



OPEN

SUBJECT AREAS:

REGULATORY
NETWORKS

SYSTEMS ANALYSIS

DYNAMIC NETWORKS

NUMERICAL SIMULATIONS

Discovery of a kernel for controlling biomolecular regulatory networks

Junil Kim, Sang-Min Park & Kwang-Hyun Cho

Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea.

Received
3 April 2013Accepted
3 July 2013Published
17 July 2013Correspondence and
requests for materials
should be addressed to
K.-H.C. (ckh@kaist.ac.
kr)

Cellular behavior is determined not by a single molecule but by many molecules that interact strongly with one another and form a complex network. It is unclear whether cellular behavior can be controlled by regulating certain molecular components in the network. By analyzing a variety of biomolecular regulatory networks, we discovered that only a small fraction of the network components need to be regulated to govern the network dynamics and control cellular behavior. We defined a minimal set of network components that must be regulated to make the cell reach a desired stable state as the control kernel and developed a general algorithm for identifying it. We found that the size of the control kernel was related to both the topological and logical characteristics of a network. Intriguingly, the control kernel of the human signaling network included many drug targets and chemical-binding interactions, suggesting therapeutic application of the control kernel.

The behavior of a cellular system is an emergent outcome of complicated interactions of molecules that work together in a highly organized manner, forming a biomolecular regulatory network in which molecules are represented by nodes and the interactions between molecules are denoted by links (an inhibiting link or an activating link, depending on the regulation type)^{1–3}. As a result, cellular behavior can be defined dynamically only by the network state that represents all the molecular states in the network collectively. The cellular behavior or phenotype can be considered a high-dimensional attractor of the biomolecular regulatory network, in which an attractor is a mathematical concept that represents a stable steady state adopted by a dynamic system, which in this case is a biomolecular regulatory network^{4–6}. Based on the concept of an attractor, a biomolecular regulatory network is mapped onto an attractor landscape, where each point in the landscape represents one state of the network, defined by a set of state values that contain the activity states of all the molecules in the network⁴. Although an attractor landscape of a biomolecular regulatory network is composed of various attractors, cellular behavior mainly reaches a dominant stable state known as the *primary attractor*^{7–8}. The area around an attractor is called the *basin of attraction* and is the region of states that lead to trajectories that converge to the attractor^{4,6}. The question then arises as to whether the cellular behavior can be controlled with minimal effort such that a desired cellular state is reached regardless of the initial state. In terms of the attractor landscape, this control corresponds to identification of a minimal set of nodes such that the regulation (or pinning) of these nodes guarantees the convergence of the network state to the desired attractor, with the basin of attraction equivalent to the whole state space.

Several studies have investigated control of complex networks, with the focus on the structural controllability needed to drive a network from any initial state to any desired final state on the basis of the network structure without considering the inherent nonlinear dynamics and detailed interaction types, such as activatory or inhibitory regulation^{9–11}. Liu *et al.*⁹ reported that there is a set of driver nodes that can offer full control over the network based on nodal dynamics (i.e., perturbing the nodes). They also found that the driver nodes tend to avoid hubs and that the total number of driver nodes depends on the degree (the number of links connected to the node) distribution. However, Nepusz *et al.*¹⁰ reported contrasting results with respect to both the structural characteristics and the number of driver nodes based on edge dynamics (i.e., perturbing the interactions between the nodes, called the switchboard dynamics (SBD)). These structure-based controllability studies have limitations when applied to biomolecular regulatory networks because they assumed linear dynamics and time-varying control of nodes, which are unrealistic and unfeasible, if not impossible, for biomolecular regulatory networks. Moreover, the different regulation types between molecules, such as activation or inhibition, have not been addressed in the framework of these previous studies.

In the present study, we investigated whether cellular behavior could be controlled by regulating a minimal set of nodes of the underlying biomolecular regulatory network, based on the concept of the attractor landscape. We



defined this minimal set of nodes as the *control kernel*, the regulation of which ensures that the network state converges to a desired attractor from any possible initial network state. For this purpose, we employed a discrete logic-based Boolean network model that can avoid the problem of parameterization, which is often a critical limitation of continuous dynamic models such as a differential equation model, especially for large complex biomolecular regulatory networks^{2,12,13}. Using the Boolean network model, we developed a general algorithm for identifying the control kernel, and we applied it to various biomolecular regulatory networks. Surprisingly, we found that the control kernel includes only a small fraction of the total number of network nodes and that the size of the control kernel depends on both topological and logical characteristics of the network. We also found that the control kernel of the human fibroblast signaling network was enriched with drug targets and chemical-binding interactions, which suggests that targeting the control kernel of a biomolecular regulatory network could be an effective therapeutic strategy. Identifying and regulating the control kernel can also be useful for controlling the differentiation of stem cells to a specific cell type, controlling the reprogramming or transdifferentiating of the differentiated cells in a systematic manner¹⁴ or converting a disease cellular state to a normal state^{15,16}.

Results

Identification of the control kernel. We defined the control kernel for any desired attractor of a given biomolecular regulatory network as the minimal set of network nodes needing to be regulated to drive

the network state to converge to any desired attractor regardless of the system's initial state. Fig. 1 illustrates the concept of the control kernel. The toy example network in Fig. 1a shows that the network state can converge to many different attractors, depending on the initial network state. However, Fig. 1b shows that the network state converges only to the desired primary attractor from any initial state if the three nodes {1,2,7} are regulated by pinning their states to those of the primary attractor. Thus, in this example network, the set of three nodes {1,2,7} is the control kernel for the primary attractor (the red nodes in Fig. 1b). To find the control kernel for any biomolecular regulatory network, we developed a general identification algorithm based on attractor landscape analysis (see METHODS).

