

RESEARCH ARTICLE

Open Access



Properties of Boolean dynamics by node classification using feedback loops in a network

Yung-Keun Kwon 

Abstract

Background: Biological networks keep their functions robust against perturbations. Many previous studies through simulations or experiments have shown that feedback loop (FBL) structures play an important role in controlling the network robustness without fully explaining how they do it. Hence, there is a pressing need to more rigorously analyze the influence of FBL structures on network robustness.

Results: In this paper, I propose a novel node classification notion based on the FBL structures involved. More specifically, I classify a node as a *no-FBL-in-upstream* (NFU) or *no-FBL-in-downstream* (NFD) node if no feedback loop is involved with any upstream or downstream path of the node, respectively. Based on those definitions, I first prove that every NFU node is eventually frozen in Boolean dynamics. Thus, NFU nodes converge to a fixed value determined by the upstream source nodes. Second, I prove that a network is robust against an arbitrary state perturbation subject to a non-source NFD node. This implies that a network state eventually sustains the attractor despite a perturbation subject to a non-source NFD node. Inspired by this result, I further propose a perturbation-sustainable probability that indicates how likely a perturbation effect is to be sustained through propagations. I show that genes with a high perturbation-sustainable probability are likely to be essential, disease, and drug-target genes in large human signaling networks.

Conclusion: Taken together, these results will promote understanding of the effects of FBL on network robustness in a more rigorous manner.

Keywords: Boolean dynamics, Robustness, Perturbations, Feedback loop, Signaling networks

Abbreviations: FBL, Feedback loop; NFD, No-FBL-in-downstream; NFU, No-FBL-in-upstream

Background

It is well known that bio-molecular networks can keep their regulatory functions robust against various types of external or internal perturbations. For instance, the fate decision mechanism in a bacteriophage life cycle [1], the chemotaxis process in *Escherichia coli* [2], and segmental polarization in *Drosophila melanogaster* [3] were shown to be robust against noisy environments. It is more interesting that the dynamics of a biological network can be highly related to its structural characteristics [4]. In particular, many recent studies have shown that a feedback loop (FBL), a circular chain of interactions, can play an important role in controlling the robustness or susceptibility of

networks [5, 6]. For instance, the negative FBL between MDM2 and p53 maintains an optimal level of p53 and creates appropriate dynamics of p53 expression level changes for a given level of DNA damage [7]. The *Xenopus* cell cycle is also robustly controlled against a certain level of perturbation with the help of several FBLs [8]. It was shown that a high proportion of coherently coupled FBLs can enhance the robustness of a network against state perturbations [9]. The number of FBLs involved at a node was also found to be positively correlated with the node's functional importance [10]. Those simulation studies, however, cannot fully explain how the FBLs influence the network robustness. Hence, there is a pressing need to more rigorously analyze the relationship between FBL structures and network dynamics.

Correspondence: kwonyk@ulsan.ac.kr

School of Electrical Engineering, University of Ulsan, 93 Daehak-ro, Nam-gu, Ulsan 44610, Republic of Korea

To measure network robustness, I herein use a synchronous Boolean network model [9, 11, 12] in which a node state is represented by a Boolean value, and the states of all nodes are synchronously updated at every discrete time step. Every network state moves to another state, and a series of consecutive transitions are represented by a network state trajectory that eventually converges to a fixed-point or cyclic attractor. The attractor can describe various dynamic behaviors in a biological system, such as multi-stability and oscillations. If a node state is perturbed, the trajectory might converge to a different attractor. Therefore, a network is considered robust if the attractor does not change against a perturbation. Some tools have been proposed to quantify the network robustness by simulating the state transitions after randomly initializing the node states [13–17]. They have a limitation in network size for analysis, though, due to the exponential complexity of attractor computation. Therefore, it is a critical issue to find analytic results that can identify trivial parts that do not require further computation of state transitions.

