkages, only, without using other information. In particular, we can discuss and check the consistency between the regulatory network and the observed dynamics of its molecular activities, without knowing further quantitative information to specify the unobserved or hidden parts of the dynamics.

2. Mathematical formalization

First, let us explain an ODE setting for the dynamics of molecular activity popularly used in mathematical biology, although we will generalize the formula later. Let $x_k(t) \in \mathbb{R}$ denote the activity of bio-molecules k at time t. Then the rate of change $\dot{x}_k = \frac{d}{dt} x_k$ can be expressed in the following form:

$$\dot{x}_k = f_k\left(\mathbf{x}_{I_k}\right) - d_k\left(x_k\right), \qquad k = 1, \ldots, N \quad . \tag{1}$$

Here f_k is any positive non-linear function showing the enhancement of activity of molecule k, which we may call regulatory function, and d_k is a positive and increasing function for the decay of activity (Glass and Kauffman, 1973; Mochizuki, 2005). The set $I_k \subseteq \{1, \ldots, N\}$ is a subset of molecules which regulate the molecule

k. In other words, I_k is the input set of k. The bold face notation $\mathbf{x}_{I_k}$ abbreviates the vector of the components x_j with $j \in I_k$.

The decay term d_k is always present, because the activities of the molecules actually measure the concentrations of materials, including mRNAs, active states of proteins or other small molecules. In addition to the decay, the suppression of activity by self-regulation is assumed to be expressed by d_k. We do not include k itself in the input set I_k when the effect of x_k on $\dot{x}_k$ is either self-repression or decay. In other words, I_k includes k, i.e. $k \in I_k$, if and only if self-activation of molecule k exceeds self-repression and decay. The collection of input sets I_k, $k = 1, \ldots, N$, specifies all regulatory relations explicitly.

More generally we consider the following broader class of ODE model of regulatory networks:

$$\dot{x}_k = F_k\left(x_k, \mathbf{x}_{I_k}\right), \qquad k = 1, \ldots, N. \tag{2}$$

Here $F_k\left(x_k, \mathbf{x}_{I_k}\right)$ are any non-linear functions. Modeling self-repression or decay we assume $\partial_1 F_k\left(x_k, \mathbf{x}_{I_k}\right) < 0$ for all k. Here ∂_1 denotes the first partial derivative with respect to the first occurrence of the argument x_k. As for the second argument set $\mathbf{x}_{I_k}$, the treatment of the variable x_k is similar to the previous form (1). We allow $k \in I_k$ if and only if the total derivative of F_k with respect to x_k is not always

negative. In this way we eliminate the monotonicity assumption for the term d_k in the formula (1). We do not have to assume that decay or self-repression terms are separated from other terms as a functions of a single variable x_k. In other words, decay or self-repression may not only be a function of x_k, but may also depend on other variables in $\mathbf{x}_{I_k}$. For a mathematically complete explanation of this form, see also the companion paper (Fiedler *et al*., 2013).

The ODE system (1) or (2) encodes the information of the regulatory network, i.e. information on dependencies between activities of bio-molecules in the system. Of course, the precise dynamical behavior depends on quantitative and qualitative details, like the precise form of the regulatory functions f_k, the parameter values of reaction rates, or the initial states as shown in Figure 1. In the following we consider the dynamics of the system (1) or (2) based on the information of the regulatory network I_k, only, without using any such further information.

To summarize, there is a graph theory aspect and a dynamical systems aspect to regulatory networks of the form (1) or (2). The regulatory graph Γ of the network simply consists of all "species" or "reactants" x_k, as vertices, and of all directed edges $j \to k$ such that j is an element of the input set I_k of k. The dynamical systems aspect, on the other hand, asks for the time-dependent behavior $x_k(t)$ of the concentrations x_k of the participating bio-molecule species or reactants. It is our purpose to explore the interplay of this dynamics with the regulatory graph structure.

3. Feedback vertex sets as determining nodes

Here we explain and combine two concepts from two different mathematical fields. The first concept are “feedback vertex sets” (FVS) from graph theory. A FVS is a subset of vertices in a directed graph, such that the removal of the set leaves the graph without directed cycles. The second concept are “determining nodes” from dynamical systems. In our setting (1), (2), we call a subset of variables $J \subseteq \{1,\ldots,N\}$ "determining nodes" if and only if the convergence of variables in J, for any two trajectories implies the convergence of all variables of these trajectories. More precisely let $\mathbf{x}(t)$ and $\tilde{\mathbf{x}}(t)$ be two solutions of (1) or (2). Then we require $\mathbf{x}(t)-\tilde{\mathbf{x}}(t)\to 0$ to hold in the limit for large time $t\to+\infty$, i.e. for all components $k\in\{1,\ldots,N\}$, if the two solutions satisfy $\mathbf{x}_J(t)-\tilde{\mathbf{x}}_J(t)\to 0$, i.e. for all components k in the subset $J\subseteq\{1,\ldots,N\}$. In other words, if the dynamics of the determining nodes J are given for large times t, then the dynamics of the whole system are determined uniquely, for large times.

The concept of "determining nodes" was first proposed in the context of the Navier-Stokes equations of hydrodynamics (Foias and Temam, 1984; Foias and Titi, 1991). The previous discussions in Navier-Stokes context focused on the existence of a finite number of spatial locations ("nodes"), the dynamics on which is sufficient to determine the potentially infinite-dimensional dynamics of the total system. The relationship between determining nodes and the argument set of an ODE, i.e. regulatory networks, has not been discussed. In the companion paper (Fiedler *et al*., 2013), we combined the above two concepts for the first time. We proved mathematically that

any feedback vertex set of a regulatory network is a set of determining nodes of the dynamics on the network. Conversely, if a vertex set is determining, for *all* choices of nonlinearities compatible with the network structure, then it is a feedback vertex set.

Let us illustrate the concept of feedback vertex sets using some examples of small networks as shown in Figure 2. We show a choice of a feedback vertex set with gray-marked vertices. Figure 2a is the directed 3-cycle. Removal of any one vertex cuts the 3-cycle and leaves a graph without any directed cycle. Thus any one vertex among the three is a minimal feedback vertex set. If the network with three vertices has reversible, bidirectional regulatory edges (Figure 2b), we need to remove two vertices to obtain a cycle-free graph. Thus any two vertices among the three provide a minimal feedback vertex sets of Figure 2b. Although the network shown in Figure 2d looks complex, the minimal feedback vertex is small. As all directed cycles traverse the vertex at the center, the set including only the central vertex is the minimal feedback vertex set. Figure 2e is another case where the minimal feedback vertex set includes only one single vertex. In the following we frequently use the term "feedback vertex set" to indicate the choice of a "minimal feedback vertex set". There may be multiple ways to select a minimal feedback vertex set. The importance of feedback vertex sets in dynamics was first mentioned by Akutsu *et al*. (1998) for the steady states of Boolean network systems. Tamura et al. (2010) have also used the concept of feedback vertex set, in a Boolean setting, to analyze metabolic networks.

The details of the mathematical proof are given in the companion paper (Fiedler *et al*., 2013). Here, we provide a brief intuitive explanation of our theory. First, let us consider a single regulation in a network. Of course, if the dynamics of the input vertices are given, the long-term dynamics of the downward vertex is determined

uniquely. If we do not know the regulatory function leading to the lower vertex, the dynamics is not determined constructively, but is still determined uniquely. Then, let us consider a system of a regulatory network including several vertices and edges as in Figure 3. Here we ignore the vertices that do not receive any regulatory input, because the dynamics of such "top" vertices converges to the trivial unique stationary point, and does not contribute to the diversity of attractors generated by the autonomous dynamics of the system. Repeating our previous argument inductively, downward through the network, the dynamics of the whole system can be determined uniquely if the dynamics of an appropriate subset of vertices is given. Of course, the dynamics of the total system can be determined uniquely only when that subset is chosen appropriately; all of the remaining vertices should be downward of the vertices in that subset. Third, let us consider a problem: how can we minimize that subset, on which the dynamics are given? Our answer is the minimal feedback vertex set I. Indeed, if I is a feedback vertex set of a graph with vertex set $\Gamma = \{1, 2, \ldots, N\}$, then all remaining vertices in $K = \Gamma \setminus I$ can be ordered from I on downward, by the definition of the feedback vertex set I. This implies that the dynamics of all vertices in the system can be determined uniquely by induction from the dynamics on the feedback vertex set I. We repeat, once again, that we have recursively ignored all vertices without direct input in this argument.

The proof gives an assurance that we can detect all possible dynamical large time behavior of a system just by measuring the trajectories on the FVS of the regulatory network. The FVS is a concept of graph theory, and is determined only from a given graph of the regulatory network, of course. In many biological systems which show a complex diversity of behaviors, regulatory networks are determined, but

other information is not readily available. Our theory makes it possible to discuss the consistency between the regulatory network and the observed dynamics of the molecular activities, without any additional information. In the following three sections, we show examples of applications of the concepts to biological networks. We analyze regulatory networks for cell differentiation, signal transduction, and circadian rhythms.

4. Cell differentiation in development of ascidiacea

4-1. Analysis of network

We consider a gene regulatory network determined by Imai *et al*. (2006), which is responsible for cell-differentiation in the development of the ascidiacea *Ciona intestinalis* from the 16-cell stage to the tail-bud stage. In the focal period, the difference in gene activities between cells progresses with time, and 13 different gene expression patterns are observed at the final, tail-bud stage, depending on the position of cells in the body. The tail-bud stage continues for a relatively longer period than the previous developmental stages and the regional gene expression patterns are kept for that longer period. The system is expected to be flexible enough to produce many steady states of gene activities, which correspond to the differentiated cell types.

In the study (Imai *et al*., 2006) 80 genes were identified to control embryogenesis of *Ciona*. The regulatory interactions between these genes were examined by perturbation analysis, where the activity of one gene is manipulated and the effects are examined. Although a few of the disruptions were not successful for experimental reasons, the systematic analysis drew a complex intermingled network of regulatory interactions between the 80 genes as shown in Figure 4a. In Appendix A we

also provide the regulatory interactions of the ascidiacea gene network in text form.

The regulatory edges in the network by Imai *et al*. (2006) are categorized into two classes, activation and repression. We do not need to distinguish them here except for self regulatory edges. There are 16 genes with self-regulatory edges, all of which are self-repressions. We remove these self-repressive edges from the network because any self-repression can be subsumed into the decay term, i.e. a negative partial derivative of the regulatory function F_k with respect to the first argument x_k.

As a preparation we removed the vertices that do not receive any regulations or do not regulate any vertices. These top or bottom genes converge to fixed inputs or provide outputs of the system which do not contribute to the diversity of attractors. The removal procedure was repeated as long as the network had vertices with no input or no output. Note that this reduction preserves all directed cycles, and hence preserves any minimal feedback vertex set. We obtained the reduced network of Figure 4b with only 7 vertices. The network clearly possesses a minimal feedback vertex set which consists of only a single feedback vertex, FoxD-a/b. All long-term dynamics on the global attractor possibly generated by this gene regulatory network can therefore be identified by measurement of the activity of the single gene FoxD-a/b, if the network information is correct.

If gene expressions observed in differentiated cells at the tail-bud stage are stable equilibria of this system, their diversity should be identified by the activity of FoxD-a/b. Let us consider the equation system $x_k = h_k\left(\mathbf{x}_{I_k}\right) \equiv d_k^{-1} \cdot f_k\left(\mathbf{x}_{I_k}\right)$ for all vertices k, that is satisfied at the equilibria of the ODE (1) as follows:

$$
\begin{aligned}
x_{Fox} &= h_{Fox}\left(x_{Otx}, x_{Twist}\right) \\
x_{FGF} &= h_{FGF}\left(x_{Fox}\right) \\
x_{nodal} &= h_{nodal}\left(x_{Fox}, x_{FGF}\right) \\
x_{NoTrlc} &= h_{NoTrlc}\left(x_{nodal}, x_{Fox}, x_{FGF}\right) \\
x_{Otx} &= h_{Otx}\left(x_{FGF}\right) \\
x_{Twist} &= h_{Twist}\left(x_{NoTrlc}, x_{Fox}, x_{FGF}, x_{Otx}, x_{ZicL}\right) \\
x_{ZicL} &= h_{ZicL}\left(x_{Fox}, x_{FGF}\right)
\end{aligned}
\qquad . \qquad (3)
$$

After a simple reduction procedure, we obtain the following equivalent system:

$$
\begin{aligned}
x_{Fox} &= h_{Fox}\left(h_{Otx}\left(h_{FGF}\left(x_{Fox}\right)\right), h_{Twist}\left(h_{NoTrlc}\left(h_{nodal}\left(x_{Fox}, h_{FGF}\left(x_{Fox}\right)\right), x_{Fox}, h_{FGF}\left(x_{Fox}\right)\right), x_{Fox}, h_{FGF}\left(x_{Fox}\right), h_{Otx}\left(h_{FGF}\left(x_{Fox}\right)\right), h_{ZicL}\left(x_{Fox}, h_{FGF}\left(x_{Fox}\right)\right)\right)\right) \\
x_{FGF} &= h_{FGF}\left(x_{Fox}\right) \\
x_{nodal} &= h_{nodal}\left(x_{Fox}, h_{FGF}\left(x_{Fox}\right)\right) \\
x_{NoTrlc} &= h_{NoTrlc}\left(h_{nodal}\left(x_{Fox}, h_{FGF}\left(x_{Fox}\right)\right), x_{Fox}, h_{FGF}\left(x_{Fox}\right)\right) \\
x_{Otx} &= h_{Otx}\left(h_{FGF}\left(x_{Fox}\right)\right) \\
x_{Twist} &= h_{Twist}\left(h_{NoTrlc}\left(h_{nodal}\left(x_{Fox}, h_{FGF}\left(x_{Fox}\right)\right), x_{Fox}, h_{FGF}\left(x_{Fox}\right)\right), x_{Fox}, h_{FGF}\left(x_{Fox}\right), h_{Otx}\left(h_{FGF}\left(x_{Fox}\right)\right), h_{ZicL}\left(x_{Fox}, h_{FGF}\left(x_{Fox}\right)\right)\right) \\
x_{ZicL} &= h_{ZicL}\left(x_{Fox}, h_{FGF}\left(x_{Fox}\right)\right)
\end{aligned}
\qquad , \qquad (4)
$$

where the right hand sides of all equations are explicit functions of x_{Fox}. The first equation of (4) depends only on x_{Fox}, i.e. possible equilibria of FoxD-a/b are given as solutions of that single equation for x_{Fox}. For each solution of x_{Fox}, the remaining variables x_k are determined uniquely from x_{Fox} using the remaining equations in (4). Thus diversity of equilibria of the system is indeed determined by the solutions of the first equation in (4).

We can verify the result by examining whether diversity of cell differentiation can be identified by the activity of FoxD-a/b only. Actually Imai *et al*. (2006)

provided data of gene expressions in differentiated cells. In Table 1 we summarized gene expressions in 13 different cells at the tail bud stage from (Imai *et al*., 2006). Unfortunately, it is usual in present developmental biology that gene expressions are interpreted in a discrete and binary manner, i.e. active (1) or inactive (0). Of course, it is impossible to identify the diversity of 13 different gene expressions by only two points of one binary variable, even without looking Table 1 in detail.

There are two possible ways to interpret the result. The first possibility is that the diversity of differentiated cells may be reflected in continuous values of activity of FoxD-a/b. In this case, we can ask experimental biologists to measure the activity of FoxD-a/b by a more precise method, and we may be able to identify the diversity of cell states by different activity levels of FoxD-a/b. The second possibility is that the present understanding of the gene regulatory network of ascidiacea is not sufficient to explain the diversity of gene expressions at the tail bud stage. In other words, there may be unknown regulatory linkages which are important to generate the observed diversity of cell differentiation of ascidiacea.