Control kernel of various biomolecular regulatory networks. We have investigated the control kernel of various biomolecular regulatory networks by applying the kernel identification algorithm to Boolean network models constructed on the basis of experimental evidence. These networks include the *S. cerevisiae* cell cycle network⁷, the *S. pombe* cell cycle network⁸, the gene regulatory network (GRN) underlying mammalian cortical area development¹⁷, the GRN underlying *A. thaliana* development¹⁸, the GRN underlying mouse myeloid development¹⁹, the mammalian cell cycle network²⁰, the cAMP response element-binding protein (CREB) signaling network²¹, and the human fibroblast signaling network¹². We aimed to identify the control kernel for the primary attractor of each of these networks (Figs. 2 and 3). We expected that we would have to regulate many of the network nodes to make all the network states converge to

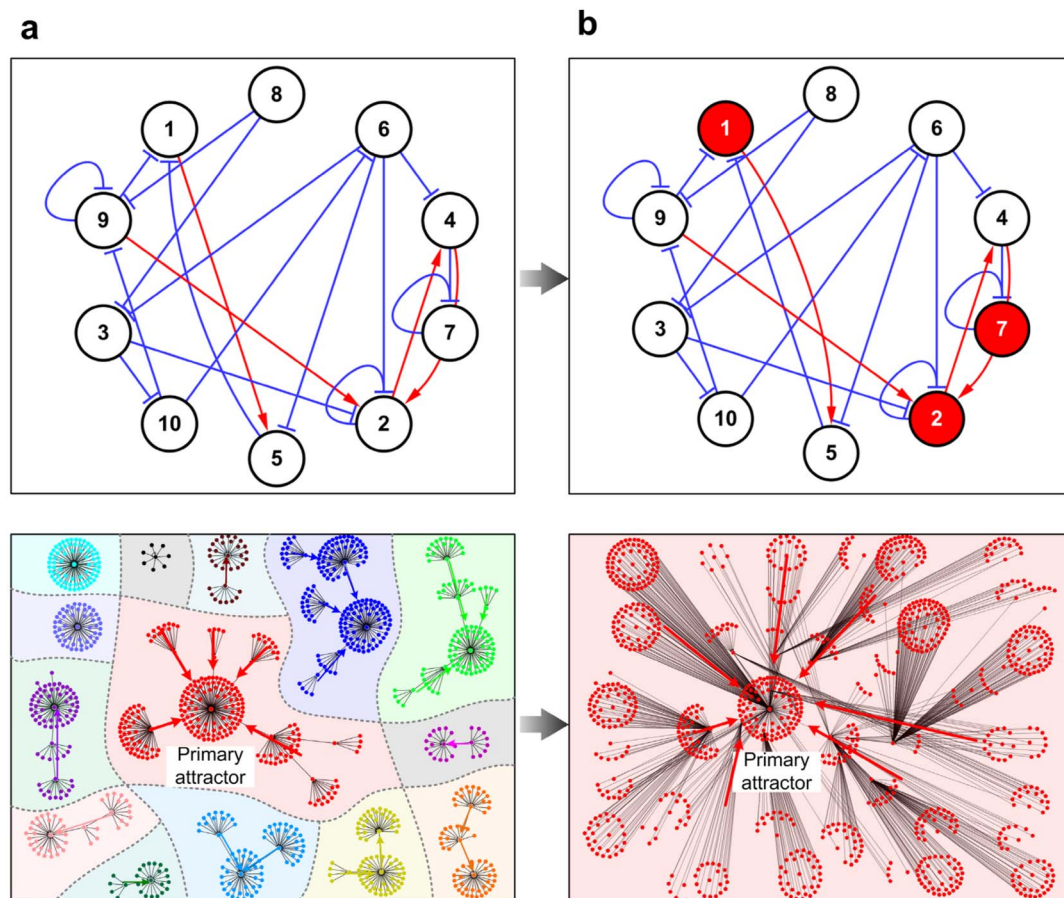


Figure 1 | An example network and its state transition diagram illustrating the concept of the control kernel. The original network is shown in (a), and the controlled network, in which the nodes of the control kernel were regulated, is shown in (b). In the upper panel, a red node, red arrow, and blue blunt line denote a control kernel node, activation link, and inhibition link, respectively. Each dot in the state transition diagram (lower panel) represents a network state that is composed of all the node values at a certain instant in time, and the same colored dots denote inclusion in the same basin of attraction.