In this paper, I focus on the effects of an FBL on Boolean converging dynamics. A state of a node is propagated to other nodes along a path in a chain of consecutive interactions. Therefore, the state cannot be fed back to the original node if it is not involved in an FBL. In other words, the current state of a node will eventually disappear unless a downstream path constructs an FBL. From that idea, I developed an FBL-based notion to classify nodes in a network. In particular, I defined two sets of nodes, *no-FBL-in-upstream* (NFU) and *no-FBL-in-downstream* (NFD), according to whether the upstream and downstream paths, respectively, involve any FBLs and proved two theorems regarding NFU and NFD nodes. One is that every NFU node is always frozen irrespective of the initial states of other nodes. This implies that the converging values of all NFU nodes are eventually fixed to a value determined by the upstream source nodes. It also means that a network is likely to be susceptible to a perturbation subject to the source nodes. The other is that a network is robust against an arbitrary perturbation subject to a non-source NFD node. In other words, a network state eventually converges to the same attractor despite a state perturbation subject to a non-source NFD node. Inspired by those results, I further developed a perturbation-sustainable probability which indicates how likely it is that a perturbation effect will be sustained through a network state trajectory and showed that it can adequately identify functionally important genes, such as essential, disease-associated, and drug-target genes, in large human signaling networks. Taken together, all of these results will promote understanding of the effects of FBLs on Boolean converging dynamics and reduce the computational costs of state transition-based simulation tools.

Methods

Structural classification of nodes in a network

In this study, a biological network is represented by a directed graph $G(V, E)$ where $V = \{v_1, v_2, \dots, v_N\}$ is a set of nodes and $E = \{e_1, e_2, \dots, e_A\}$ is a set of directed edges (interactions); an edge $e \in E$ is an ordered pair of nodes (v_i, v_j) where $v_i, v_j \in V$. I use some notions from graph theory, including FBL and upstream/downstream paths, to represent the biological networks as follows.

Definition A node u is an input node of v if there exists an interaction from u to v (i.e., $(u, v) \in E$). In addition, in-degree of v means the number of input nodes of v .

Definition A node v is a source node if in-degree of v is zero. It is assumed that the state of a source node is fixed to its initial value over all the time.

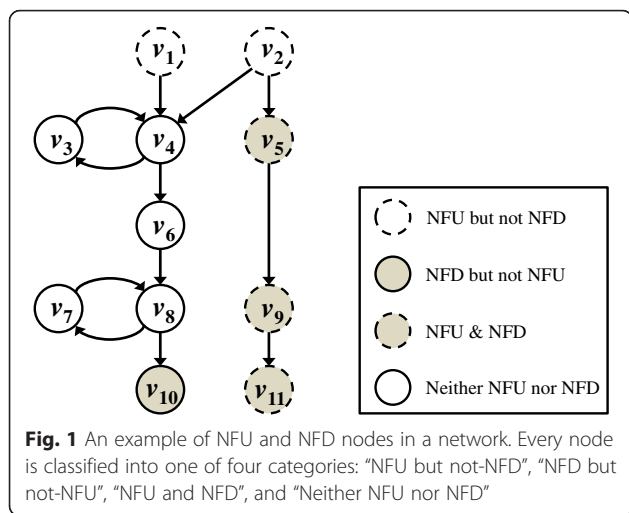
Definition Given a network $G(V, E)$, a *path* P of a length $L (\geq 1)$ is represented by a sequence of ordered nodes $u_1 u_2 \dots u_{L+1}$ with interactions from u_i to u_{i+1} ($(u_i, u_{i+1}) \in E$ for $\forall i \in \{1, 2, \dots, L\}$) with no repeated nodes except u_1 and u_{L+1} . In addition, P is called a *feedback loop* (FBL) if $u_1 = u_{L+1}$.

Definition Given a network $G(V, E)$, an *upstream* (resp., *downstream*) *path* $P = u_1 u_2 \dots u_{L+1}$ of a node $v \in V$ is a path in which the last (resp., *first*) node u_{L+1} (resp., u_1) is v . Note that if P is a feedback loop, then it is both an upstream and downstream path of v . In addition, $P = u_1 u_2 \dots u_{L+1}$ is a *maximal upstream* (resp., *downstream*) *path* if there is no longer path such as $w u_1 u_2 \dots u_{L+1}$ (resp., $u_1 u_2 \dots u_{L+1} w$) for some $w \in V$.

Based on those terms, I define no-FBL-in-upstream and no-FBL-in-downstream nodes as follows:

Definition Given a network $G(V, E)$, a node v is called a *no-FBL-in-upstream* (NFU) (resp., *no-FBL-in-downstream*; NFD) node if there is no upstream (resp., *downstream*) path $P = u_1 u_2 \dots u_L v$ (resp., $P = v u_1 u_2 \dots u_L$) such that for some $i \in \{1, 2, \dots, L\}$, u_i is involved in any feedback loop.