4-2. Diversity of dynamical behaviors

Here we demonstrate the meaning of determining nodes directly by an in silico numerical simulation based on the gene regulatory network for ascidiacea development. We explained that the regulatory functions are unknown, although the regulatory relations between genes are determined experimentally. We consider purely hypothetical regulatory functions of the product form:

$$\dot{x}_k = f_k\left(\mathbf{x}_{I_k}\right) - x_k \qquad (5)$$

$$f_k = \prod_{j \in I_k} g_{j \to k}\left(x_j\right) \ ,$$

where $g_{j \to k}\left(x_j\right)$ is chosen randomly with probability 0.5 from the following two functions:

$$g^s_{j \to k}\left(x\right) = \begin{cases} 0.1 & \left(0 \le x < 0.2\right) \\ 0.3 & \left(0.2 \le x < 0.4\right) \\ 0.5 & \left(0.4 \le x < 0.6\right) \\ 0.7 & \left(0.6 \le x < 0.8\right) \\ 0.9 & \left(0.8 \le x\right) \end{cases} \qquad \text{(6a)}$$

or

$$g^c_{j \to k}\left(x\right) = 1. \qquad \text{(6b)}$$

We calculate the dynamics of the resulting gene activity of this system numerically using form (5), (6) and starting from randomly chosen initial values. The system may have multiple steady states depending on the choice of regulatory functions. However, the diversity of steady states may not be captured by measurement of some genes. The dynamics of some genes may show convergence to a single steady state while others show multiple solutions depending on the initial values.

We repeated the procedure: construction of regulatory functions by choosing (6a) or (6b) for each regulatory edge randomly, calculation of dynamics by changing initial states randomly, and search for steady states of the systems. We examined 1000 sets of random choices of regulatory functions, and 1000 different initial states for each set of functions. We omitted the cases where the whole system converged to a single

stationary point, i.e. the case that the system did not support steady state multiplicity. From this exhaustive numerical simulation we calculated the diversity of steady states captured by the observation of each gene in the network. The obtained results are summarized in Figure 5.

We can see that only the FoxD-a/b gene reflects steady state diversity of the whole system with 100% reliability. The observations of any other genes encounter a positive risk to miss the diversity of the total system. We confirm again that FoxD-a/b is the minimal feedback vertex set of this network. Note the difference in the reliability among the other vertices, which are not in the minimal feedback set. The Twist-like-1 gene rather strongly reflects the diversity of total system. On the other hand, the observation of the Otx gene runs a much higher risk to misrepresent the diversity of the total system. Our analysis indicates the Twist-like-1 gene as a second best candidate to detect the total diversity of steady states.

We also tried different $g_{j\to k}\left(x_j\right)$ given in the following form:

$$g_{j\to k}\left(x_j\right)=\begin{cases}0.5 & \left(0\le x<T_{j\to k}\right)\\ 1.0 & \left(T_{j\to k}\le x\right)\end{cases}, \qquad (6')$$

where $T_{j\to k}$ is a threshold of a step function, which is chosen randomly between 0 and 1 for each regulatory edge. We obtained qualitatively similar results with (6') replacing (6).

5. Signal transduction network

A variety of cell responses are induced by the surrounding environment or the signals from outside the cells. The signaling pathway downstream of the epidermal growth factor (EGF) receptor has been studied in mammalian cells. It has been shown to regulate a large diversity of cell responses including proliferation, migration, oncogenesis, and apoptosis. The process by which the growth factor signals induce cell reactions can be described roughly as follows. By the ligand-binding to the EGF receptor, the tyrosine-kinase activity of the receptor is induced. The activated EGF receptors phosphorylate and activate the target proteins. The activation of proteins causes activation of other species of molecules and the activation signal is transferred through a series of species of molecules, sequentially. The signal is finally transferred into the nucleus, regulates gene expressions and causes changes in macroscopic cell behavior. The process is called "signal transduction" and the pathway of transfer has been studied well. The whole pathway constitutes a complex system including famous and important sub-pathways such as the MAP kinase cascade, the PIP pathway, and the Ca^{2+} signaling cascade. Rather diverse reactions of the cell are produced by this signal transduction system. In other words, the signal transduction network is a system for the determination of the macroscopic behavior of mammalian cells after receiving the signal molecule stimulus from outside. Many studies of signal transduction focus on the crucial question how a single system produces multiple output responses depending on the stimulus inputs from outside the cells.

We analyze a regulatory network of signal transduction summarized by (Oda *et al*. 2005). The authors collected information on pathways of signal transduction from published papers, which are determined by various experimental methods. They summarized the information of regulation between molecules, and constructed a

complex regulatory network of 113 species of molecules including kinases, phosphatase, or ions like Ca^{2+}, and many regulatory edges between them. There is no self-regulatory edge in this network.

As a preprocessing we removed vertices without input or output as described in our above analysis for ascidiacea network. The resulting reduced network is still complex and possesses 61 vertices (Figure 6). We adopt a method of computer-aided search to determine a feedback vertex set of minimal size. Our search algorithm is simple, using only the definition of a feedback vertex set directly. We repeated the following examinations one by one exhaustively starting from smaller sizes of candidate sets to larger: (1) select a subset of vertices, (2) examine whether the network after the removal of the subset is cycle-free or not. We found 36 ways to select the feedback vertex set with a minimum of five vertices. Table 2 shows the possible choices of vertices to form a minimal feedback vertex set. The vertices in these sets are categorized into five groups, which are shown in different colors in Figure 6.

The equilibria of the system should be expressed as solutions of a system of equations including only a feedback vertex set. For example, if we select the feedback vertex set as $I = \{\text{ErbB11, SOS, c-Src, cyt } Ca^{2+}\text{, PI45-P2}\}$, the equilibria of the system are solutions of an equation system written in the form:

$$\begin{aligned}
x_{\text{ErbB11}} &= H_{\text{ErbB11}}\left(x_{\text{ErbB11}}, x_{\text{SOS}}, x_{\text{c-Src}}, x_{\text{PI4,5-P2}}\right) \\
x_{\text{SOS}} &= H_{\text{SOS}}\left(x_{\text{ErbB11}}, x_{\text{SOS}}, x_{\text{c-Src}}, x_{\text{cyt Ca2+}}, x_{\text{PI4,5-P2}}\right) \\
x_{\text{c-Src}} &= H_{\text{HB-EGF}}\left(x_{\text{ErbB11}}, x_{\text{c-Src}}, x_{\text{cyt Ca2+}}\right) \qquad . \qquad (7) \\
x_{\text{cyt Ca2+}} &= H_{\text{cyt Ca2+}}\left(x_{\text{cyt Ca2+}}, x_{\text{PI4,5-P2}}\right) \\
x_{\text{PI4,5-P2}} &= H_{\text{PI4,5-P2}}\left(x_{\text{ErbB11}}, x_{\text{SOS}}, x_{\text{c-Src}}, x_{\text{cyt Ca2+}}, x_{\text{PI4,5-P2}}\right)
\end{aligned}$$

The derivation of the precise form of the reduced equilibrium system (7) follows the same idea as the derivation of the reduced equilibrium system (4) in the ascidiacea network of section 4. In fact, the first equation of (4) would simply take the form $x_{Fox} = H_{Fox}(x_{Fox})$, in the present notation. Analogously, let I denote any minimal feedback vertex set of the signal transduction network and let $i \in I$ denote any feedback vertex. Then $x_i = H_i$, where the function H_i only depends on feedback vertex variables x_j for a subset $j \in J(i) \subseteq I$. What is this subset $J = J(i)$ of dependencies $x_i = H_i(\mathbf{x}_J)$, in the reduced system of equilibrium equations?

We define the elements $j \in J(i)$ by the following path property. We require there exists a directed path γ from j to i which does not traverse any other feedback vertices of I before reaching i. Let $J = J(i) \subseteq I$ denote the set of such feedback vertices j. For example $i \in J(i)$, because otherwise the set $I \setminus \{i\}$ would be a feedback vertex set more "minimal" than I: any directed cycle through i would also have to pass through another feedback vertex of $I \setminus \{i\}$.

Now we show how $x_i = H_i(\mathbf{x}_J)$ can be written as a function of only the feedback vertex variables x_j, $j \in J$, where the paths γ start. Let U denote the union of all such paths γ for all $j \in J(i)$. Then the regulatory network allows us to successively determine all equilibrium values x_k on vertices $k \in U \setminus J$, by successive evaluation of their regulatory functions, ultimately as functions of $\mathbf{x}_J$. In a final step

this determines $x_i = H_i\left(\mathbf{x}_J\right)$, as claimed. For the particular choice of $I = \{$ErbB11, SOS, c-Src, cyt Ca^{2+}, PI45-P2$\}$ above, this proves claim (7).

The dependencies of the reduced equilibrium system $x_i = H_i\left(\mathbf{x}_{J(i)}\right)$ on the feedback vertices $i \in I$ can be summarized in a reduced network. As vertices we choose a feedback vertex set I. We draw a directed edge from vertex j to vertex i if and only if $j \in J(i)$. In this sense, and only for steady states, we can reduce the network to smaller one including only the particular choices of feedback vertex sets as shown in Figure 7. In these figures, the edges indicate dependence between variables in the steady state systems analogous to (4). We emphasize that these reduced networks for steady states each have the same potential to generate the same diversity of steady states as the original network. However, they are not appropriate for considering other classes of dynamical behaviors, like periodic oscillations or quasi-periodic oscillations. We call them "steady state networks". As a first example we note how $i \in J(i)$ implies that every vertex set of the steady state network possesses a self-loop. See (4) and all examples in Figure 7.

The topology of the steady state network depends on the choice of a minimal FVS. We show these topologies of networks in Figure 7, and the correspondence between the choices of FVS and topologies in Table 3. Among them the network 1 is the simplest, with a minimal number of edges in its FVS. We may say that the FVS of ID number 1 in Table 2 gives the simplest reduced network, counting resulting edges. The vertices in one of these FVS are marked by circles in Figure 6.

The regulatory network of signal transduction is expected to show a broad variety of dynamic responses in the global attractor depending on the stimulus signals

from outside the cell. Measuring the time tracks of five feedback vertices experimentally in different environments, after receiving the stimulus signals, will faithfully represent the diversity of the dynamical response of the whole system. If we discover, on the other hand, that the time tracks of five feedback vertices are not sufficient to explain all of the observed behaviors, then we will be forced to conclude that the original network again missed some important edges or molecules.

6. Control of mammalian circadian rhythms.

In this example, we explore a control aspect of our theory. We demonstrate that the dynamics of the whole system can be controlled by prescribing the dynamics of only a feedback vertex set. Mammalian circadian rhythms in mice have been studied well, experimentally. Four major genes are involved in the system: Per1, Per2, Cry1 and Cry2. The regulations between genes and the interactions between these proteins have been examined in detail. The system in a normal animal exhibits periodically oscillating gene activities. Many mathematical models for the system have been proposed and studied. Of course mathematical models include assumptions on experimentally unverified facts, in particular in the specific formulae for the precise regulatory functions f_k. In some studies models were analyzed mathematically or numerically and conditions for periodic oscillations were determined.

For our numerical experiments we use a mathematical model proposed by Mirsky *et al*. (2009), which includes 21 variables and hundreds of parameters. The ordinary differential equations and our choice of parameter values are detailed in Appendix B. The regulatory network is shown in Figure 8a. Our choice of a minimal feedback vertex set with size 7 is $I = \{$PER1, PER2, CRY1, CRY2, RORc,

CLK, BMAL1}, which is different from our companion paper (Fiedler *et al*., 2013). We found that the dynamics of the model possesses several invariant sets including two stable periodic oscillations (P1 and P2), one unstable periodic oscillation (UP), and one unstable stationary point (USS) under a choice of parameter values, which are different from the original values used in Mirsky *et al*. (2009). Figure 8c shows the trajectories of these asymptotic behaviors in two dimensional phase space, *Per1* mRNA and *Per2* mRNA.

We performed four numerical experiments, controlling "from P1 to P2", "from P2 to P1", "from P1 to UP" and "from P1 to USS" (Figure 9). We examined whether the system is controlled by prescribing the time tracks of the 7 informative variables $\mathbf{x}_I(t)$ on the feedback vertex set I. As a preparation we calculated the time tracks of each informative variable x_k, $k \in I$, on the four invariant sets, P1, P2, UP and USS, by direct numerical simulation, i.e. $x_k^{\mathrm{P1}}(t)$, $x_k^{\mathrm{P2}}(t)$, $x_k^{\mathrm{UP}}(t)$ and $x_k^{\mathrm{USS}}(t)$ $(0 \le t \le T)$.

The control protocol of our numerical in silico experiment called "from P1 to P2" is the following: the time tracks of the 7 informative variables are prescribed to follow their values $\mathbf{x}_I^{\mathrm{P2}}(t)$, as on P2. The dynamics of the remaining 14 variables $x_k(t)$, $k \notin I$, are calculated by the remaining 14 ODEs of the system, and the initial state of these remaining variables is chosen to coincide with a point on the P1 trajectory. We used different points on the P1 orbit as initial states. The results did not depend on the particular choice of the initial state as much as we examined. We observed how the dynamical trajectory of the remaining variables starting from P1 left that stable periodic orbit, immediately, and quickly converged to the competing P2 orbit. The

total system finally shows periodic oscillation on the P2 orbit.

Similarly we examined the opposite control protocol “from P2 to P1”, where the tracks of the 7 informative variables in the feedback vertex set are now prescribed to follow their values $\mathbf{x}_I^{\mathrm{P1}}(t)$ on P1, and the dynamics of the remaining 14 variables are calculated by the remaining ODEs with an initial state on P2. In the experiment we observed that the remaining system immediately left P2, this time, and quickly converged to P1 following the prescribed informative dynamics $\mathbf{x}_I^{\mathrm{P1}}(t)$.

Next we examined the control protocol "from P1 to UP", where the tracks of the 7 informative variables in the feedback vertex set I are prescribed to follow their values $\mathbf{x}_I^{\mathrm{UP}}(t)$ on the unstable periodic orbit UP, and the dynamics of the remaining 14 variables are calculated by the remaining ODEs with an initial state on P1. It is interesting that the total system turned out to converge to UP even though periodic oscillation UP was unstable, originally.

Analogously we examined the numerical control protocol “from P1 to USS”: the 7 informative variables $\mathbf{x}_I(t)$ in the feedback vertex set are fixed at their constant values $\mathbf{x}_I^{\mathrm{USS}}$ of the unstable stationary point USS, and the remaining 14 variables are calculated by the remaining ODEs with an initial state on the stable periodic orbit P1. Even though USS was unstable, originally, we again found that the remaining variables left the stable periodic orbit P1, immediately. The total system then converged to the unstable stationary point $\mathbf{x}^{\mathrm{USS}}$ and remained there, by the continued clamping control of all 7 informative vertices.

We conclude that control of the feedback vertex set is indeed sufficient to

control the total system towards a previously existing target state, even when the target state is unstable, originally. This is a direct consequence of our theorem on determining nodes via the following interpretation: "control of the feedback vertex set implies control of the whole system". In the companion paper (Fiedler *et al*., 2013) this interpretation is proved in a setting where the remaining ODEs are viewed as a nonautonomous regulatory network with an empty (remaining) feedback vertex set. The original feedback vertex set I then serves as a nonautonomous input.

We next examined whether the system can be controlled equally well by a non-full set of vertices. We control only 6 vertices I' among the 7 of the full feedback vertex set I and tried to control the system “from P1 to P2”. We prescribed the time tracks $\mathbf{x}_{I'}^{\mathrm{P2}}(t)$, $I' = \{$PER1, PER2, CRY1, CRY2, RORc, BMAL1$\}$, i.e. skipping CLK. The dynamics of the remaining 15 variables, including the skipped informative node CLK, are calculated by ODEs with an initial state on P1. The result is shown in Figure 10: the trajectory converged to an unknown spurious periodic oscillation, and not to P2. Similarly we performed numerical experiments “from P2 to P1”, “from P1 to UP” and “from P1 to USS”. The control of the system succeeded in "from P2 to P1" only, and failed in the cases “from P1 to UP” and "from P1 to USS". This demonstrates that controlling a non-informative set of feedback vertices set may not be sufficient to control the system. We caution our reader that the above results do depend on the choice of prescribed variables. If we control the time tracks $\mathbf{x}_{I''}(t)$ of suitable choice of 6 variables among 7 informative variables $I'' = \{$PER2, CRY1, CRY2, RORc, CLK, BMAL1$\}$, we can control the dynamics of the whole system.

Our in silico numerical control experiments demonstrate that the full feedback

vertex set is a sufficient set to control the whole regulatory network. Our result provides a rational criterion to select variables if we consider controlling complex systems which involve many variables. Our criterion does not depend on the particular choices of regulatory functions, and is based on the dependency structure of the regulatory network as a directed graph, only. The feedback vertex set criterion is quite powerful for biological systems, because biological systems are usually very complex and, in many cases, the regulatory edges are the only available information.

7. Other approaches

In this section we contrast our approach with two alternative recent view points which we find particularly interesting and illuminating. From the large literature on the subject of gene regulation and regulatory networks we focus on descriptions by continuous, quantitative variables. We ignore mere Boolean approaches. As more and more quantitative data are becoming available, the regulatory dependencies will be most appropriately represented in a regulatory network setting as described above.