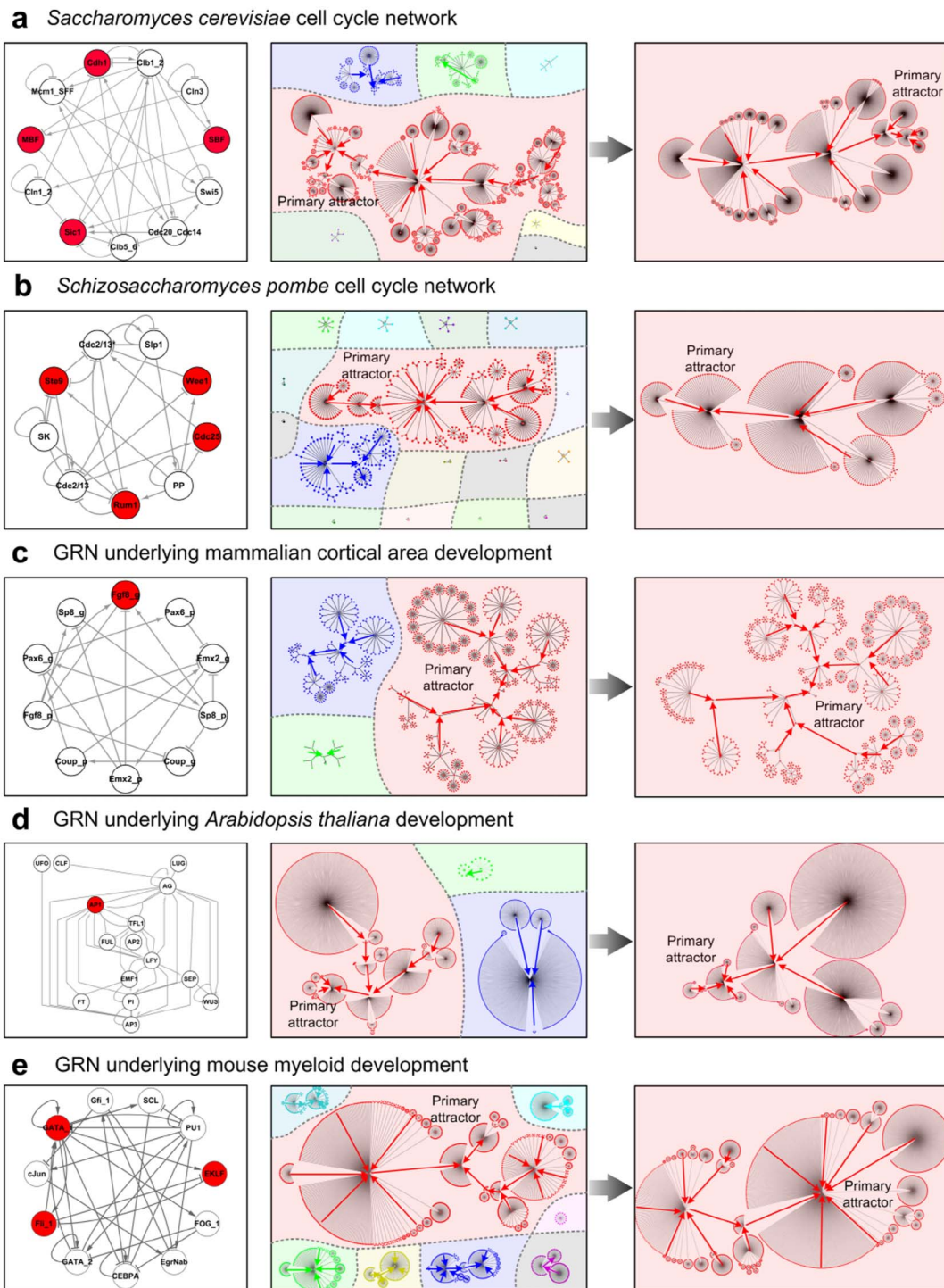


Figure 2 | Five biomolecular regulatory networks and their state transition diagrams before and after regulating the nodes of the control kernel. (a) *Saccharomyces cerevisiae* cell cycle network. (b) *Schizosaccharomyces pombe* cell cycle network. (c) Gene regulatory network (GRN) underlying mammalian cortical area development. (d) GRN underlying *Arabidopsis thaliana* development. (e) GRN underlying mouse myeloid development. In (a)–(e), the left, center, and right panels show the original network with the control kernel denoted by red nodes, the state transition diagram of the original network, and the state transition diagram of the controlled network, respectively. In each state transition diagram, the same colored dots represent inclusion in the same basin of attraction.

a desired attractor because cellular phenotypes are known to be robust to external perturbations²². Surprisingly, however, we found that we needed to regulate only a small fraction of network nodes to drive the network state to the primary attractor (36% for the *S. cerevisiae* cell cycle network, 44% for the *S. pombe* cell cycle network, 10% for the GRN underlying mammalian cortical area development, 6.7% for the GRN underlying *A. thaliana* development, 30% for the GRN

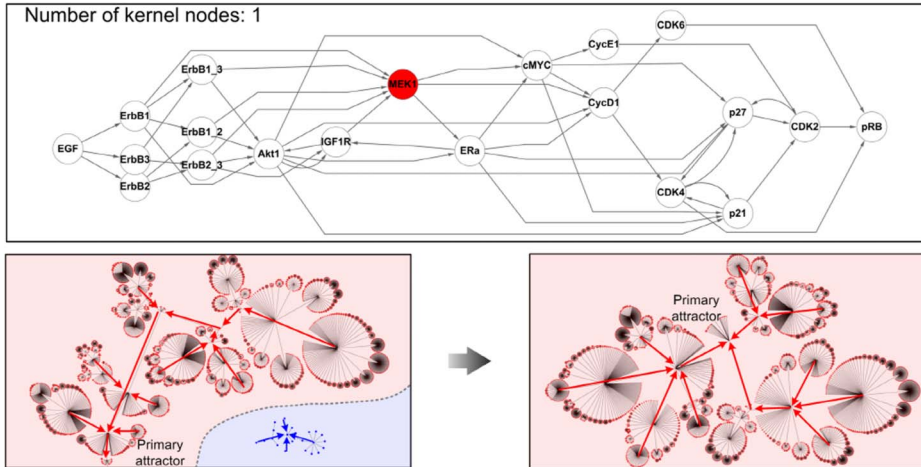
underlying mouse myeloid development, 5% for the mammalian cell cycle network, 7.8% for the CREB signaling network, and 8.6% for the human fibroblast signaling network (Table 1)).

The size of the control kernel. We found that the control kernel is, in general, relatively small in view of the total number of network nodes, but we also found that its size varies depending on the network. This