Figure 1 shows an example of NFU and NFD nodes in a network. This network contains five NFU nodes v_1, v_2, v_5, v_9 , and v_{11} and four NFD nodes v_5, v_9, v_{10} and v_{11} . Note that v_5, v_9 and v_{11} are both NFU and NFD nodes. On the other hand, it also contains five nodes that are neither NFU nor NFD nodes because an FBL is both upstream and downstream of each of them. In particular, note that v_6 is not directly involved in an FBL. In this paper, I will show that the NFU and NFD nodes can induce interesting dynamic properties in a perturbation analysis.



A perturbation analysis in a Boolean network model

To define the robustness or sensitivity of a network, I use a synchronous Boolean network model used in previous studies [9, 12, 16]. In a Boolean network $G(V, E)$, each $v_i \in V$ has a value of 1 (on) or 0 (off) that represent the possible states of the corresponding elements. For example, the values 1 and 0 represent the “turn-on” and “turn-off” states of a gene, respectively. A directed interaction (v_i, v_j) can represent a positive (activating) or negative (inhibiting) relationship from v_i to v_j . The value of each variable v_i at time $t + 1$ is determined by the values of k_i other variables $v_{i_1}, v_{i_2}, \dots, v_{i_{k_i}}$ with an interaction to v_i at time t by a Boolean update function $f_i : \{0, 1\}^{k_i} \rightarrow \{0, 1\}$ where f_i is a constant value if v_i is a source node. All the Boolean variables are synchronously updated by a set of update functions $F = \{f_1, f_2, \dots, f_N\}$, and each update rule can be written as $v_i(t + 1) = f_i(v_{i_1}(t), v_{i_2}(t), \dots, v_{i_{k_i}}(t))$.

Many studies have been performed to elucidate the dynamic behaviors of biological networks. In particular, I address robustness against perturbations in terms of Boolean dynamics. In a Boolean network $G(V, E)$, a network state at time t is represented by an ordered list $\mathbf{v}(t) = [v_1(t), v_2(t), \dots, v_N(t)] \in \{0, 1\}^N$. Then, for a subset $U = \{u_1, u_2, \dots, u_M\} \subseteq V$, a subset state of U at time t denoted by $\mathbf{v}_U(t) = [u_1(t), u_2(t), \dots, u_M(t)] \in \{0, 1\}^M$ is an ordered list consisting of values u_1 through u_M at time t . A state trajectory of V starts from an initial state and eventually converges to either a fixed-point or limit-cycle attractor. These attractors can represent diverse biological network behaviors, such as multi-stability and oscillation [18–20]. This notion of an attractor introduces robustness in terms of the Boolean network dynamics as follows. If a network sustains an attractor against a perturbation, it is called robust against the perturbation. This concept has been widely used [5, 21–24]. Here, I consider an initial-state perturbation. Given an initial state

$\mathbf{v}(0) = [v_1(0), v_2(0), \dots, v_N(0)]$ at $t = 0$, an initial-state perturbation subject to node $v_x \in V$ represents a situation in which $\mathbf{v}(0)$ is mutated to $\mathbf{v}'(0) = [v_1(0), \dots, 1 - v_x(0), \dots, v_N(0)]$, i.e., the corresponding initial value is switched to $\bar{v}_x(0)$ (the negation of $v_x(0)$). An initial-state perturbation represents the abnormal (or malfunctioning) status of a protein or gene caused by a mutation. The attractors to which $\mathbf{v}(0)$ and $\mathbf{v}'(0)$ will converge can be compared to each other. The network is called robust or sensitive against the perturbation according to whether the attractors are the same as or different from each other, respectively. Based on this concept, I define some terms more rigorously with respect to the Boolean dynamics, as follows.

Definition The sequence of states to which $\mathbf{v}(0)$ eventually converges is called the attractor induced from \mathbf{v} , which is denoted by an ordered list of network states $\zeta(\mathbf{v}) = [\mathbf{v}(\tau), \mathbf{v}(\tau + 1), \dots, \mathbf{v}(\tau + p - 1)]$ where $\mathbf{v}(t) = \mathbf{v}(t + p)$ for $\forall t \geq \tau$ (p is a length of the attractor), and $\mathbf{v}(i) \neq \mathbf{v}(j)$ for $\forall i \neq j \in \{\tau, \tau + 1, \dots, \tau + p - 1\}$. In addition, for a subset $U \subseteq V$, $\zeta(\mathbf{v}_U) = [v_U(\tau), v_U(\tau + 1), \dots, v_U(\tau + p - 1)]$ represents the states sequence of U in the attractor induced from \mathbf{v} .