The first alternative view point is that of linear, or linearization-based, standard control theory as adapted to regulatory networks by Liu, Slotine, and Barabasi (Liu *et al*., 2011; 2013). The other alternative which we discuss below is a recent adaptation of the celebrated 1971 Takens embedding theorem (Takens, 2010) to regulatory network by Joly (Joly, 2012).

The linearized setting of Liu *et al*. (2011; 2013) is the setting of standard linear control theory in the form

$$\begin{aligned} \dot{\mathbf{x}} &= A\mathbf{x} + B\mathbf{u} \\ \mathbf{y} &= C\mathbf{x} \end{aligned} \quad . \tag{8}$$

Here $\mathbf{x} = \mathbf{x}(t) \in \mathbb{R}^N$ describes the state vector of the system, $\mathbf{u} = \mathbf{u}(t) \in \mathbb{R}^K$ are the input (alias control, actuator) variables and $\mathbf{y} = \mathbf{y}(t) \in \mathbb{R}^M$ are the accessible observables (alias sensors, measurements) on which a successful control strategy $\mathbf{u}(t)$ is to be based. The matrices A, B, C are assumed constant. The regulatory graph Γ is defined such that the matrix $A = (a_{kj})$ of the uncontrolled network $\dot{\mathbf{x}} = A\mathbf{x}$ becomes the adjacency matrix of Γ. More precisely, Γ possesses an edge from vertex j to vertex k if and only if $a_{kj} \neq 0$. Our decay condition $d_k > 0$ or $\partial_1 F_k(x_k, \mathbf{x}_{I_k}) < 0$ requires $a_{kk} < 0$ for vertices k without self-loops – a restriction not imposed by (Liu *et al.*, 2011, 2013).

The nonlinear setting of Joly (2012) takes the general form

$$\dot{x}_k = G_k(\mathbf{x}_{I_k}). \qquad (9)$$

Self-loops $k \to k$ are allowed but not required. Self-decay is not required either. The strong results of Joly detailed below, however, do not hold for all nonlinearities G. Rather, they hold for some generic subset of nonlinearities G which is not explicitly known. In particular it is not known, but could be tested numerically or even experimentally, whether any specific nonlinearities of Michaelis-Menten or switching type fall into the class addressed by this mathematical result.

The observation objective of standard linear control theory in the setting (8) is the instantaneous observation and reconstruction of the complete system state

$\mathbf{x}(t) \in \mathbb{R}^N$ from the observation vector $\mathbf{y}(t) \in \mathbb{R}^M$ and its time derivatives $\dot{\mathbf{y}}(t)$,…, $\mathbf{y}^{(N-1)}(t)$. The standard sufficient condition for observability is the full rank N condition for the $NM \times N$ matrix with N blocks C, CA,…, CA^{N-1}, each of size $M \times N$. Liu *et al*. (2013) suggest a graphical approach (GA) and a maximal matching (MM) approach to achieve this. The maximal matching approach starts from a maximal set of edges $j \to k$ such that no pair of them shares the same start node j or end node k. The MM approach requires all terminal nodes of the maximal paths in the maximal edge set to be sensors. (We have reversed arrows in Liu *et al*. (2013) for consistency with our present paper.) The more parsimonious GA variant selects only one sensor per terminal strong linkage class of the network. Here any two vertices j, k are called strongly linked if they are connected by a directed path in the graph Γ, in either direction. A maximal strongly linked set is called class. A strong linkage class is called terminal if there do not exist any paths leaving it. In Liu *et al*. (2013) numerical evidence based on random networks was given to indicate that, in absence of "symmetries", the GA is sufficient to establish observability via the full rank condition of C, CA,…, CA^{N-1}. The precise meaning of "symmetries" was not specified.

Joly (2012) has obtained a precise mathematical version of a similar result in the nonlinear setting (9). Although his language is somewhat different, his results are based on the same choice of sensors as in the GA of Liu *et al*. (2013). Unlike the Takens embedding theorem (Takens, 2010), however, his results do not reconstruct all phase space $\mathbf{x} \in \mathbb{R}^N$. In fact, his sensors are proven to faithfully detect only asymptotically stationary and periodic orbits $\mathbf{x}(t)$ from finite time tracks $\mathbf{x}_I(t)$, $0 \le t \le T$, on the sensor set I. This drawback might be overcome in future work, in

our opinion. The genericity limitation mentioned above is more serious. In principle it might exclude linear problems, wholesale. Like (Liu *et al*., 2013), on the other hand, it promises strongly reduced sensor sets at least in systems with few output classes, alias terminal strong linkage classes.

The control objective of standard linear control theory in the setting (8) requires to solve any task of the following form: given any initial state $\mathbf{x}(0)=\mathbf{x}_0$ and any target state $\mathbf{x}(T)=\mathbf{x}_T$ at a target time $T>0$, steer $\mathbf{x}(t)$ from $\mathbf{x}_0$ to $\mathbf{x}_T$ according to $\dot{\mathbf{x}}=A\mathbf{x}+B\mathbf{u}$, for $0\le t\le T$, by a suitable choice of the control function $\mathbf{u}(t)\in\mathbb{R}^K$. The standard sufficient condition for controllability is the full rank N Kalman condition for the $N\times KN$ matrix with N blocks B, AB,…, $A^{N-1}B$ of size $N\times K$. The notion of structural controllability requires the system to be controllable, in this sense, for most choices of A compatible with the graph structure. The $N\times K$ control matrix B is assumed to consist of K unit column vectors e_{i_1},…, e_{i_K} with prescribed input $u_k(t)$ at vertex i_k of the network. As an input set $I=\{i_1,\ldots,i_K\}$ the unmatched vertices of the maximal matching (MM) approach are suggested in Liu *et al*. (2011). (In Liu *et al*. (2013) the direction of arrows coincides with our present paper.) By definition these are all vertices which do not appear as the end vertex k of any oriented edge $j\to k$ in the maximal matching. With this notion, some 37 networks from a large variety of applications are scanned for the resulting required number $K=|I|$ of controlled inputs. An interesting heuristics compares the results numerically and favorably to randomized directed graphs with the

same average number of in/out-degrees per vertex. An analogous generic result in the nonlinear setting is not available, at present.

Our nonlinear approach via feedback vertex sets is quite complementary to the above recent results in several respects. First, we do not impose any non-explicit genericity constraints on our regulatory network (1), (2). Instead we only impose a self-decay condition which can be circumvented by introducing self-loops $k \to k$, see Fiedler *et al*. (2013).

Second, we use the same feedback vertex set FVS for, both, observation and control. Our control philosophy, however, is quite different from standard control theory. In fact it is much closer to the idea of "reprogramming" regulatory networks as discussed by (Müller and Schuppert, 2011) and the reply of the authors of (Liu *et al*., 2011). We do not seek to steer the network state $\mathbf{x}(t)$ from any initial state $\mathbf{x}_0$ to any target state $\mathbf{x}_T$ in time $0 \le t \le T$. Instead we seek to steer $\mathbf{x}(t)$ from any initial state $\mathbf{x}_0$ to any stable or unstable target trajectory $\mathbf{x}^*(t) \in \mathbb{R}^N$, $t \in \mathbb{R}$, of the original regulatory network (1), (2), such that

$$\mathbf{x}(t) - \mathbf{x}^*(t) \to 0 \qquad (10)$$

for large times. We achieve this by clamping the input vertices $\mathbf{x}_I(t) := \mathbf{x}_I^*(t) \in \mathbb{R}^{|I|}$ to their previously observed trajectory on the feedback vertex set I, only. Then the remaining dynamics $\mathbf{x}(t)$ of the whole network must follow, as we proved in (Fiedler *et al*., 2013).

Third, our control is noninvasive as we reach the target trajectory $t \mapsto \mathbf{x}^*(t)$. This follows from the fact that we only seek to follow true, existing solutions $\mathbf{x}^*(\tau) \in \mathbb{R}^N$ in the regulatory network (1), (2). But, we have demonstrated our ability to choose freely among those, at least for stable and unstable stationary or periodic targets $\mathbf{x}^*(\tau)$, in the numerical example on circadian rhythms of section 6.

Forth, and perhaps most importantly, our control is model-free. Indeed we may just measure the required data $\mathbf{x}_I^*(\tau)$ on the same sensor set I which figures as input in our control version, at least when the target trajectory $\mathbf{x}^*$ is stable. The success (10) of our control is then guaranteed by the structure (1), (2) of the network alone, independently of any specific choices of nonlinearities. Only for an in silico demonstration of the method did we have to choose a specific model, with specific nonlinearities and specific parameter values, in section 6. But, the success by clamping $\mathbf{x}_I(t) := \mathbf{x}_I^*(t) \in \mathbb{R}^{|I|}$ on the full FVS I was guaranteed a priori, independently of any such numerical quibble.

Fifth and finally, observation of the feedback vertex set I is necessary to faithfully reconstruct all trajectories $\mathbf{x}^*(\tau)$, $\tau \in \mathbb{R}$ in the global attractor, from observations $\mathbf{x}_I^*(\tau)$ on any vertex subset I alone, if this is to be achieved for all nonlinearities F_k in (2) which are compatible with the network structure. This is the contents of the converse parts to our determinacy result in (Fiedler et al., 2013), as stated in section 3. We recall that any feedback vertex set I is determining. Conversely, any vertex set I which is determining for all choices of the nonlinearities f_k in (1) or F_k in (2) is necessarily a feedback vertex set. This converse aspect is

somewhat analogous, in the nonlinear case, to the notation of structural controllability advocated in (Liu *et al*., 2013) for the linear setting.

In summary, our nonlinear method is complementary to the linear approaches to observation and control by (Liu *et al*., 2011; 2013) and to the observational nonlinear genericity approach of Joly (2012). In particular we do not seek to steer the network dynamics $\mathbf{x}(t) \in \mathbb{R}^N$ from any initial state $\mathbf{x}_0$ to any target state $\mathbf{x}_T$. Instead we seek to steer the network dynamics $\mathbf{x}(t)$ to any pre-selected target trajectory $\mathbf{x}^*(\tau)$ of the regulatory network without any modeling knowledge on the system other than the regulatory network graph Γ.

8. Discussion

We presented three biological applications of a mathematical theory which distills an important aspect of the global dynamics from information on regulatory linkages alone. For complex regulatory networks we identified small feedback vertex sets (FVS), by measurements of which any recurrent dynamical behavior of whole system is assured to be identified. This includes steady states, periodic oscillations or quasi-periodic oscillations. In some examples, we discussed the possible inability of the networks to explain the diversity of the observed dynamics of the actual biology involved. Our theory also provides a rational criterion to select key molecules to control a system: the dynamics of whole system is sufficiently controlled by prescribing the dynamics on a FVS.

For mathematical proofs see the companion paper (Fiedler *et al*., 2013). We combine two mathematical concepts from different mathematical fields: "determining

nodes", i.e. a sufficient subset of variables which faithfully describe the long-term dynamics of the whole system, are determined as a "feedback vertex set" (FVS) of a regulatory network, in the sense of graph theory. Our theory does not require any quantitative information on the functions F_k of the regulatory ODE system (2). We only distinguish the case that the first partial derivative of F_k with respect to x_k is not strictly negative. This implies that the analysis of experimental data according to our theory returns experimental biology, directly. Any inability of the regulatory network to explain experimentally observed dynamics, directly indicates flawed experimental data, or incompleteness of the regulatory network in experimentally relevant parts.

In the previous paper (Mochizuki and Saito, 2010), we introduced the concept of "informative nodes" determined from regulatory linkages. We showed that all possible steady states of the whole system are identified by observing the stationary activity levels of the informative nodes, only. Now, and in the companion paper (Fiedler *et al*., 2013), we have improved three different aspects of that theory.

First, the theory applies to much wider classes of dynamical behaviors of ordinary differential equation (ODE) systems. We showed that the informative nodes faithfully trace, not only the steady states but also, the "full dynamics" of the network, including periodic oscillations, quasi-periodic oscillations, or bounded chaos.

The second improvement concerns the definition of "informative nodes". In the previous paper (Mochizuki and Saito, 2010), the key "informative nodes" vertices were not yet clearly connected to the directed graph structure of a regulatory network. We continue to use the term "informative nodes", for their dynamic relevance. The term is equivalent, however, to the term "feedback vertex set" which is originally a concept in graph theory. The "feedback vertex set" is directly defined from the

structure of a directed graph: a subset of vertices whose removal leaves the directed graph without directed cycles. We also introduced the term "determining nodes", which is a concept in dynamical systems. It is a subset of variables in a system such that the dynamical trajectories on that subset determine the long-term dynamics of all remaining variables. In the companion paper (Fiedler *et al*., 2013), we proved mathematically that a "feedback vertex set" of a regulatory network, alias a set of "informative nodes", is also a set of "determining nodes" for the full dynamics on the network.

The third improvement is the class of mathematical formulae of ODE systems. In the previous paper, we assumed linear functions for the decay of molecular activities. Here and in the companion (Fiedler *et al*., 2013), the decay of activities is expressed by any decreasing function F_k of the concentration of the molecules. In addition, we do not need to assume any separation of the decay term, which was used in the previous paper. Rather, the decay of activity and self-repression are now coherently expressed as a negative first partial derivative of the nonlinearity $F_k\left(x_k, \mathbf{x}_{I_k}\right)$ with respect to x_k. In other words, we do not distinguish between decay and self-repression in our dynamical point of view. This implies a significant practical improvement for the analysis of complex networks because we can ignore any loops of self-repression. This results in a substantial reduction of the size of the FVS. More importantly, it implies that our theory is applicable to any ODE systems, in principle, and is applicable, but not limited, to biological regulatory networks.

Our new concepts are useful to analyze regulatory networks in modern biology, in particular, because the information of dependence is abundant in biology, whereas further quantitative information remains elusive in many, if not most, systems. Some

biological systems are expected to present diverse dynamical behaviors depending on initial states or environmental factors. We can aim to measure the dynamics on a (small) feedback vertex set of the system, which should reflect the diversity of dynamics of the total system. If we find that the diversity observed on the feedback vertex set is insufficient to explain the expected diversity of the total system, it directly indicates some inconsistencies or some incompleteness of the available information. For example, there may be unknown regulatory cycles which require additional feedback vertices.

We presented three examples of our analysis of biological networks: a gene regulatory network for cell differentiation of ascidiacea, a signal transduction network, and a network for mammalian circadian rhythms. The dynamics of the gene regulatory network of ascidiacea is expected to produce many steady states corresponds to differentiated cell types. Actually the network has only one determining node. The present regulatory network in Imai *et al*. (2006) seems to be possibly incomplete, and may require unaccounted regulatory edges that would produce feedback loops which are not cut by FoxD-a/b.

The possible direct verification for our prediction is a knockdown experiment of FoxD-a/b gene. The regulatory network of ascidiacea will have an empty set of FVS by the removal of FoxD-a/b. The dynamics based on a network with empty FVS will converge to a unique equilibrium and will not show any diversity. If the present network of ascideacea is correct, the embryo with FoxD-a/b knockdown will not show any diversity of cell types.

In silico simulations with artificial regulatory functions for the ascidiacea network exhibited marked differences in the reliability to reflect diversity of steady

states, among nodes outside the feedback vertex set {FoxD-a/b}. The differences may reflect characteristics of the regulatory network. On the regulatory graph, more reliable genes seemed to be on shorter cycles through the feedback vertex, while less reliable genes seemed to be located on longer cycles. The relation between the structure of a network and the reliability of its vertices to faithfully predict the diversity of its dynamical behavior certainly deserves further study.

The system of signal transduction is expected to exhibit diverse dynamical responses depending on the initial state and corresponding to the stimulus signals from outside the cell. Many studies of signal transduction therefore focus on the question how a single system produces multiple outputs depending on the input stimulus from outside. Actually this interpretation of signal transduction networks is still hypothetical and is not yet confirmed by experiments. We need time-series data of the activities of key molecules to faithfully represent the dynamics of the full system. For this purpose we have to select molecules to track, because it is still difficult to simultaneously measure the activities of many molecules with a time-resolution sufficient for discussing their dynamics. Our theory again provides rational criteria to select those key molecules.