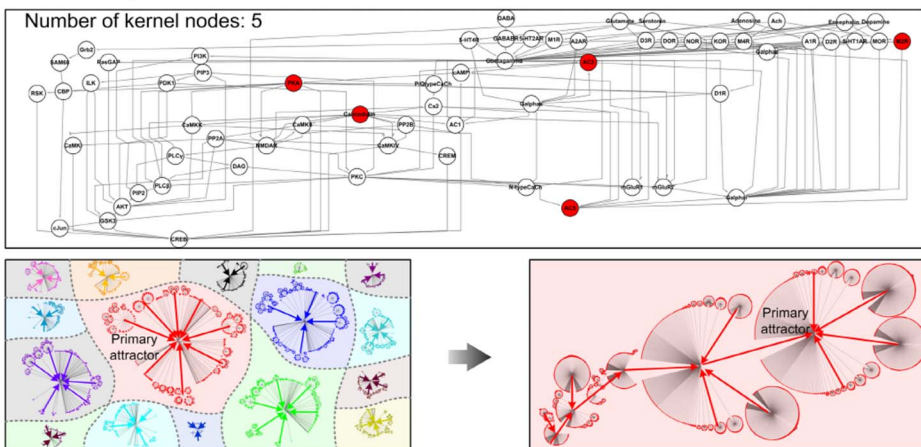
a Mammalian cell cycle network

Number of kernel nodes: 1



b CREB signaling network

Number of kernel nodes: 5



c Human fibroblast signaling network

Number of kernel nodes: 12

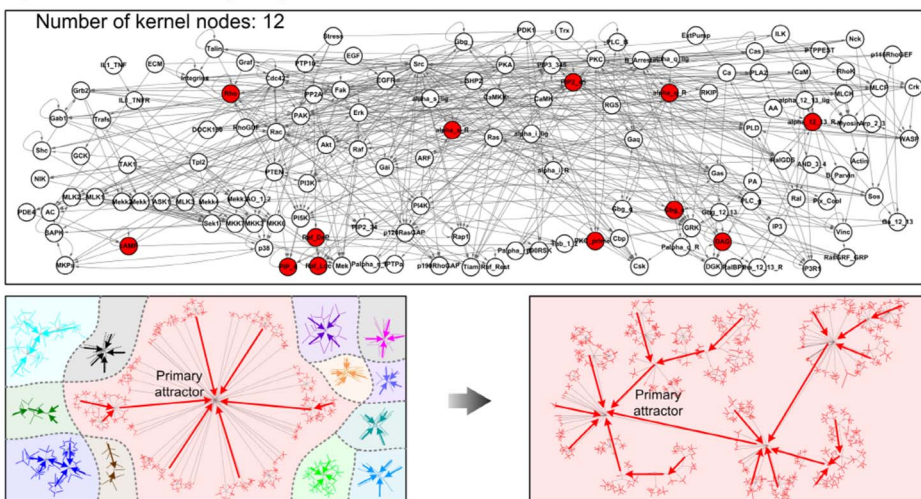


Figure 3 | Examples of larger biomolecular regulatory networks and their state transition diagrams before and after the regulation of control kernel nodes. (a) Mammalian cell cycle network. (b) CREB signaling network. (c) Human fibroblast signaling network. In (a)–(c), the upper, lower left, and lower right panels show the original network with the control kernel denoted by red nodes, the state transition diagram of the original network, and the state transition diagram of the controlled network, respectively. In each state transition diagram, the same colored dots represent inclusion in the same basin of attraction.

promoted the question of what determines the size of the control kernel. To investigate the relationship between the size of the control kernel and the topological characteristics of a network, we considered the most representative five topological characteristics:

the average degree, the proportion of inhibitory links, the network heterogeneity^{23,24}, the clustering coefficient²³, and the characteristic path length²³. We conducted a linear correlation test and found that only the proportion of inhibitory links is significantly positively



correlated with the size of the control kernel (Table 1 and Supplementary Fig. S1). To investigate the general relationship between the topological characteristics of a network and the size of the control kernel, we examined their correlations in random networks of different sizes and with different types of logical rules for state transitions (see Supplementary Methods). Only the proportion of inhibitory links was found to be positively correlated with the size of the control kernel in these random networks, for all sizes (Supplementary Fig. S2); however, this correlation was dependent on the logical rules.

To further investigate why the size of the control kernel has such a correlation with the proportion of inhibitory links, we explored the relationship between the size of the control kernel and the dynamical characteristics represented by the number of attractors and the basin size of the desired attractor. We can expect that, if a network has many attractors and the basin size of the desired attractor is small, then many nodes should be regulated, because most of the network states would not converge to the desired attractor without external intervention. To examine this expectation, the average size of the control kernel was calculated over the number of attractors and the basin size of the desired primary attractor. We found that the size of the control kernel was smaller when the network had fewer attractors and the primary attractor had a larger basin (Supplementary Fig. S3). We also found that the number of attractors and the basin size of the primary attractor of a network depended on the proportion of inhibitory links of the network for most of the logical rules (Supplementary Fig. S4). Thus, the correlation between the size of the control kernel and the proportion of inhibitory links (Supplementary Fig. S2) can be explained by the relationship between the proportion of inhibitory links and number of attractors, as well as the basin size of the primary attractor (Supplementary Fig. S4).

Topological and dynamical characteristics of the control kernel.

Previous studies have indicated that the degree of a node is an important topological parameter that determines whether the node is a “driver node”^{9,10}. Thus, we compared the degree characteristics of the nodes included in the control kernel with those of the remaining non-kernel nodes. In contrast to the earlier studies, we found that the nodes of the control kernel do not have any specific degree characteristics in either the eight biomolecular regulatory networks (Supplementary Fig. S5) or in the random networks (Supplementary Fig. S6). However, we found an interesting characteristic for the state value distribution of the control kernel nodes. For a quantitative comparison, we defined the “state coherency” of a node to be a measure of the different state value distributions over all the attractors (see METHODS for details). We found that the average state coherency of the control kernel nodes is lower than that of the non-kernel nodes in both biomolecular regulatory networks (Supplementary Fig. S5 and Table S1–S8) and random networks (Supplementary Fig. S6). This result indicates that a node that has different state values depending on the attractors is more likely to be included in the control kernel.

The ratio of drug targets and the number of chemical-binding interactions of the control kernel. Attractors of a biomolecular regulatory network often represent cellular phenotypes^{4–6}. Thus, controlling the network state transition to a desired attractor by regulating the nodes of a control kernel is related to controlling the dynamic behavior of a cellular system, which suggests that the nodes of the control kernel might be related to drug targets. We compared the ratio of drug targets among the nodes of the control kernel with that of non-kernel nodes for the human fibroblast signaling network. Surprisingly, we found that all the control kernel nodes were drug targets, whereas only 29% of the non-kernel nodes were drug targets (Fig. 4a). This difference was statistically very significant (p -value = $1.67\text{E}-4$). We further investigated the number of chemical-binding interactions of the nodes of the control kernel compared with the

number for non-kernel nodes and found that the chemical-binding interactions were much more enriched for the control kernel nodes (p -value = $5.76\text{E}-6$) (Fig. 4b). Figure 4c and d show that the control kernel contains a larger number of drug targets and chemical-binding interactions. Taking these findings together, our study shows that identifying the control kernel might be therapeutically beneficial by unraveling a pool of novel drug targets.

Discussion

In this paper, we introduce the new concept of the control kernel, and examine it for various biomolecular regulatory networks on the basis of attractor landscape analysis. Intriguingly, we found that only a small fraction of the network nodes belong to the control kernel. This finding might seem counter-intuitive because cellular systems must have evolved to be robust against such a small number of molecular perturbations. However, the combination of the small number of molecular perturbations in the control kernel is crucial, and most other combinations of the same number of molecules are not effective. Thus, in this respect, we can still consider that the cellular system is robust to random perturbations of a small number of molecules. Moreover, this result suggests that the cellular system is adaptable because the perturbation of the control kernel allows the cellular system to change its dynamic behavior and acquire a new phenotype.