Definition For $U \subseteq V$, $\zeta(\mathbf{v}_U)$ is frozen if there exists a time step τ such that $v(t) = v(t + 1)$ for $\forall v \in U$ and $\forall t \geq \tau$.

Definition Given two attractors with a same length, $\zeta(\mathbf{v}) = [\mathbf{v}(\tau), \mathbf{v}(\tau + 1), \dots, \mathbf{v}(\tau + p - 1)]$ and $\zeta(\mathbf{v}') = [\mathbf{v}'(\tau'), \mathbf{v}'(\tau' + 1), \dots, \mathbf{v}'(\tau' + p - 1)]$, they are equivalent to each other if there exists a time step offset $t \geq 0$ such that $\mathbf{v}(\tau + i) = \mathbf{v}'(\tau' + (i + t) \bmod p)$ for $\forall i \in \{0, 1, \dots, p - 1\}$.

Definition Consider an arbitrary initial state $\mathbf{v}(0) = [v_1(0), v_2(0), \dots, v_N(0)]$ and its perturbed state at $v_x \in V$, $\mathbf{v}'(0) = [v_1(0), \dots, 1 - v_x(0), \dots, v_N(0)]$. The Boolean network is called robust against the perturbation subject to v_x if $\zeta(\mathbf{v})$ is equivalent to $\zeta(\mathbf{v}')$. Otherwise, it is called sensitive or susceptible.

Datasets of signaling networks and functionally important genes

In this study, I derive an estimated probability with which a perturbation effect is sustained in a network. To show the usefulness of it, I used two large-scale human signaling networks. One is the signaling network of 1659 genes and 7964 interactions constructed in a previous study [25] by integrating all the human signaling pathways in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database [26] (see Additional file 1: Table S1). The other signaling network consists of 6306 genes and 62,937 interactions (version 6) downloaded from <http://www.bri.nrc.ca/wang> [27] (see Additional file 1: Table S2). In this work, I call them the KEGG and WANG networks, respectively. In addition, I considered essential, disease,

and drug-targeted genes to represent functionally important genes. By using the DEG (Database of Essential Genes, version 5.4) database [28], I found 473 and 1519 essential genes included in the KEGG and WANG networks, respectively. I also found 403 and 1557 disease genes in the KEGG and WANG networks, respectively, using the OMIM (Online Mendelian Inheritance in Man) database [29]. Finally, I identified 353 and 1116 drug target genes in the KEGG and WANG networks, respectively, using the DrugBank database [30].

Results

Dynamic properties of NFU nodes

In this section, I show that NFU nodes are eventually frozen in Boolean network dynamics. A Boolean network $G(V, E)$ with a set of update functions F is given.

Lemma 1 Given an initial state $\mathbf{v}(0)$ and a node $v_k \in V$, let $U = \{u_1, u_2, \dots, u_M\}$ be the set of input nodes of v_k . If $\zeta(\mathbf{v}_U)$ is frozen, then $\zeta(\mathbf{v}_{\{v_k\}})$ is also frozen.

Proof For $\forall u_i \in U$, there exists a time step τ_i such that $u_i(t) = u_i(t+1)$ for $\forall t \geq \tau_i$ since $\zeta(\mathbf{v}_{\{u_i\}})$ is frozen. Let $\tau = \max_{i \in \{1, 2, \dots, M\}}(\tau_i)$. Then, $v_k(t+2) = f_k(u_1(t+1), \dots, u_M(t+1)) = f_k(u_1(t), \dots, u_M(t)) = v_k(t+1)$ for $\forall t \geq \tau$. Thus, $\zeta(\mathbf{v}_{\{v_k\}})$ is also frozen ■

Lemma 2 An initial state $\mathbf{v}(0)$ and a node $v \in V$ are given. If every maximal upstream path of v includes at least one node u where $\zeta(\mathbf{v}_{\{u\}})$ is frozen, then $\zeta(\mathbf{v}_{\{v\}})$ is also frozen.