Our experiments demonstrate that the informative, or feedback vertex set is a sufficient set to control the whole regulatory network. Our result provides a criterion to select variables if we consider controlling complex systems which involve many variables. It is quite powerful for biological systems, because biological systems are usually very complex and available information is limited in many cases. Recent life sciences are trying to control biological system for medical purposes. The problems of circadian rhythms in human, for example, cause physiological or mental diseases,

including sleep difficulty or mental depression. Such problems may be solved if we successively control the activities of some carefully selected genes. Of course it will remain impossible to control all molecules in a circadian rhythm system. Thus we have to select minimal but sufficient sets of accessible molecules to control the system. Our theory may contribute to this ambitious goal, providing a rational criterion to identify key controlling molecules based on the graph information of their regulatory edges, alone.

We showed that our theory is useful to understand the dynamics of complex biological systems in several directions: to determine molecules whose activity should be measured, to understand the function of the system from time series of measurements of activities of some key molecules, and to identify unknown molecules or unknown regulatory linkages. By applying our theory to networks we can directly determine informative molecules, whose activity should be measured to understand the dynamics of the full system. By combining data of molecular activities, our theory would be a quite powerful tool in molecular biology to derive predictions on the existence of unknown molecules or of unknown regulatory linkages. We hope that our theory will contribute to molecular biology as a strong tool to derive such predictions, to support the rational selective acquisition of still missing biological data, and to further elucidate the complex and fascinating biological mechanisms of life.

Acknowledgements

This work was supported in part by the JST PRESTO program, CREST program of Japan, an HFSP program grant, and the Deutsche Forschungsgemeinschaft, SFB 910 "Control of Self-Organizing Nonlinear Sytems". Generous hospitality during extended very fruitful and enjoyable working visits which made this work possible is mutually acknowledged with particular gratitude. We express our sincere thanks to the following people for their useful comments: T. Akutsu, A. Azouani, G. Craciun, N. Go, Y. Iwasa, S. Ishihara, A. Ishimatsu, H. Kishino, H. Kokubu, S. Mano, H. Matano, T. Nishimori, Y. Nishiura, H. Oka, K. Okada, G. Rangel, N. Shigesada, M. Wakayama, and the referees.

Appendix A

Regulatory linkages of ascidiacea are summarized in the following:

ADMP -> {Achaete-Scute a-like2, CAGF9, Dll-C, FGF8/17/18, Irx-C, msxb, NK4, ZF (C2H2)-24}

AP-2-like2 -> {Emc}

Dll-B -> {Emx, FoxC, FoxH-a, GATA-b, MYTF, Six12, Six36, SOCS1/2/3/CIS, ZF (C2H2)-24}

DMRT1 -> {FoxC, meis, Six12, Six36}

ets/pointed2 -> {Brachyury, chordin, DMRT1, DUSP1.2.4.5, ELK, EphrinA-c, Fli/ERG1, Fos, FoxC, Mist, Mnx, msxb, MYTF, nodal, Otx, SMYD1, TWIST-like-1a/b, TWIST-like-2, ZF (C3H)}

FGF9/16/20 -> {Brachyury, chordin, COE, Delta-like, DMRT1, DUSP1.2.4.5, ELK, Emc, EphrinA-c, FGF8/17/18, Fli/ERG1, Fos, FoxB, FoxC, Hex, Jun, LAG1-like 5, Mist, Mnx, msxb, MyoD, MYTF, Neurogenin, nodal, noggin, NoTrlc, Otx, Pax3/7, Pax6, TTF1, TWIST-like-1a/b, TWIST-like-2, ZF (C3H), ZicL}

FoxA-a -> {Brachyury, chordin, Delta-like, DMRT1, DUSP1.2.4.5, Emx, Eph1, FoxB, FoxC, Fz4, GATA-a, GATA-b, Lhx3, Mnx, msxb, MYTF, nodal, NoTrlc, Otx, Pax3/7, Pax6, sFRP1/5, sFRP3/4-b, SoxB1, TTF1, TWIST-like-1a/b, TWIST-like-2, ZF (C2H2)-33, ZicL}

FoxB -> {Cdx, FGF8/17/18, Mnx, Pax6}

FoxC -> {ZF (C2H2)-2}

FoxD-a/b -> {Brachyury, chordin, COE, Delta-like, Dll-B, DMRT1,

FGF8/17/18, FGF9/16/20, FoxB, Mnx, MYTF, Neurogenin, nodal, NoTrlc, Pax6, TWIST-like-1a/b, Wnt5, ZF (C2H2)-33, ZicL}

Mesp -> {FoxF, NK4, NoTrlc, Tolloid}

msxb -> {Achaete-Scute a-like2, CAGF9, Dll-C, SoxB2, ZF (C2H2)-24}

MyoD -> {Mox, Otp, SMYD1}

Neurogenin -> {COE, Delta-like, FGF8/17/18, MYTF}

nodal -> {chordin, COE, Delta-like, E(spl)/hairy-b, Emc, FGF8/17/18, FoxC, Lmx, msxb, MYTF, Neurogenin, NoTrlc, Pax3/7, Pax6, Snail, SOCS1/2/3/CIS}

NoTrlc -> {TWIST-like-1a/b}

Otx -> {Fli/ERG1, Fos, FoxC, FoxD-a/b, GATA-a, Hex, Jun, LAG1-like 5, meis, Mist, MYTF, Six12, Six36, TWIST-like-1a/b, TWIST-like-2}

Snail -> {Mnx, MYTF}

SoxC -> {Cdx, Delta-like, DMRT1, FGF8/17/18, FoxC, GATA-b, MYTF, Neurogenin, nodal, Pax6, Snail}

Tbx6b/c/d -> {Mnx, MyoD, Otp, SMYD1, Snail}

TWIST-like-1a/b -> {Fli/ERG1, Fos, FoxD-a/b, Hex, Hlx, Hox4, LAG1-like 5, Mist, TWIST-like-2}

ZicL -> {Brachyury, Cdx, chordin, COE, Delta-like, Fos, Lhx3, Lmx, meis, Mnx, MyoD, MYTF, Neurogenin, Otp, Pax6, Six12, Six36, SMYD1, Snail, Tbx6a, Tbx6b/c/d, TWIST-like-1a/b, Wnt5, ZF (C2H2)-33}

Appendix B

Model equations for mammalian circadian rhythm.

The mathematical model which we used in Section 6 is written as a system of ODEs including 21 variables, *Per1*, *Per2*, *Cry1*, *Cry2*, *Rev-erb*, *Clk*, *Bmal1*, *Rorc*, PER1, PER2, CRY1, CRY2, REV-ERB, CLK, BMAL1, RORc, PER1/CRY1, PER2/CRY1, PER1/CRY2, PER2/CRY2 and CLK/BMAL1.

$$\begin{aligned}\frac{dPer1}{dt} = &\left(v_{0,Per1} + v_{1,Per1} \frac{\text{CLK/BMAL1}^{na_{1,Per1}}}{KA_{1,Per1}^{na_{1,Per1}} + \text{CLK/BMAL1}^{na_{1,Per1}}} \right) \\ &\times \frac{KI_{1,Per1}^{ni_{1,Per1}}}{KI_{1,Per1}^{ni_{1,Per1}} + \text{PER1/CRY1}^{ni_{1,Per1}}} \frac{KI_{2,Per1}^{ni_{2,Per1}}}{KI_{2,Per1}^{ni_{2,Per1}} + \text{PER1/CRY2}^{ni_{2,Per1}}} \\ &\times \frac{KI_{3,Per1}^{ni_{3,Per1}}}{KI_{3,Per1}^{ni_{3,Per1}} + \text{PER2/CRY1}^{ni_{3,Per1}}} \frac{KI_{4,Per1}^{ni_{4,Per1}}}{KI_{4,Per1}^{ni_{4,Per1}} + \text{PER2/CRY2}^{ni_{4,Per1}}} \\ &- k_{m,Per1} Per1\end{aligned} \tag{B.1}$$

$$\begin{aligned}\frac{dPer2}{dt} = &\left(v_{0,Per2} + v_{1,Per2} \frac{\text{CLK/BMAL1}^{na_{1,Per2}}}{KA_{1,Per2}^{na_{1,Per2}} + \text{CLK/BMAL1}^{na_{1,Per2}}} \right) \\ &\times \frac{KI_{1,Per2}^{ni_{1,Per2}}}{KI_{1,Per2}^{ni_{1,Per2}} + \text{PER1/CRY1}^{ni_{1,Per2}}} \frac{KI_{2,Per2}^{ni_{2,Per2}}}{KI_{2,Per2}^{ni_{2,Per2}} + \text{PER1/CRY2}^{ni_{2,Per2}}} \\ &\times \frac{KI_{3,Per2}^{ni_{3,Per2}}}{KI_{3,Per2}^{ni_{3,Per2}} + \text{PER2/CRY1}^{ni_{3,Per2}}} \frac{KI_{4,Per2}^{ni_{4,Per2}}}{KI_{4,Per2}^{ni_{4,Per2}} + \text{PER2/CRY2}^{ni_{4,Per2}}} \\ &- k_{m,Per2} Per2\end{aligned} \tag{B.2}$$

$$\frac{dCry1}{dt} = \left(v_{0,Cry1} + v_{1,Cry1} \frac{\text{CLK/BMAL1}^{na_{1,Cry1}}}{KA_{1,Cry1}^{na_{1,Cry1}} + \text{CLK/BMAL1}^{na_{1,Cry1}}} + v_{2,Cry1} \frac{\text{RORc}^{na_{2,Cry1}}}{KA_{2,Cry1}^{na_{2,Cry1}} + \text{RORc}^{na_{2,Cry1}}} \right) \frac{KI_{1,Cry1}^{ni_{1,Cry1}}}{KI_{1,Cry1}^{ni_{1,Cry1}} + \text{PER1/CRY1}^{ni_{1,Cry1}}} \times \frac{KI_{2,Cry1}^{ni_{2,Cry1}}}{KI_{2,Cry1}^{ni_{2,Cry1}} + \text{PER1/CRY2}^{ni_{2,Cry1}}} \frac{KI_{3,Cry1}^{ni_{3,Cry1}}}{KI_{3,Cry1}^{ni_{3,Cry1}} + \text{PER2/CRY1}^{ni_{3,Cry1}}} \times \frac{KI_{4,Cry1}^{ni_{4,Cry1}}}{KI_{4,Cry1}^{ni_{4,Cry1}} + \text{PER2/CRY2}^{ni_{4,Cry1}}} \frac{KI_{5,Cry1}^{ni_{5,Cry1}}}{KI_{5,Cry1}^{ni_{5,Cry1}} + \text{REV-ERB}\alpha^{ni_{5,Cry1}}} - k_{m,Cry1} Cry1 \quad \text{(B.3)}$$

$$\frac{dCry2}{dt} = \left(v_{0,Cry2} + v_{1,Cry2} \frac{\text{CLK/BMAL1}^{na_{1,Cry2}}}{KA_{1,Cry2}^{na_{1,Cry2}} + \text{CLK/BMAL1}^{na_{1,Cry2}}} + v_{2,Cry2} \frac{\text{RORc}^{na_{2,Cry2}}}{KA_{2,Cry2}^{na_{2,Cry2}} + \text{RORc}^{na_{2,Cry2}}} \right) \frac{KI_{1,Cry2}^{ni_{1,Cry2}}}{KI_{1,Cry2}^{ni_{1,Cry2}} + \text{PER1/CRY1}^{ni_{1,Cry2}}} \times \frac{KI_{2,Cry2}^{ni_{2,Cry2}}}{KI_{2,Cry2}^{ni_{2,Cry2}} + \text{PER1/CRY2}^{ni_{2,Cry2}}} \frac{KI_{3,Cry2}^{ni_{3,Cry2}}}{KI_{3,Cry2}^{ni_{3,Cry2}} + \text{PER2/CRY1}^{ni_{3,Cry2}}} \times \frac{KI_{4,Cry2}^{ni_{4,Cry2}}}{KI_{4,Cry2}^{ni_{4,Cry2}} + \text{PER2/CRY2}^{ni_{4,Cry2}}} \frac{KI_{5,Cry2}^{ni_{5,Cry2}}}{KI_{5,Cry2}^{ni_{5,Cry2}} + \text{REV-ERB}\alpha^{ni_{5,Cry2}}} - k_{m,Cry2} Cry2 \quad \text{(B.4)}$$

$$\frac{d\,\text{Re}v\text{-}erv\alpha}{dt} = v_{1,\text{Re}v\text{-}erv\alpha} \frac{\text{CLK/BMAL1}^{na_{1,\text{Re}v\text{-}erv\alpha}}}{KA_{1,\text{Re}v\text{-}erv\alpha}^{na_{1,\text{Re}v\text{-}erv\alpha}} + \text{CLK/BMAL1}^{na_{1,\text{Re}v\text{-}erv\alpha}}} \times \frac{KI_{1,\text{Re}v\text{-}erv\alpha}^{ni_{1,\text{Re}v\text{-}erv\alpha}}}{KI_{1,\text{Re}v\text{-}erv\alpha}^{ni_{1,\text{Re}v\text{-}erv\alpha}} + \text{PER1/CRY1}^{ni_{1,\text{Re}v\text{-}erv\alpha}}} \frac{KI_{2,\text{Re}v\text{-}erv\alpha}^{ni_{2,\text{Re}v\text{-}erv\alpha}}}{KI_{2,\text{Re}v\text{-}erv\alpha}^{ni_{2,\text{Re}v\text{-}erv\alpha}} + \text{PER1/CRY2}^{ni_{2,\text{Re}v\text{-}erv\alpha}}} \times \frac{KI_{3,\text{Re}v\text{-}erv\alpha}^{ni_{3,\text{Re}v\text{-}erv\alpha}}}{KI_{3,\text{Re}v\text{-}erv\alpha}^{ni_{3,\text{Re}v\text{-}erv\alpha}} + \text{PER2/CRY1}^{ni_{3,\text{Re}v\text{-}erv\alpha}}} \frac{KI_{4,\text{Re}v\text{-}erv\alpha}^{ni_{4,\text{Re}v\text{-}erv\alpha}}}{KI_{4,\text{Re}v\text{-}erv\alpha}^{ni_{4,\text{Re}v\text{-}erv\alpha}} + \text{PER2/CRY2}^{ni_{4,\text{Re}v\text{-}erv\alpha}}} - k_{m,\text{Re}v\text{-}erv\alpha}\, \text{Re}v\text{-}erv\alpha \quad \text{(B.5)}$$

$$\frac{dClk}{dt} = \left(v_{0,Clk} + v_{1,Clk} \frac{\mathrm{RORc}^{na_{1,Clk}}}{KA_{1,Clk}^{na_{1,Clk}} + \mathrm{RORc}^{na_{1,Clk}}} \right) \times \frac{KI_{1,Clk}^{ni_{1,Clk}}}{KI_{1,Clk}^{ni_{1,Clk}} + \text{REV-ERB}\alpha^{ni_{1,Clk}}} - k_{m,Clk} Clk \tag{B.6}$$

$$\frac{dBmal1}{dt} = \left(v_{0,Bmal1} + v_{1,Bmal1} \frac{\mathrm{RORc}^{na_{1,Bmal1}}}{KA_{1,Bmal1}^{na_{1,Bmal1}} + \mathrm{RORc}^{na_{1,Bmal1}}} \right) \times \frac{KI_{1,Bmal1}^{ni_{1,Bmal1}}}{KI_{1,Bmal1}^{ni_{1,Bmal1}} + \text{REV-ERB}\alpha^{ni_{1,Bmal1}}} - k_{m,Bmal1} Bmal1 \tag{B.7}$$

$$\begin{aligned} \frac{dRorc}{dt} = & \left(v_{0,Rorc} + v_{1,Rorc} \frac{\text{CLK/BMAL1}^{na_{1,Rorc}}}{KA_{1,Rorc}^{na_{1,Rorc}} + \text{CLK/BMAL1}^{na_{1,Rorc}}} \right. \\ & \left. + v_{2,Rorc} \frac{\mathrm{RORc}^{na_{2,Rorc}}}{KA_{2,Rorc}^{na_{2,Rorc}} + \mathrm{RORc}^{na_{2,Rorc}}} \right) \frac{KI_{1,Rorc}^{ni_{1,Rorc}}}{KI_{1,Rorc}^{ni_{1,Rorc}} + \text{PER1/CRY1}^{ni_{1,Rorc}}} \\ & \times \frac{KI_{2,Rorc}^{ni_{2,Rorc}}}{KI_{2,Rorc}^{ni_{2,Rorc}} + \text{PER1/CRY2}^{ni_{2,Rorc}}} \frac{KI_{3,Rorc}^{ni_{3,Rorc}}}{KI_{3,Rorc}^{ni_{3,Rorc}} + \text{PER2/CRY1}^{ni_{3,Rorc}}} \\ & \times \frac{KI_{4,Rorc}^{ni_{4,Rorc}}}{KI_{4,Rorc}^{ni_{4,Rorc}} + \text{PER2/CRY2}^{ni_{4,Rorc}}} \frac{KI_{5,Rorc}^{ni_{5,Rorc}}}{KI_{5,Rorc}^{ni_{5,Rorc}} + \text{REV-ERB}\alpha^{ni_{5,Rorc}}} \\ & - k_{m,Rorc} Rorc \end{aligned} \tag{B.8}$$