Previous studies that addressed the controllability problem of complex networks primarily focused on the structural characteristics of driver nodes, which are nodes whose control is sufficient to fully control the system’s dynamics^{9,10}. The main findings were that degree and heterogeneity are the most important determinants of driver nodes. By contrast, our results showed that not only the topological characteristics, such as the proportion of inhibitory interactions, but also the logical rules for state transitions are important in determining the size of the control kernel because these aspects together define the dynamical characteristics of a network with respect to the attractors and basin sizes. We further found that state coherency is a critical parameter that characterizes the control kernel. This finding is noteworthy because state coherency better represents the dynamical characteristics of a node rather than the degree, which represents only the topological characteristics.

It is remarkable that drug targets and chemical-binding interactions were highly enriched in the control kernel of the human fibroblast signaling network. Such enrichments were much less significant for the set of driver nodes^{9,10} (Supplementary Fig. S7 and Table S9). This indicates that the control kernel is a better representation of the biologically meaningful node set than the driver nodes of previous studies, which suggests that the control kernel could be useful for identifying new drug targets.

Recently, Liu *et al.* suggested the concept of an optimal sensor set defined by the minimal set of network nodes allowing us to reconstruct the system’s complete internal state (i.e. observation point of view)²⁵, whereas we proposed the concept of control kernel defined by the minimal set of network nodes needing to be regulated to drive the network state to converge to any desired attractor regardless of the system’s initial state (i.e. control point of view). We compared the control kernel and the optimal sensor set of the human fibroblast signaling network and showed that they are composed of very different nodes (see Supplementary Table S9).

As the number of network nodes increases, the size of whole state space increases exponentially. So, there is a practical limitation in exactly searching for the whole state space due to computational complexity. But, we can still investigate the state space of any large Boolean network model by estimating its attractor landscape on the basis of random sampling of state transition trajectories (as demonstrated by one of our example networks, the human fibroblast signaling network which consists of 139 nodes and 575 links). It has been known that biologically significant attractors have a relatively large basin^{26,27} and that they represent distinct cell phenotypes. So, it is

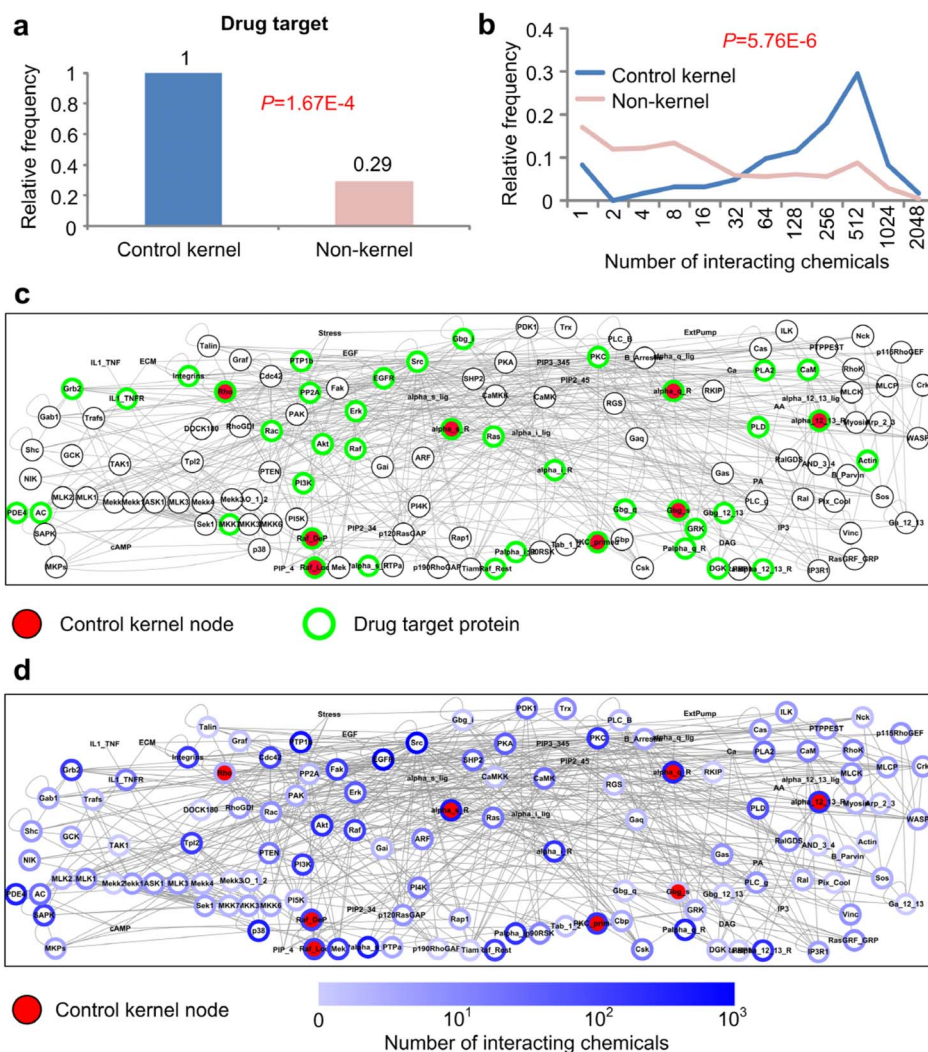


Figure 4 | The enrichment of drug targets and chemical-binding interactions in the control kernel of the human fibroblast signaling network. (a) The enrichment of drug targets in the control kernel and non-kernel. (b) The distribution of the number of chemical-binding interactions in the control kernel and non-kernel. (c) Network representation of the enrichment of drug targets that shows the overlapping of control kernel nodes and drug targets. (d) Network representation of the enrichment of chemical-binding interactions that shows the overlapping of control kernel nodes and the nodes that have a larger number of chemical-binding interactions.

computationally tractable to search for such biologically meaningful attractors and to estimate their basin of attraction. The proposed method is therefore applicable to any large scale biological network.