Proof Let P_1, P_2, \dots, P_M be the list of maximal upstream paths of v , and let $u_i \in P_i$ ($i = 1, 2, \dots, M$) be the node where $\zeta(\mathbf{v}_{\{u_i\}})$ is frozen. Then consider a sub-path P'_i of P_i starting from u_i and ending at v . Define $W = \{w \in V \mid w \in P'_i \text{ for some } i\}$ and let $l(w)$ be the length of the longest path from w to v . Assuming that $L = \max_{w \in W} l(w)$, W can be divided into $L+1$ disjoint subsets W_0, W_1, \dots and W_L where $W_k = \{w \in W \mid l(w) = k\}$. Then $\zeta(\mathbf{v}_{\{w\}})$ for $\forall w \in W$ is frozen by mathematical induction with respect to $l(w)$, as follows. When $k=L$, it is obvious that $\zeta(\mathbf{v}_{\{w\}})$ of every $w \in W_L$ is frozen because $w \in \{u_1, u_2, \dots, u_M\}$. Assume that $\zeta(\mathbf{v}_{\{w\}})$ of every $w \in W_{k+1}$ is frozen and consider an arbitrary element $w' \in W_k$. Then every input node of w' is an element of W_{k+1} . By lemma 1, $\zeta(\mathbf{v}_{\{w'\}})$ is frozen. Thus, $\zeta(\mathbf{v}_{\{w\}})$ of every $w \in W_k$ is also frozen. By mathematical induction, $\zeta(\mathbf{v}_{\{w\}})$ of every $w \in W$ is frozen. Therefore, $\zeta(\mathbf{v}_{\{v\}})$ is also frozen because $v \in W$ ■

Lemma 2 provides a sufficient condition for the frozenness of a node. This can be extended to the case of NFU nodes as follows.

Theorem 1 An initial state $\mathbf{v}(0)$ is given. If $v \in V$ is an NFU node, then $\zeta(\mathbf{v}_{\{v\}})$ is frozen.

Proof By the definition of NFU nodes, every maximal upstream of v starts with a source node u whose $\zeta(\mathbf{v}_{\{u\}})$ is frozen. By Lemma 2, $\zeta(\mathbf{v}_{\{v\}})$ is also frozen ■

This theorem implies that the states of NFU nodes are dependent on the states of source nodes, and this might make the network tend to be susceptible to perturbations subject to source nodes.

Corollary An initial state $\mathbf{v}(0)$ is given. If there is no FBL then $\zeta(\mathbf{v})$ is frozen.

Proof Since there are no FBLs, every $v \in V$ is an NFU node. By Theorem 1, it follows that $\zeta(\mathbf{v})$ is always frozen irrespective of the initial states ■

Theorem 1 and its corollary explain the effect of FBLs on the frozenness of the converging state sequences. More specifically, every NFU node is frozen to a value determined by the set of source nodes included in its upstream paths. I also note that this result is strongly related to previous studies based on synchronous or asynchronous Boolean network models [31–33]. In particular, the corollary corresponds to a previous result having stated that the Boolean dynamics converges to a unique fixed point in an acyclic Boolean network [31]. It is also relevant to the previous results showed that limit-cycle attractors can be induced by negative feedback loops [32, 33]. In addition, Theorem 1 can reduce the computation of attractors in a large scale network by easily obtaining the converging values of the NFU nodes.

Dynamic properties of NFD nodes

In the lemmas and a theorem of this section, I investigate the effect of FBLs on robustness. I consider an arbitrary initial state $\mathbf{v}(0) = [v_1(0), v_2(0), \dots, v_N(0)]$ and a perturbed state at $v_x \in V$ from $\mathbf{v}(0)$, $\mathbf{v}'(0) = [v_1(0), \dots, v_x - 1(0), 1 - v_x(0), v_x + 1(0), \dots, v_N(0)]$ in the following lemmas 3 and 4, and theorem 2. I denote the value of a node $w \in V$ at time t in the trajectories starting from $\mathbf{v}(0)$ and $\mathbf{v}'(0)$ by $\mathbf{v}_{\{w\}}(t)$ and $\mathbf{v}'_{\{w\}}(t)$, respectively.