$$\begin{aligned} \frac{d\text{PER1}}{dt} = & t_{Per1} Per1 - a_{\text{PER1,CRY1}} \text{PER1} \times \text{CRY1} - a_{\text{PER1,CRY2}} \text{PER1} \times \text{CRY2} \\ & + d_{\text{PER1/CRY1}} \text{PER1/CRY1} + d_{\text{PER1/CRY2}} \text{PER1/CRY2} - k_{p,\text{PER1}} \text{PER1} \end{aligned} \tag{B.9}$$

$$\begin{aligned} \frac{d\text{PER2}}{dt} = & t_{Per2} Per2 - a_{\text{PER2,CRY1}} \text{PER2} \times \text{CRY1} - a_{\text{PER2,CRY2}} \text{PER2} \times \text{CRY2} \\ & + d_{\text{PER2/CRY1}} \text{PER2/CRY1} + d_{\text{PER2/CRY2}} \text{PER2/CRY2} - k_{p,\text{PER2}} \text{PER2} \end{aligned} \tag{B.10}$$

$$\frac{d\text{CRY1}}{dt} = t_{Cry1}Cry1 - a_{\text{PER1,CRY1}}\text{PER1}\times\text{CRY1} - a_{\text{PER2,CRY1}}\text{PER2}\times\text{CRY1} + d_{\text{PER1/CRY1}}\text{PER1/CRY1} + d_{\text{PER2/CRY1}}\text{PER2/CRY1} - k_{p,\text{CRY1}}\text{CRY1} \quad \text{(B.11)}$$

$$\frac{d\text{CRY2}}{dt} = t_{Cry2}Cry2 - a_{\text{PER1,CRY2}}\text{PER1}\times\text{CRY2} - a_{\text{PER2,CRY2}}\text{PER2}\times\text{CRY2} + d_{\text{PER1/CRY2}}\text{PER1/CRY2} + d_{\text{PER2/CRY2}}\text{PER2/CRY2} - k_{p,\text{CRY2}}\text{CRY2} \quad \text{(B.12)}$$

$$\frac{d\text{REV-ERB}\alpha}{dt} = t_{\text{Re}v\text{-}erb\alpha}\ \text{Re}v\text{-}erb\alpha - k_{p,\text{REV-ERB}\alpha}\text{REV-ERB}\alpha \quad \text{(B.13)}$$

$$\frac{d\text{CLK}}{dt} = t_{Clk}Clk - a_{\text{CLK,BMAL1}}\text{CLK}\times\text{BMAL1} + d_{\text{CLK/BMAL1}}\text{CLK/BMAL1} - k_{p,\text{CLK}}\text{CLK} \quad \text{(B.14)}$$

$$\frac{d\text{BMAL1}}{dt} = t_{Bmal1}Bmal1 - a_{\text{CLK,BMAL1}}\text{CLK}\times\text{BMAL1} + d_{\text{CLK/BMAL1}}\text{CLK/BMAL1} - k_{p,\text{BMAL1}}\text{BMAL1} \quad \text{(B.15)}$$

$$\frac{d\text{RORc}}{dt} = t_{Rorc}Rorc - k_{p,\text{RORc}}\text{RORc} \quad \text{(B.16)}$$

$$\frac{d\text{PER1/CRY1}}{dt} = a_{\text{PER1,CRY1}}\text{PER1}\times\text{CRY1} - d_{\text{PER1/CRY1}}\text{PER1/CRY1} \quad \text{(B.17)}$$

$$\frac{d\text{PER2/CRY1}}{dt} = a_{\text{PER2,CRY1}}\text{PER2}\times\text{CRY1} - d_{\text{PER2/CRY1}}\text{PER2/CRY1} \quad \text{(B.18)}$$

$$\frac{d\text{PER1/CRY2}}{dt} = a_{\text{PER1,CRY2}}\text{PER1}\times\text{CRY2} - d_{\text{PER1/CRY2}}\text{PER1/CRY2} \quad \text{(B.19)}$$

$$\frac{d\text{PER2/CRY2}}{dt} = a_{\text{PER2,CRY2}}\text{PER2}\times\text{CRY2} - d_{\text{PER2/CRY2}}\text{PER2/CRY2} \qquad \text{(B.20)}$$

$$\frac{d\text{CLK/BMAL1}}{dt} = a_{\text{CLK,BMAL1}}\text{CLK}\times\text{BMAL1-}d_{\text{CLK/BMAL1}}\text{CLK/BMAL1} \qquad \text{(B.21)}$$

Our choice of parameter values are as follows:

$v_{0,Per1}$=0.000001, $v_{1,Per1}$=3.0, $v_{0,Per2}$=0.09, $v_{1,Per2}$=3.29, $v_{0,Cry1}$=0.26, $v_{1,Cry1}$=2.44, $v_{2,Cry1}$=2.89, $v_{0,Cry2}$=1.29, $v_{1,Cry2}$=2.72, $v_{2,Cry2}$=0.1, $v_{1,\text{Re}v\text{-}erb\alpha}$=11.06, $v_{0,Clk}$=3.98, $v_{1,Clk}$=3.36, $v_{0,Bmal1}$=1.98, $v_{1,Bmal1}$=4.12, $v_{0,Rorc}$=0.06, $v_{1,Rorc}$=3.55, $v_{2,Rorc}$=0.46,

$na_{1,Per1}$=2.0, $ni_{1,Per1}$=2.0, $ni_{2,Per1}$=1.0, $ni_{3,Per1}$=2.0, $ni_{4,Per1}$=4.0, $na_{1,Per2}$=10.0, $ni_{1,Per2}$=1.0, $ni_{2,Per2}$=1.0, $ni_{3,Per2}$=9.0, $ni_{4,Per2}$=8.0, $na_{1,Cry1}$=4.91, $na_{2,Cry1}$=3.01, $ni_{1,Cry1}$=1.0, $ni_{2,Cry1}$=1.0, $ni_{3,Cry1}$=6.0, $ni_{4,Cry1}$=4.0, $ni_{5,Cry1}$=2.24, $na_{1,Cry2}$=4.39, $na_{2,Cry2}$=4.43, $ni_{1,Cry2}$=1.0, $ni_{2,Cry2}$=1.0, $ni_{3,Cry2}$=4.0, $ni_{4,Cry2}$=8.0, $ni_{5,Cry2}$=1.75, $na_{1,\text{Re}v\text{-}erb\alpha}$=4.40, $ni_{1,\text{Re}v\text{-}erb\alpha}$=0.15, $ni_{2,\text{Re}v\text{-}erb\alpha}$=0.3, $ni_{3,\text{Re}v\text{-}erb\alpha}$=7.0, $ni_{4,\text{Re}v\text{-}erb\alpha}$=7.0, $na_{1,Clk}$=3.50, $ni_{1,Clk}$=1.96, $na_{1,Bmal1}$=4.13, $ni_{1,Bmal1}$=0.02, $na_{1,Rorc}$=1.57, $na_{2,Rorc}$=0.56, $ni_{1,Rorc}$=1.0, $ni_{2,Rorc}$=1.0, $ni_{3,Rorc}$=7.0, $ni_{4,Rorc}$=7.0, $ni_{5,Rorc}$=4.33,

$KA_{1,Per1}$=1.98, $KI_{1,Per1}$=1.07, $KI_{2,Per1}$=3.96, $KI_{3,Per1}$=1.68, $KI_{4,Per1}$=3.11, $KA_{1,Per2}$ =1.90, $KI_{1,Per2}$=4.51, $KI_{2,Per2}$=2.98, $KI_{3,Per2}$=2.24, $KI_{4,Per2}$=3.31, $KA_{1,Cry1}$=1.46, $KA_{2,Cry1}$=3.76, $KI_{1,Cry1}$=0.03, $KI_{2,Cry1}$=0.77, $KI_{3,Cry1}$=3.59, $KI_{4,Cry1}$=3.44, $KI_{5,Cry1}$ =2.82, $KA_{1,Cry2}$=0.69, $KA_{2,Cry2}$=2.96, $KI_{1,Cry2}$=4.63, $KI_{2,Cry2}$=2.95, $KI_{3,Cry2}$=3.57, $KI_{4,Cry2}$=2.75, $KI_{5,Cry2}$=3.97, $KA_{1,\mathrm{Rev\text{-}erb}\alpha}$=3.15, $KI_{1,\mathrm{Rev\text{-}erb}\alpha}$=3.56, $KI_{2,\mathrm{Rev\text{-}erb}\alpha}$=3.62, $KI_{3,\mathrm{Rev\text{-}erb}\alpha}$=4.71, $KI_{4,\mathrm{Rev\text{-}erb}\alpha}$=1.23, $KA_{1,Clk}$=1.59, $KI_{1,Clk}$=0.83, $KA_{1,Bmal1}$=2.59, $KI_{1,Bmal1}$=2.47, $KA_{1,Rorc}$=4.30, $KA_{2,Rorc}$=4.89, $KI_{1,Rorc}$=3.49, $KI_{2,Rorc}$=2.34, $KI_{3,Rorc}$ =2.71, $KI_{4,Rorc}$=2.09, $KI_{5,Rorc}$=3.36,

$k_{m,Per1}$=2.18, $k_{m,Per2}$=0.20, $k_{m,Cry1}$=0.22, $k_{m,Cry2}$=0.41, $k_{m,\mathrm{Rev\text{-}erv}\alpha}$=0.60, $k_{m,Clk}$=3.19, $k_{m,Bmal1}$=1.42, $k_{m,Rorc}$=1.50, $k_{p,\mathrm{PER1}}$=2.58, $k_{p,\mathrm{PER2}}$=3.0, $k_{p,\mathrm{CRY1}}$=0.312, $k_{p,\mathrm{CRY2}}$=5.9, $k_{p,\mathrm{REV\text{-}ERB}\alpha}$=0.31, $k_{p,\mathrm{CLK}}$=1.52, $k_{p,\mathrm{BMAL1}}$=2.28, $k_{p,\mathrm{RORc}}$=3.33, t_{Per1}=3.05, t_{Per2}=2.38, t_{Cry1}=3.94, t_{Cry2}=1.69, $t_{\mathrm{Rev\text{-}erb}\alpha}$=1.60, t_{Clk}=3.04, t_{Bmal1}=4.00, t_{Rorc}=1.39, $a_{\mathrm{PER1,CRY1}}$ =3.57, $a_{\mathrm{PER1,CRY2}}$=3.12, $a_{\mathrm{PER2,CRY1}}$=3.81, $a_{\mathrm{PER2,CRY2}}$=4.0, $a_{\mathrm{CLK,BMAL1}}$=1.98, $d_{\mathrm{PER1/CRY1}}$ =1.32, $d_{\mathrm{PER1/CRY2}}$=1.85, $d_{\mathrm{PER2/CRY1}}$=1.37, $d_{\mathrm{PER2/CRY2}}$=2.42, $d_{\mathrm{CLK/BMAL1}}$=0.97.

We calculated the dynamics of the model by Euler time steps Δt=0.001.

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 Informatics* **9**, 151-160.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C. H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust, A. G., Pan, Z. J., Schilstra, M. J., Clarke, P. J. C., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L., Bolouri, H. (2002). A genomic regulatory network for development. *Science* 295, 1669-1678.

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. Dyn. Differential Eqns.* (In Press)

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.

Foias, C. and Titi, E.S. (1991) Determining nodes, finite differences schemes and inertial manifolds. *Nonlinearity* **4**, 135-153.

Glass, L., Kauffman, S. A. (1973) The logical analysis of continuous, non-linear biochemical control networks. *J. theor. Biol.* **39**, 103-129.

Imai, K. S., Levine, M., Satoh, N. and Satou Y. (2006). Regulatory Blueprint for a Chordate Embryo. *Science* **312**, 1183-1187.

Joly B. (2012) Observation and inverse problems in coupled cell networks. *Nonlinearity* **25**, 657-676.

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

Liu, Y.Y., Slotine, J.J. and Barabasi, A.L. (2013) Observability of complex systems, *Proc. Natl. Acad. Sci. USA* **110**, 2460-2465.

Mirsky, H. P., Liu, A. C., Welsh, D. K., Kay, S. A. and Doyle III, F. J. (2009) A model of the cell-autonomous mammalian circadian clock. *Proc. Natl. Acad. Sci. USA* **106**, 11107-11112.

Mochizuki, A. (2005) An Analytical Study of The Number of Steady States in Gene Regulatory Networks. *J. theor. Biol.* **236**, 291-310.

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

Müller, F. J. and Schuppert, A. (2011) Few inputs can reprogram biological networks. *Nature* **478**, E4

Shen-Orr, S. S., Milo, R., Mangan, S. & Alon. U. (2002) Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genet.* **31**, 64–68.

Takens, F. (2010) Reconstruction theory and nonlinear time series analysis. In: H. Broer *et al*. (eds.), *Handbook of Dynamical Systems* **3**. Elsevier Amsterdam, 345-377.

Tamura, T., Takemoto, K., and Akutsu, T. (2010) Finding minimum reaction cuts of metabolic networks under a Boolean model using integer programming and feedback vertex sets. *International Journal of Knowledge Discovery in Bioinformatics*, **1**, 14-31.

Table and Table Caption

Table 1

	FoxD-a/b	NoTrlc	Otx	Twist-like-1	ZicL
palp	0	0	1	0	0
a8.21-a8.22-derived-epidermis	0	0	1	0	0
a-line-lateral-epidermis	0	0	1	0	0
a-line-ventral-epidermis	0	0	0	0	0
b8.18-b8.20-derived-epidermis	0	0	0	0	0
b-line-lateral-epidermis	0	0	0	0	0
b7.14-b7.16-derived-epidermis	0	0	0	0	0
CNS	1	0	0	0	0
muscle	0	0	0	0	0
B8.5-B7.7-mesenchyme	0	0	0	1	0
TLC	0	0	0	1	0
TVC	0	1	0	0	0
endoderm	0	0	0	0	0

Summary of expression of 5 genes in 13 differentiated cells at tailbud stage of Ascidiacea. The expressions of 5 genes among 7 shown in Figure 4b are obtained from Imai *et al*. (2006). Top row indicates names of genes, and leftmost column indicates identified differentiated cells or organs. The Boolean entries 0 and 1 indicates inactive and active state of genes, respectively.

Table 2

①	②	③	④	⑤	# combinations
ErbB11	SOS ERK1/2	HB-EGF c-Src ADAMS	cyt Ca^{2+} CaM CaMKII	PI4,5-P2	2*3*3=18

ErbB11	SOS	HB-EGF c-Src ADAMS	cyt Ca^{2+}	PI4-P DAG PKC PLD phosphatidyl acid PI5K	3*6=18

List of minimal feedback vertex sets of molecules in the signal transduction network. There are 36 possible choices of minimal feedback vertex set.

Table 3

①	②	③	④	⑤	ID reduced network
ErbB11	SOS ERK1/2	c-Src	cyt Ca2+	PI4,5-P2	1
ErbB11	SOS ERK1/2	HB-EGF ADAMS	cyt Ca2+	PI4,5-P2	2
ErbB11	SOS ERK1/2	c-Src	CaM CaMKII	PI4,5-P2	3
ErbB11	SOS ERK1/2	HB-EGF ADAMS	CaM CaMKII	PI4,5-P2	4
ErbB11	SOS	c-Src	cyt Ca2+	DAG PKC	5
ErbB11	SOS	HB-EGF ADAMS	cyt Ca2+	PI4-P DAG PKC PLD phosphatidyl acid PI5K	6
ErbB11	SOS	c-Src	cyt Ca2+	PI4-P PLD phosphatidyl acid PI5K	6

Choices of minimal feedback vertex sets and corresponding structure of steady state networks. See Figure 7 for the ID and structure of steady state networks.