Biological variables are continuous and, when we represent such continuous variables with discrete-state Boolean variables, we are assuming that discrete-states can still describe the important states of the system, which has been proved true for biological networks in many previous studies^{7,12,26–28}. So, the Boolean network model has been widely used for modeling biological systems and has been proved very useful for analyzing complex biological phenomena. Continuous dynamic modeling such as ordinary differential equation modeling requires parameter estimation and the simulation result will heavily depend on the estimated parameter value. For a large scale biological network, such parameter estimation becomes an ill-posed problem (i.e. we have too many parameters to estimate compared to the available measurements of the data)^{13,29} and it is actually very hard to find optimal parameter estimates unless we have enough time-series measurements of all variables. On the other hand, a discrete logic-based Boolean network model can avoid such a problem of parameterization. In this point of view, a discrete logic-based Boolean network model could be a more appropriate model for the large complex biomolecular

regulatory networks than a continuous dynamic model for practical application.

We envision that our study can be used to understand and control the cell fate determination process. For example, some pioneering studies on cellular reprogramming showed that only a few inputs are sufficient to reprogram complex biological processes^{30–32}. However, the actual event of reprogramming at the level of individual cells still shows a very low efficiency²⁷. One possible explanation for this observation is that the initial states of the cells in these experiments are not the same, which results in convergence to different attractors that represent different cell phenotypes³³. Thus, if we choose the nodes of the control kernel as reprogramming factors, then we might be able to improve the efficiency of reprogramming, regardless of the heterogeneity in the initial cellular states (Fig. 5a). Another example for an application of our study is developing a novel therapeutic strategy by identifying and regulating the control kernel for a normal attractor such that a pathogenic network state is perturbed and converges to the normal attractor (Fig. 5b).

Methods

A general algorithm for the identification of the control kernel. To identify the control kernel of a given biomolecular regulatory network, we must first construct its



state transition diagram, which includes all of the state transition trajectories from any possible initial state to the corresponding final attractor state. On the basis of this state transition diagram, we investigated the resulting attractor landscape that consists of attractors and their basins. For small-scale networks (with the number of nodes ≤ 16), the whole state transition diagram can be constructed, and all the state transition trajectories can be investigated by a full search. However, for large-scale networks (with the number of nodes > 16), it is difficult, if not impossible, to construct the whole state transition diagram because of the resulting computational limitations. Therefore, in this case, we estimated the attractor landscape by investigating a number of sampled state transition trajectories from many different random initial states (10,000 initial states were randomly sampled in this study).

If the state transition diagram contains only a single attractor, then no control action is required to ensure convergence to this attractor because all the states will eventually converge to it without any control. However, if there are multiple attractors, we can identify the control kernel for any desired attractor through the following algorithm. Overall, the proposed algorithm repeats the followings: (i) random selection of a candidate node set, (ii) regulation of the candidate node set by pinning the state value of each node in the set to the corresponding final state value of the node in the desired attractor, and (iii) investigation of the resulting attractor landscape of the controlled network to examine the convergence of all possible initial network states to the desired attractor. We determined the minimal set of nodes that satisfies (iii) as the control kernel, and we employed a genetic algorithm (GA) to find this minimal set. GA is a computational optimization algorithm that encodes a candidate solution with a chromosome and obtains an optimal solution by artificially evolving the chromosomes using a set of operations, such as selection, crossover, and mutation³⁴. In this study, we defined the chromosome as a candidate node set, and for each chromosome, its fitness function was evaluated as the resulting basin size of the desired attractor and number of nodes in the set (see Supplementary Methods, Fig. S8 and S9, for details on the parameter settings, the fitness value trajectories of the example biomolecular regulatory networks, and the flow diagram of the overall algorithm).

Random networks. We generated a number of random networks (24,300 in total) to investigate the general relationship between the control kernel and the network properties by varying the network parameters (three different network sizes, nine different average degrees, nine different proportions of inhibitory links, and 100 randomizations) and the logical rules for state transitions (four different types of rules were considered) (see Supplementary Methods for details).

State coherency. The state coherency of a node is defined as follows:

$$C_i = \left| \sum_{j=1}^m \sum_{k=1}^{l_j} \frac{x_{ijk} B_j}{m l_j} - 0.5 \right| + 0.5$$

where C_i denotes the state coherency of node i , m denotes the number of attractors, l_j denotes the period length of attractor j ($l_j = 1$ for a point attractor), x_{ijk} denotes the k -th state value of node i in attractor j , B_j denotes the relative basin size of attractor j , and $|\cdot|$ denotes the absolute value. The state coherency varies between 0.5 and 1.

Drug targets and the number of chemical-binding interactions. The drug targets, which were retrieved from the DrugBank database³⁵, included 1,330 proteins that are targets of the drugs approved by the US Food and Drug Administration (FDA). We investigated the enrichment of those targets of FDA-approved drugs. The number of chemical-binding interactions for each protein in the network were investigated on the basis of the Stith database³⁶, which provides the experimentally verified chemical-binding interactions of 6,849 proteins.

Statistical analysis. We performed a one-sided two-sample t-test to compare the number of chemical-binding interactions of the control kernel nodes with that of non-kernel nodes, the number of chemical-binding interactions of the driver nodes with that of the complements of driver nodes, and the number of chemical-binding interactions of the driver nodes on SBD with the complements of the driver nodes on SBD in the human fibroblast signaling network. To compare the enrichment of the