Lemma 3 Let $\mathbf{v}(0)$ be an initial state, $\mathbf{v}'(0)$ a perturbed state at $v_x \in V$ from $\mathbf{v}(0)$, and $w \in V$ an arbitrary node. If there is no path from v_x to w then $\mathbf{v}_{\{w\}}(t) = \mathbf{v}'_{\{w\}}(t)$ for $\forall t \geq 0$.

Proof The state value of node w is updated irrespective of that of node v_x because there is no path from v_x to w . Thus, the lemma holds ■

Lemma 4 Let $\mathbf{v}(0)$ be an initial state, $\mathbf{v}'(0)$ a perturbed state at $v_x \in V$ from $\mathbf{v}(0)$, $w \in V$ an arbitrary node. Let $Y = \{y \in V \mid y \text{ is included in some path from } v_x \text{ to } w\}$ and $l(y)$

the length of a longest path from v_x to $y \in Y$. If v_x is a non-source node and no node in Y is involved with any FBL, then $\mathbf{v}_{\{w\}}(t) = \mathbf{v}'_{\{w\}}(t)$ for $\forall t \geq l(w) + 1$.

Proof By mathematical induction with respect to $l(y)$, I show that for every $y \in Y$, $\mathbf{v}_{\{y\}}(t) = \mathbf{v}'_{\{y\}}(t)$ for $\forall t \geq l(y) + 1$, as follows. When $l(y) = 0$, it is obvious that $y = v_x$. Then $\mathbf{v}_{\{v_x\}}(t) = \mathbf{v}'_{\{v_x\}}(t)$ for $\forall t \geq 1$ because v_x is a non-source node and involved with no FBL. To prove the inductive step, I assume that the property holds for $l(y) \leq k - 1$. Consider an arbitrary $y \in Y$ such that $l(y) = k$ and let U be the set of input nodes of y . For every $u \in U$, there are two cases to consider: either $u \in Y$ or $u \notin Y$. In case of $u \in Y$, it is obvious that $l(u) \leq k - 1$ by the definition of $l(\cdot)$. By the induction hypothesis, $\mathbf{v}_{\{u\}}(t) = \mathbf{v}'_{\{u\}}(t)$ for $\forall t \geq l(u) + 1$. In case of $u \notin Y$, it means that there is no path from v_x to u . Then $\mathbf{v}_{\{u\}}(t) = \mathbf{v}'_{\{u\}}(t)$ for $\forall t \geq 0$ by lemma 3. From both cases, for $\forall u \in U$, $\mathbf{v}_{\{u\}}(t) = \mathbf{v}'_{\{u\}}(t)$ for $\forall t \geq k$. Then $\mathbf{v}_{\{y\}}(t) = \mathbf{v}'_{\{y\}}(t)$ for $\forall t \geq k + 1$, thereby showing the property holds when $l(y) = k$. Since $w \in Y$, the lemma holds ■

Theorem 2 Let $\mathbf{v}(0)$ be an initial state, $\mathbf{v}'(0)$ a perturbed state at $v_x \in V$ from $\mathbf{v}(0)$, and $w \in V$ an arbitrary node. If v_x is an NFD and non-source node, then the network is robust against a state perturbation subject to v_x .

Proof I show that there exists a constant time T such that $\mathbf{v}_{\{w\}}(t) = \mathbf{v}'_{\{w\}}(t)$ for $\forall t \geq T$ in the following three cases. (i) Case that $w = v_x$: Because w is an NFD and non-source node, $\mathbf{v}_{\{w\}}(t) = \mathbf{v}'_{\{w\}}(t)$ for $\forall t \geq 1$. (ii) Case that w is not connected by any path from v_x : By lemma 3, $\mathbf{v}_{\{w\}}(t) = \mathbf{v}'_{\{w\}}(t)$ for $\forall t \geq 0$. (iii) Case that w is connected by at least one path from v_x : Let $Y = \{y \in V | y \text{ is included in some path from } v_x \text{ to } w\}$ and $l(w)$ be a longest length of those paths, respectively. Because v_x is an NFD node, no node included in Y is involved with any FBL. By lemma 4, $\mathbf{v}_{\{w\}}(t) = \mathbf{v}'_{\{w\}}(t)$ for $\forall t \geq l(w) + 1$. By (i), (ii), and (iii), there exists a constant time T such that $\mathbf{v}_{\{w\}}(t) = \mathbf{v}'_{\{w\}}(t)$ for $\forall t \geq T$ and $\forall w \in V$. Accordingly, the attractors starting at $\mathbf{v}(0)$ and $\mathbf{v}'(0)$ are equivalent to each other. Therefore, the network is robust against the state perturbation subject to v_x ■