Figure Legends

Figure 1 An example of a two-vertex regulatory network and possible dynamical behaviors depending on regulatory functions. (a) Schematic representation of a regulatory network with two vertices. The directed edges show inhibitory regulatory interactions between nodes. (b, d) Examples of regulatory functions in which the genes are controlled by two different transcription factors. (b): $f_A = \left\{1+\exp\left[20\left(x_B-0.5\right)\right]\right\}^{-1}$, $f_B = \left\{1+\exp\left[20\left(x_A-0.5\right)\right]\right\}^{-1}$, $d_k\left(x_k\right) = x_k$, (d): $f_A = \left\{1+\exp\left[-20\left(x_A-0.5\right)\right]\right\}^{-1}\left\{1+\exp\left[20\left(x_B-0.5\right)\right]\right\}^{-1}$, $f_B = \left\{1+\exp\left[-20\left(x_B-0.5\right)\right]\right\}^{-1}\left\{1+\exp\left[20\left(x_A-0.5\right)\right]\right\}^{-1}$, $d_k\left(x_k\right) = x_k$. (c) and (e) shows dynamical trajectories and null-clines on two-dimensional state space using regulatory functions (b) and (d), respectively. Red and blue curves are null-clines of dynamics of x_A and x_B, respectively. Open and solid circle is unstable and stable stationary point, respectively. Broken curves are trajectories from different initial state.

Figure 2 Examples of regulatory networks with few vertices. The directed edges show regulatory interactions between nodes. The gray vertices are one choice of a minimal feedback vertex set, in each of the cases (a) – (e).

Figure 3 Intuitive explanation of the theory. If the dynamics of the gray vertices are given, the dynamics of the remaining vertices are determined uniquely. The set of vertices on which the dynamics are given can be further reduced to the

minimal feedback vertex set which consist of only the single vertex marked by a red circle. See text for detailed explanation.

Figure 4 Gene regulatory network of ascidiacea development. (a) We draw the network based on Imai *et al*. (2006). The original network includes 16 genes with self-repression. We removed these repressive self-loops because self-repression can be subsumed under degradation in our formulation (2) of the network. (b) Reduced network obtained by successive removal of nodes without input or without output. The reduced network possesses a minimal feedback vertex set with a single vertex, FoxD-a/b.

Figure 5 Diversity of steady states as captured by observation of single genes, for the ascidiacea network and randomly chosen hypothetical regulatory functions f_k. The results of nodal, NoTrlc, FoxD-a/b, FGF9/15/20, Otx, Twist-like-1 and ZicL are shown in separate graphics. For each graphics the vertical axis indicates the percentage of actual multiple steady states correctly identified by observation of the respective gene, only. The horizontal axis enumerates the random choices of regulatory functions, ordered from lower to higher scores for each gene.

Figure 6 Signal transduction network downstream of the EGF receptor, based on Oda *et al*. (2005). We successively removed nodes without input or without output from the original. The network is still complex and intermingled, including many cycles. We marked a choice of a 5-element minimal FVS.

Figure 7 Simplified steady state networks of signal transduction based on various choices of a 5-element FVS. There are 6 possible topologies of simplified networks depending on the choice of the FVS. The correspondence of ID numbers between the topology and the choice of a FVS is listed in Table 3. The difference of edges of each network of ID 2 – 5 from that of ID 1 is shown in red bold arrows.

Figure 8 Dynamical system of mammalian circadian rhythms. (a) A regulatory network with 21 variables, redrawn after Mirsky *et al*. (2009). Our choice of a minimal feedback vertex set I with seven elements is marked by circles. (b) Trajectories of two stable periodic orbits, period1 (P1) and period2 (P2), one unstable periodic orbit, (UP), and one unstable stationary state (USS), represented by time tracks of the variable *Per2*. Vertical axis: *Per2*, horizontal axis: time *t*. Dotted and broken curve: P1, dotted curve: P2, broken curve: UP, solid line: USS. (c) Trajectories of the same solutions in the phase plane of the two variables *Per1* and *Per2*. These two variables are not in the feedback vertex set.

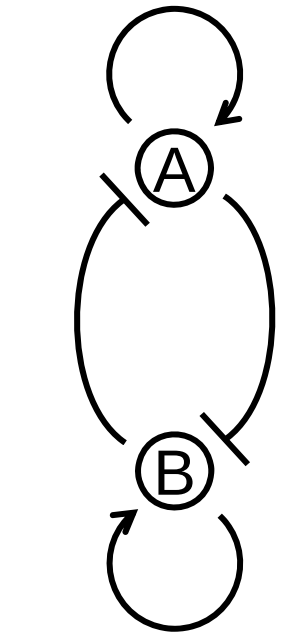
Figure 9 Numerical trajectories of successful open loop controls of circadian rhythms via the full feedback vertex set I. The horizontal and vertical axes are *Per1* and *Per2*, respectively, which are not in the chosen feedback vertex set I. Zooms into P2, UP, and USS are shown as top-right inserts. The resulting trajectory of the control experiment is always the solid red curve. (a) "From P1 to P2". The stable cycles P1 and P2 are shown by gray solid curves. (b) "From P2 to P1". Gray solid: P1 and P2. (c) "From P1 to UP". Gray solid: P1 and UP. (d) "From P1 to USS". Gray solid: P1, open dot: USS.

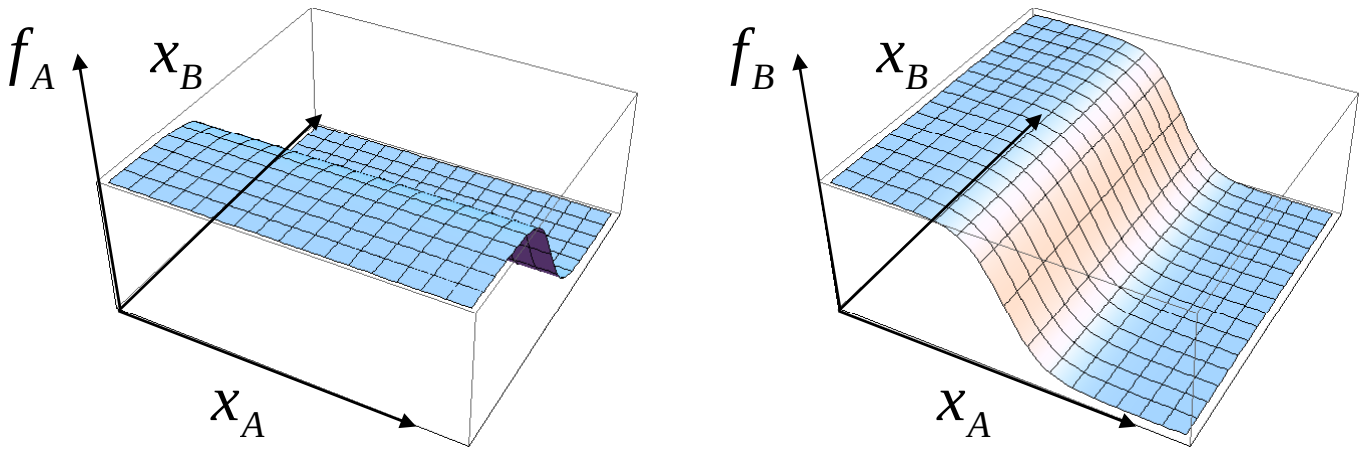
Figure 10 Numerical trajectories of failed open loop controls of circadian rhythms by the reduced vertex set $I' = \{\text{PER1, PER2, CRY1, CRY2, RORc, BMAL1}\} = I \setminus \{\text{CLK}\}$. Trajectories of the failed control experiment are always solid red. (a) "From P1 to P2". Gray solid: stable cycles P1 and P2. Bottom-center: zoom into P2. Top-right insert: trajectory for the same range of *Per1* and *Per2* as in Figures 8c and 9. Top-left insert: Poincaré section of (*Per1*, *Per2*) at CLK=0.65. The insert indicates the presence of an invariant 2-torus with quasi periodic dynamics. (b) "From P1 to USS". Gray solid: P1, open dot: USS. Top-right insert: zoom into UP.

4. Figure

Click here to download 4. Figure: Figures.pdf

Fig. 1a-c





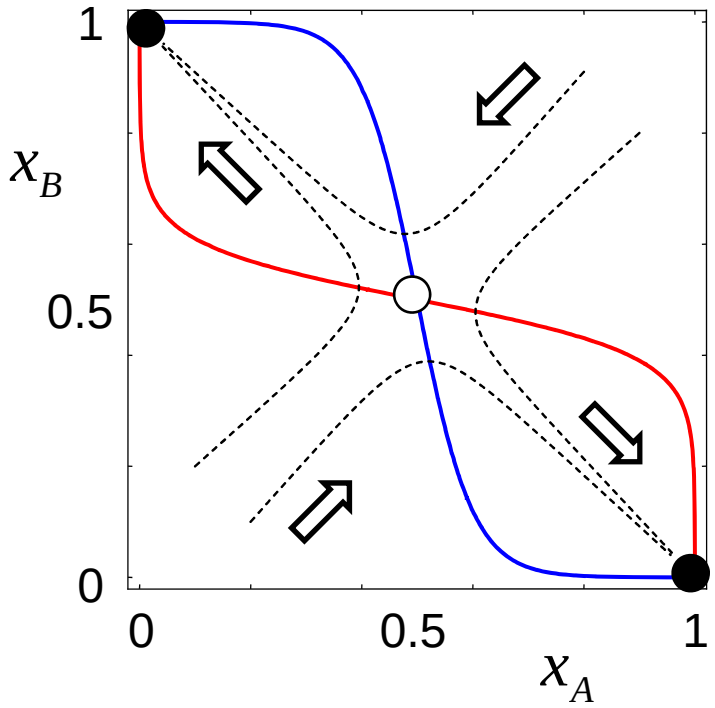
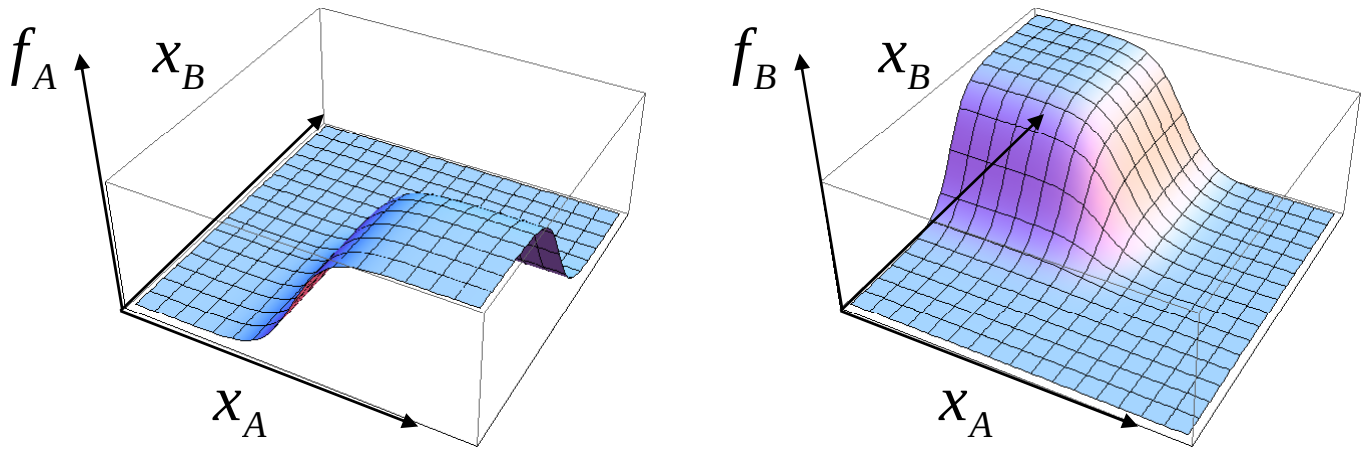