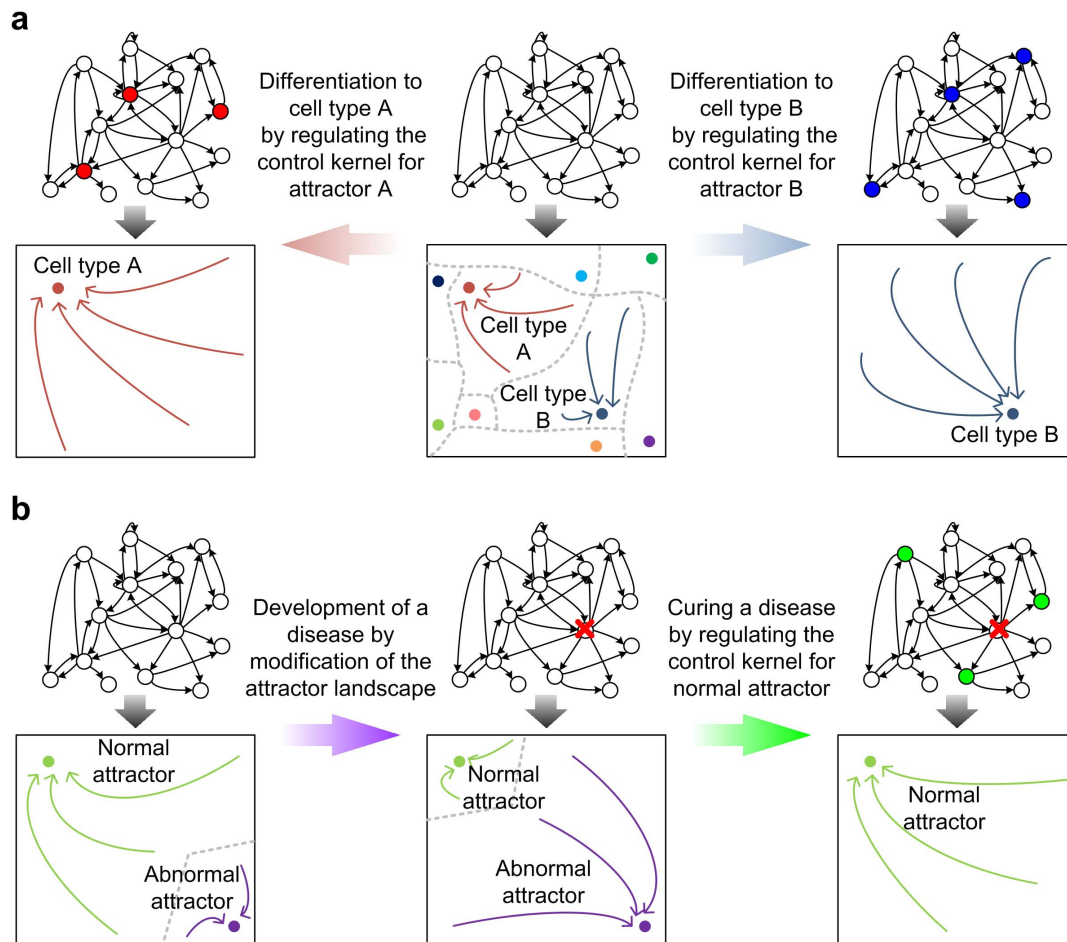


Figure 5 | A conceptual diagram that illustrates the potential applications of the control kernel. The nodes of the control kernel can be used as reprogramming factors that improve the efficiency of the cellular reprogramming, regardless of the heterogeneity of the initial cellular states (a). The control kernel can also be used for developing a novel therapeutic strategy in such a way that the pathogenic network state is perturbed to return to the normal attractor (b).



Table 1 | Network parameters of the biomolecular regulatory networks

| | N | L | k | p_i | p_c | H | C | L | n_a | $p_{b,p}$ | k_c | k_h | $k_{c,i}$ | $k_{h,i}$ | $k_{c,o}$ | $k_{h,o}$ |
|--|-----|-----|------|-------|-------|------|-------|-------|-------|-----------|-------|-------|-----------|-----------|-----------|-----------|
| <i>S. cerevisiae</i> cell cycle network | 11 | 34 | 6.18 | 0.56 | 0.36 | 0.38 | 0.277 | 2.12 | 7 | 0.86 | 4.5 | 7.14 | 3.25 | 2.29 | 1.25 | 3.43 |
| <i>S. pombe</i> cell cycle network | 9 | 26 | 5.78 | 0.69 | 0.44 | 0.27 | 0 | 1.984 | 16 | 0.67 | 4.5 | 6.8 | 3 | 2 | 1.5 | 2.33 |
| GRN underlying mammalian cortical area development | 10 | 19 | 3.8 | 0.47 | 0.1 | 0.26 | 0 | 2.711 | 3 | 0.77 | 4.33 | 3 | 2 | 1.75 | 3.85 | 1.25 |
| GRN underlying <i>A. thaliana</i> development | 15 | 41 | 5.47 | 0.42 | 0.067 | 0.67 | 0.287 | 2.189 | 3 | 0.71 | 8.67 | 5.89 | 3 | 3.11 | 5.67 | 1.89 |
| GRN underlying mouse myeloid development | 11 | 30 | 5.45 | 0.47 | 0.27 | 0.43 | 0.439 | 2.081 | 8 | 0.53 | 6.33 | 5.13 | 2.33 | 2.38 | 3.33 | 2 |
| Mammalian cell cycle network | 20 | 51 | 5.1 | 0.24 | 0.05 | 0.43 | 0.112 | 2.209 | NA | 0.998 | 6.5 | 5.06 | 4 | 2.53 | 2.5 | 2.53 |
| CREB signaling network | 64 | 152 | 4.75 | 0.22 | 0.078 | 0.75 | 0.051 | 3.878 | NA | 0.29 | 7.5 | 4.59 | 4.5 | 2.37 | 3 | 2.22 |
| Human fibroblast signaling network | 139 | 575 | 8.27 | 0.23 | 0.086 | 0.71 | 0.104 | 4.486 | NA | 0.56 | 9.67 | 8.62 | 3.5 | 4.07 | 4.17 | 3.86 |

N , number of nodes; L , number of links; k , average degree; p_i , proportion of inhibitory links; p_c , proportion of control kernel nodes; H , degree heterogeneity; C , clustering coefficient; L , characteristic path length; n_a , number of attractors; $p_{b,p}$, proportion of the basin size of the primary attractor; k_c , average degree of the control kernel nodes; $k_{c,i}$, average in-degree of the control kernel nodes; $k_{c,o}$, average out-degree of the control kernel nodes; k_h , average degree of the non-kernel nodes; $k_{h,i}$, average in-degree of the non-kernel nodes; $k_{h,o}$, average out-degree of the non-kernel nodes; NA, not available. The exact numbers of attractors for the mammalian cell cycle network, the CREB signaling network, and the human fibroblast signaling network could not be obtained because there were large numbers of attractors with small basins.

drug targets between the control kernel nodes (the driver nodes, the driver nodes on SBD) and non-kernel nodes (the complements of driver nodes, the complements of driver nodes on SBD) in the human fibroblast signaling network, we performed one-sided two-sample chi-squared tests on a 2 (control kernel nodes versus non-kernel nodes, driver nodes versus the complements of driver nodes, or driver nodes on SBD versus the complements of driver nodes on SBD) \times 2 (drug targets versus the complements of drug targets) contingency table.