Fig. 1de

d



e

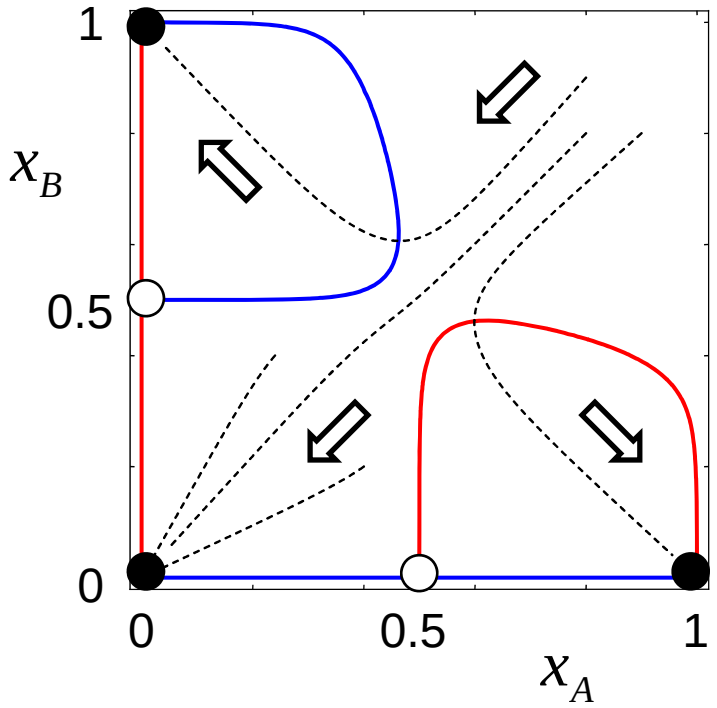


Fig. 2

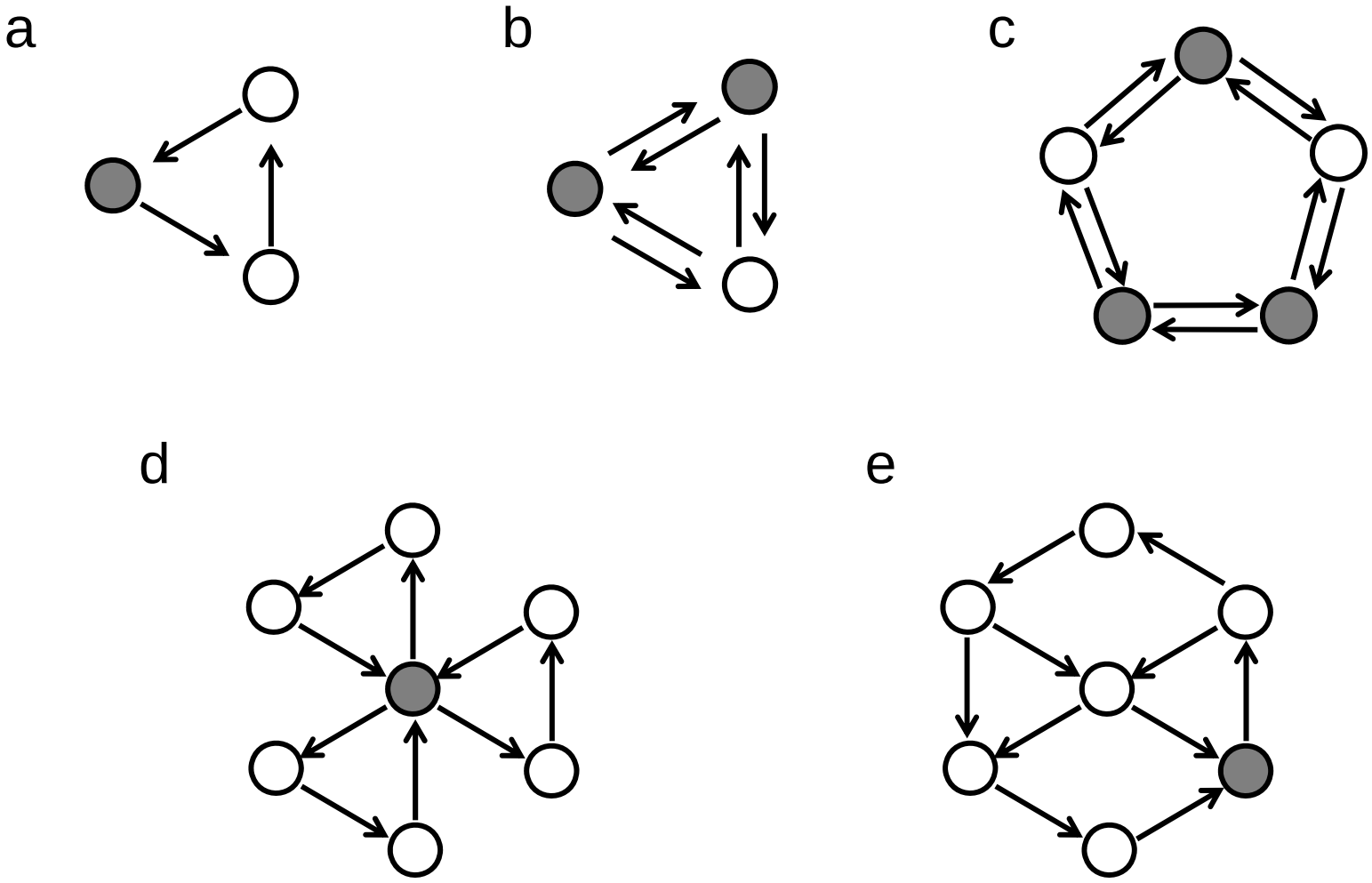


Fig. 3

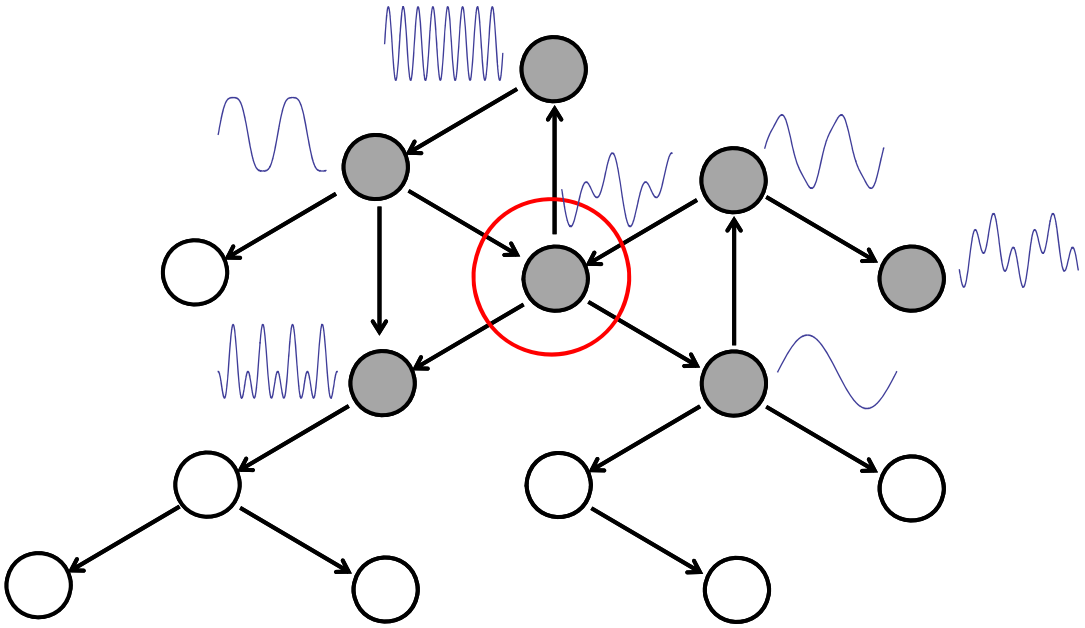


Fig. 4a

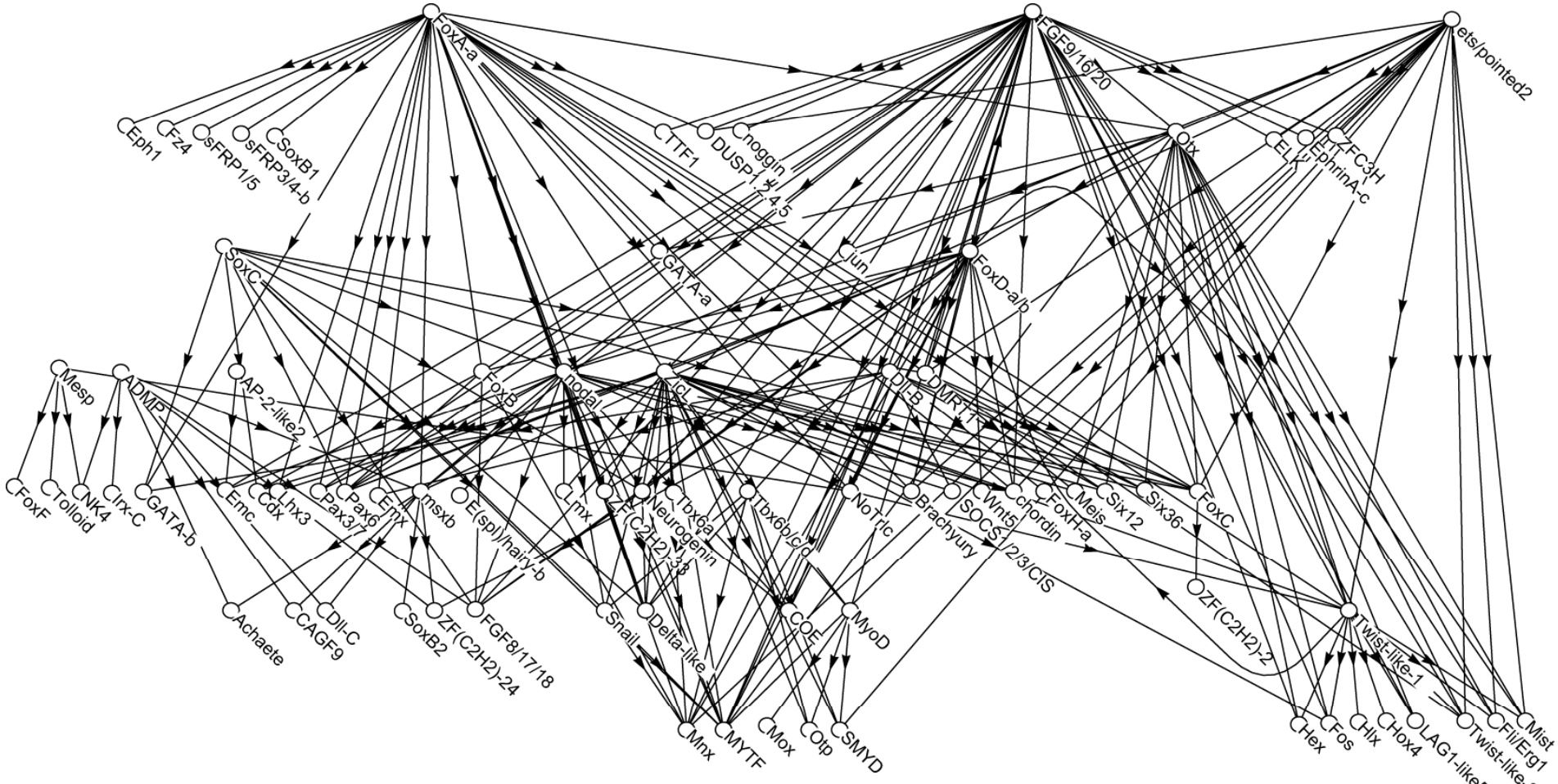

FoxA-a
FGF9/16/20
ets/pointed2
Eph1
Fz4
sFRP1/5
sFRP3/4-b
SoxB1
TTF1
noggin
DUSP1/2/4/5
Itx
EL1
EphrinA-c
FC3H
SoxC
GATA-a
Jun
FoxD-a/b
Mesp
DMP
AP-2-like2
FoxB
nodal1
cL
LB
MRF1
FoxF
Tolloid
NK4
Irx-C
GATA-b
Emc
Cdx
Lhx3
Pax3/7
Pax6
Emx
msxb
E(spl)/hairy-b
Lmx
ZF(C2H2)-38
Neurogenin
Tbx6a
Tbx6b/c/d
NoTrlc
Brachyury
SOCS1/2/3/CIS
Wnt5
chordin
FoxH-a
Meis
Six12
Six36
FoxC
ZF(C2H2)-2
Twist-like-1
Achaete
CAGF9
Dll-C
SoxB2
ZF(C2H2)-24
FGF8/17/18
Snail
Delta-like
COE
MyoD
Mnx
MYTF
Mox
Otp
SMYD
Hex
Fos
Hlx
Hox4
LAG1-like5
Twist-like-2
Fli/Erg1
Mist

Fig. 4b

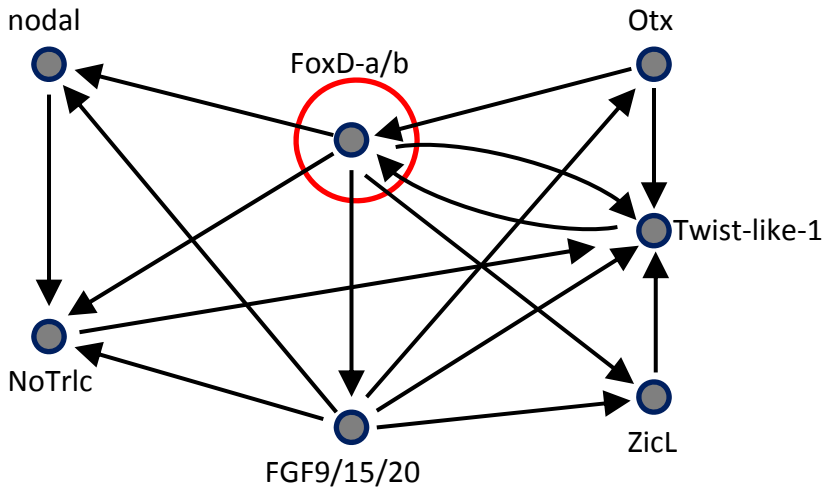


Fig. 5

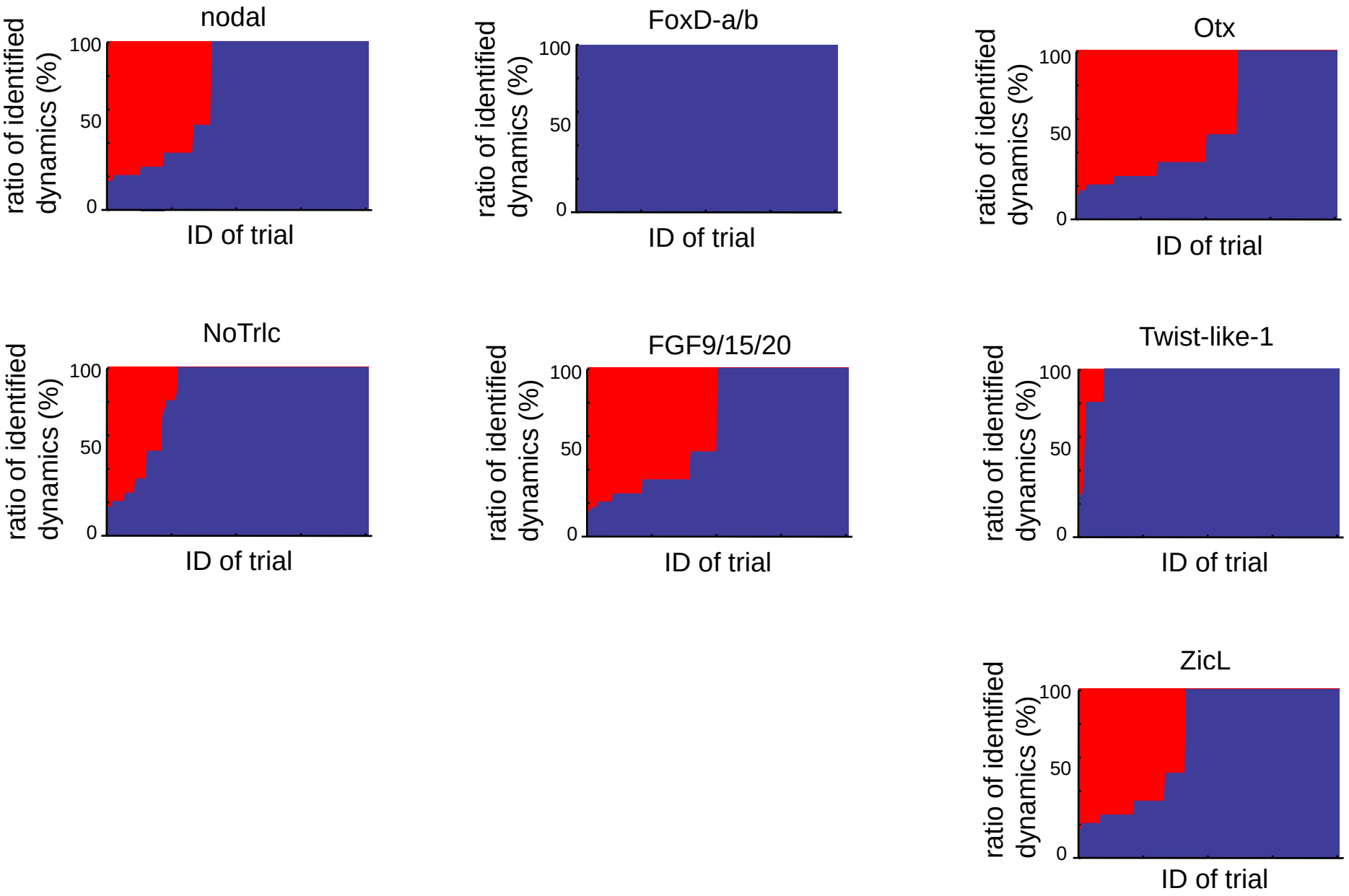

nodal
FoxD-a/b
Otx
NoTrlc
FGF9/15/20
Twist-like-1
ZicL
ratio of identified dynamics (%)
ID of trial
0
50
100