- Kim, J. *et al.* Evolutionary design principles of modules that control cellular differentiation: consequences for hysteresis and multistationarity. *Bioinformatics* **24**, 1516–1522 (2008).
- Kim, J. R. *et al.* Reduction of complex signaling networks to a representative kernel. *Sci Signal* **4**, ra35 (2011).
- Kim, T. H., Kim, J., Heslop-Harrison, P. & Cho, K. H. Evolutionary design principles and functional characteristics based on kingdom-specific network motifs. *Bioinformatics* **27**, 245–251 (2011).
- Choi, M., Shi, J., Jung, S. H., Chen, X. & Cho, K.-H. Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to DNA damage. *Science Signaling* **5**, ra83 (2012).
- Ding, S. & Wang, W. Recipes and mechanisms of cellular reprogramming: a case study on budding yeast *Saccharomyces cerevisiae*. *BMC Syst Biol* **5**, 50 (2011).
- Huang, S., Eichler, G., Bar-Yam, Y. & Ingber, D. E. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett* **94**, 128701 (2005).
- Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci U S A* **101**, 4781–4786 (2004).
- Wang, G. *et al.* Process-based network decomposition reveals backbone motif structure. *Proc Natl Acad Sci U S A* **107**, 10478–10483 (2010).
- Liu, Y. Y., Slotine, J. J. & Barabasi, A. L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
- Nepusz, T. & Vicsek, T. Controlling edge dynamics in complex networks. *Nat Phys* **8**, 568–573 (2012).
- Liu, Y. Y., Slotine, J. J. & Barabasi, A. L. Control centrality and hierarchical structure in complex networks. *PLoS One* **7**, e44459 (2012).
- Helikar, T., Konvalina, J., Heidel, J. & Rogers, J. A. Emergent decision-making in biological signal transduction networks. *Proc Natl Acad Sci U S A* **105**, 1913–1918 (2008).
- Saez-Rodriguez, J. *et al.* Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol Syst Biol* **5**, 331 (2009).
- MacArthur, B. D., Ma'ayan, A. & Lemischka, I. R. Systems biology of stem cell fate and cellular reprogramming. *Nat Rev Mol Cell Biol* **10**, 672–681 (2009).
- Chen, L., Liu, R., Liu, Z. P., Li, M. & Aihara, K. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci Rep* **2**, 342 (2012).
- Shin, S. Y. *et al.* Functional roles of multiple feedback loops in extracellular signal-regulated kinase and Wnt signaling pathways that regulate epithelial-mesenchymal transition. *Cancer Res* **70**, 6715–6724 (2010).
- Giacomantonio, C. E. & Goodhill, G. J. A Boolean model of the gene regulatory network underlying Mammalian cortical area development. *PLoS Comput Biol* **6** (2010).
- Alvarez-Buylla, E. R. *et al.* Gene regulatory network models for plant development. *Curr Opin Plant Biol* **10**, 83–91 (2007).
- Krumsiek, J., Marr, C., Schroeder, T. & Theis, F. J. Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network. *PLoS One* **6**, e22649 (2011).
- Sahin, O. *et al.* Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. *BMC Syst Biol* **3**, 1 (2009).
- Abdi, A., Tahoori, M. B. & Emamian, E. S. Fault diagnosis engineering of digital circuits can identify vulnerable molecules in complex cellular pathways. *Sci Signal* **1**, ra10 (2008).
- Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J., 3rd & Doyle, J. Robustness of cellular functions. *Cell* **118**, 675–685 (2004).
- Assenov, Y., Ramirez, F., Schelhorn, S. E., Lengauer, T. & Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics* **24**, 282–284 (2008).
- Dong, J. & Horvath, S. Understanding network concepts in modules. *BMC Syst Biol* **1**, 24 (2007).
- Liu, Y. Y., Slotine, J. J. & Barabasi, A. L. Observability of complex systems. *Proc Natl Acad Sci U S A* **110**, 2460–2465 (2013).
- Kauffman, S., Peterson, C., Samuelsson, B. & Troein, C. Genetic networks with canalizing Boolean rules are always stable. *Proc Natl Acad Sci U S A* **101**, 17102–17107 (2004).
- Huang, S. Reprogramming cell fates: reconciling rarity with robustness. *Bioessays* **31**, 546–560 (2009).
- Kauffman, S., Peterson, C., Samuelsson, B. & Troein, C. Random Boolean network models and the yeast transcriptional network. *Proc Natl Acad Sci U S A* **100**, 14796–14799 (2003).
- Aldridge, B. B., Burke, J. M., Lauffenburger, D. A. & Sorger, P. K. Physicochemical modelling of cell signalling pathways. *Nat Cell Biol* **8**, 1195–1203 (2006).
- Muller, F. J. & Schuppert, A. Few inputs can reprogram biological networks. *Nature* **478**, E4; discussion E4–5 (2011).



31. Ieda, M. *et al.* Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010).
32. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
33. Zhou, J. X. & Huang, S. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. *Trends Genet* **27**, 55–62 (2011).
34. Beasley, D., Bull, D. R. & Martin, R. R. An Overview of Genetic Algorithms.1. Fundamentals. *U Comput* **15**, 58–69 (1993).
35. Wishart, D. S. *et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* **36**, D901–906 (2008).
36. Kuhn, M., von Mering, C., Campillos, M., Jensen, L. J. & Bork, P. STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res* **36**, D684–688 (2008).

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea Government, the Ministry of Education, Science & Technology

(MEST) (2009-0086964 and 2010-0017662). The authors thank Tae-Hwan Kim for his critical reading of this manuscript.

Author contributions

K.-H.C. designed the project and supervised the research; J.K., S.-M.P. and K.-H.C. devised the algorithm and carried out the network analysis; J.K., S.-M.P. and K.-H.C. wrote the manuscript.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Kim, J., Park, S. & Cho, K. Discovery of a kernel for controlling biomolecular regulatory networks. *Sci. Rep.* **3**, 2223; DOI:10.1038/srep02223 (2013).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0>