Fig. 6

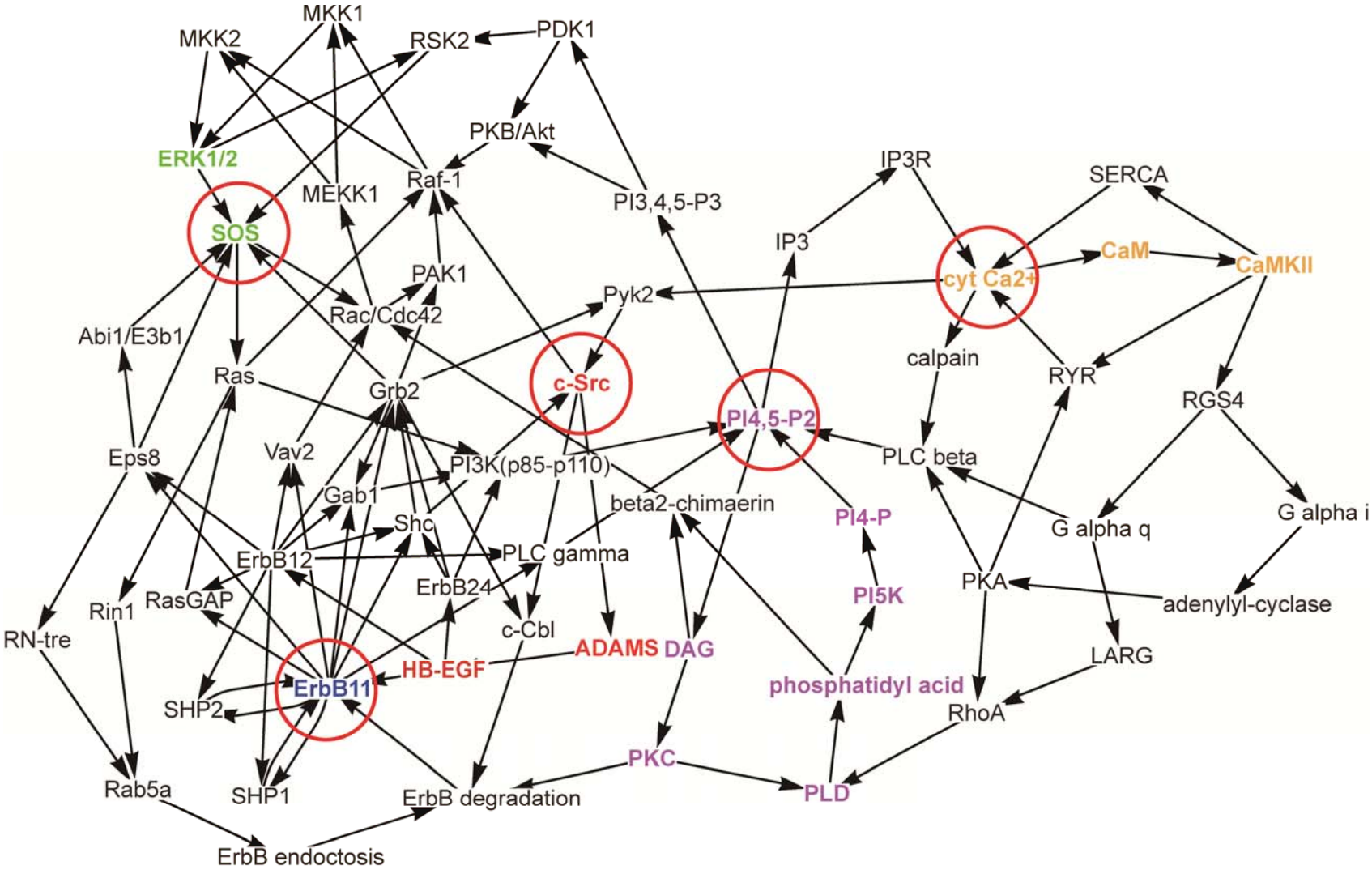
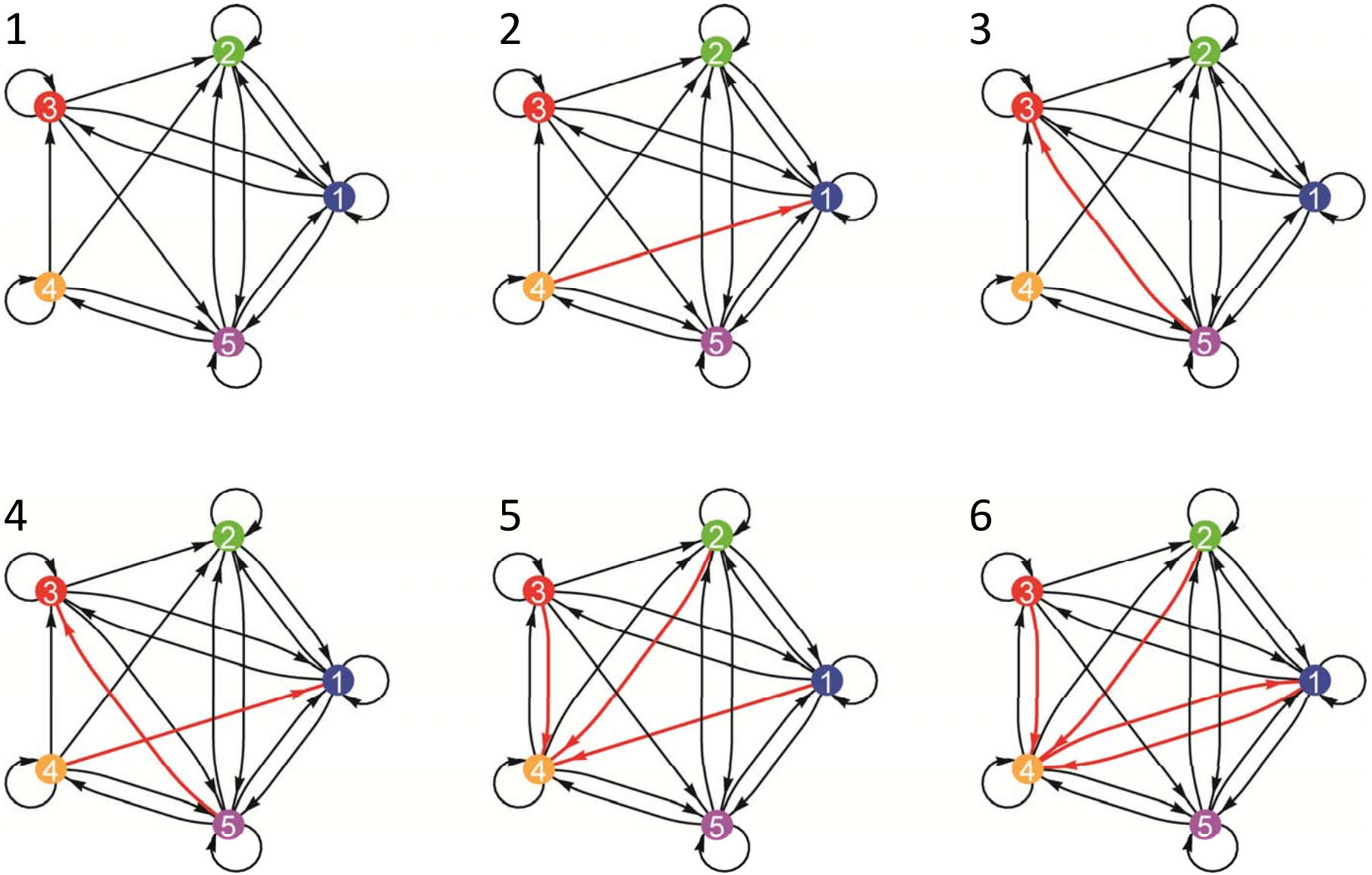
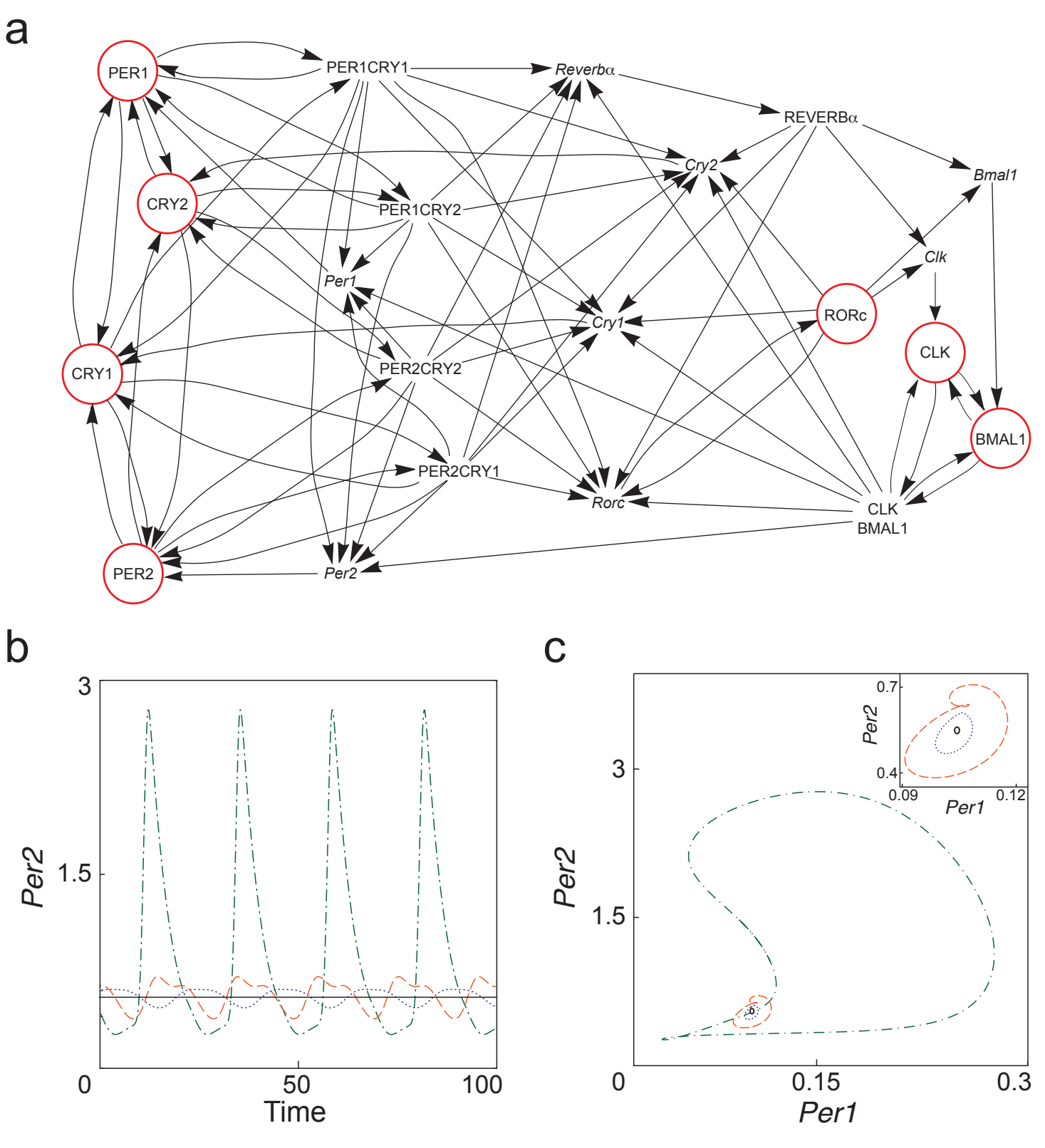
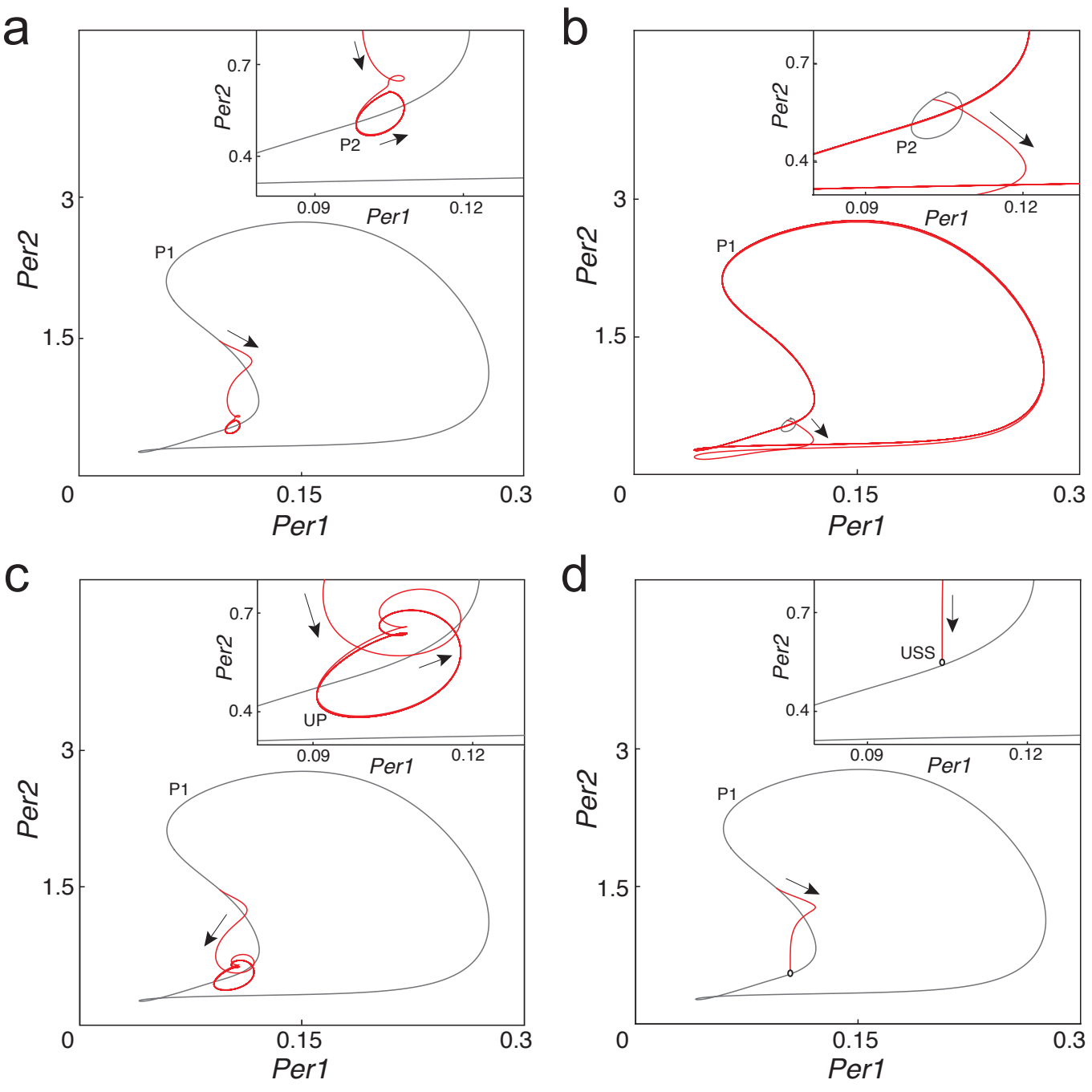
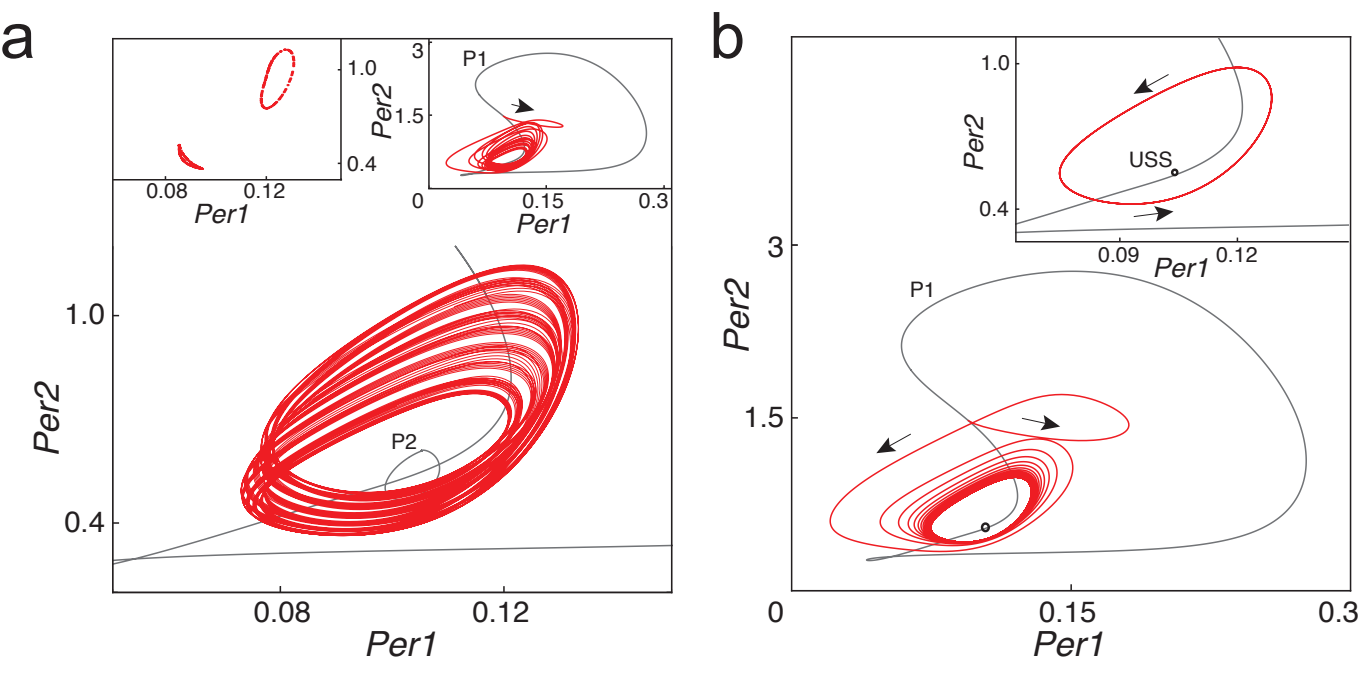

MKK1
MKK2
RSK2
PDK1
PKB/Akt
ERK1/2
MEKK1
Raf-1
PI3,4,5-P3
IP3R
SERCA
IP3
SOS
CaM
CaMKII
cyt Ca2+
PAK1
Pyk2
Rac/Cdc42
Abi1/E3b1
calpain
RYR
Ras
Grb2
c-Src
RGS4
PI4,5-P2
Eps8
Vav2
PI3K(p85-p110)
PLC beta
Gab1
beta2-chimaerin
PI4-P
G alpha i
Shc
G alpha q
ErbB12
PLC gamma
PKA
ErbB24
PI5K
Rin1
RasGAP
adenylyl-cyclase
RN-tre
c-Cbl
ADAMS
DAG
LARG
HB-EGF
ErbB11
phosphatidyl acid
SHP2
RhoA
PKC
Rab5a
SHP1
ErbB degradation
PLD
ErbB endoctosis

Fig. 7




a
PER1
PER1CRY1
Reverbα
REVERBα
Cry2
Bmal1
CRY2
PER1CRY2
Clk
Per1
Cry1
RORc
CRY1
PER2CRY2
CLK
BMAL1
PER2CRY1
Rorc
CLK
BMAL1
PER2
Per2
b
3
1.5
0
50
100
Time
Per2
c
0.7
0.4
0.09
0.12
Per1
Per2
3
1.5
0
0.15
0.3
Per1


a
b
c
d
Per2
Per1
P1
P2
UP
USS
0
0.15
0.3
1.5
3
0.4
0.7
0.09
0.12


a
b
Per1
Per2
P1
P2
USS
0.08
0.12
0.4
1.0
1.5
3
0
0.15
0.3
0.09