Theorem 2 indicates that biological networks might be robust against perturbations subject to NFD nodes. To support this result, I compared NFD and non-NFD gene groups with respect to the proportions of essential genes, disease genes, and drug targets in two human signaling networks, the KEGG and WANG networks (Fig. 2; see Additional file 1: Tables S3 and S4 for details). As shown in Fig. 2, the proportions of essential genes, disease genes, and drug targets among NFD genes were significantly smaller than those among non-NFD genes in both networks (all p -values $< 10^{-10}$). I assume that essential genes, disease genes, and drug targets are likely to be susceptible to mutations, perturbations, or

other external changes. In this regard, the relatively low proportions of essential genes, disease genes, and drug targets in the NFD group in the large-scale signaling networks support Theorem 2. In addition, I further examined the proportions of essential genes, disease genes, and drug targets in NFD group in random networks to examine if the observed result is specific to the signaling networks (see Additional file 2: Figure S1). I created each set of 100 random networks by rewiring the interactions of the KEGG (Additional file 2: Figure S1(A)) and WANG (Additional file 2: Figure S1(B)) networks so that the in-degree and the out-degree of the nodes are conserved, and observed that there is little difference between the NFD and non-NFD groups with respect to the proportions of essential genes, disease genes, and drug targets. This implies that the functionally important genes in the real signaling networks are not randomly distributed in terms of NFD classification.

Estimation of sustainability of a perturbation effect

In the previous section, Theorem 2 showed that a network state is robust against a perturbation as long as the perturbation effect is not sustained by downstream FBLs. In other words, the existence of downstream FBLs is a necessary condition to make a network susceptible to a perturbation. Inspired by that result, I have derived an estimated probability that a perturbation effect will be sustained. Given a node $v_x \in V$ subject to a perturbation, Lemma 3 shows that only downstream paths of v_x need to be considered, and Lemma 4 shows that only those involved with an FBL need to be considered. I first estimate the probability with which a perturbation subject to v_x is sustained through a single path involved with a FBL. Figure 3 shows an example of a downstream path $P = v_x u_1 u_2 \dots u_L$ of v_x which includes an FBL, and I consider $\mathbf{v}(0)$ and $\mathbf{v}'(0)$ which are an initial state and a perturbed state at $v_x \in V$ from $\mathbf{v}(0)$, respectively. It is said that the effect of a perturbation starting at v_x at the initial time is sustained through propagations in a sequence of u_1, u_2, \dots, u_L if $\mathbf{v}_{\{u_i\}}(i) \neq \mathbf{v}'_{\{u_i\}}(i)$ for $\forall i \in \{1, \dots, L\}$. Herein, it is assumed that a probability with which $u_i(i)$ is differently updated by the flipped value of $u_{i-1}(i-1)$, denoted by $\Pr(\mathbf{v}_{\{u_i\}}(i) \neq \mathbf{v}'_{\{u_i\}}(i) | \mathbf{v}_{\{u_{i-1}\}}(i-1) \neq \mathbf{v}'_{\{u_{i-1}\}}(i-1))$, is the inverse of the in-degree of u_i because of the following reason (for simplicity of explanation, $u_0 = v_x$ is assumed). Let $W = \{w_1, w_2, \dots, w_{d_i}\}$ be the set of input nodes of u_i where d_i is the in-degree of u_i . By assuming that the input nodes have an even degree of influence on updating u_i , i.e., $\Pr(Y|X_1) = \dots = \Pr(Y|X_{d_i})$ where Y and $X_k(k \in \{1, \dots, d_i\})$ denote two events $\mathbf{v}_{\{u_i\}}(i) \neq \mathbf{v}'_{\{u_i\}}(i)$ and $\mathbf{v}_{\{w_k\}}(i-1) \neq \mathbf{v}'_{\{w_k\}}(i-1)$, respectively. It is also assumed that u_i is always differently updated given a perturbation has occurred at one of the input nodes. Accordingly, $\Pr(Y | \bigcup_{k \in \{1, \dots, d_i\}} X_k) = \Pr(Y|X_1) + \dots + \Pr(Y|X_{d_i}) = 1$,

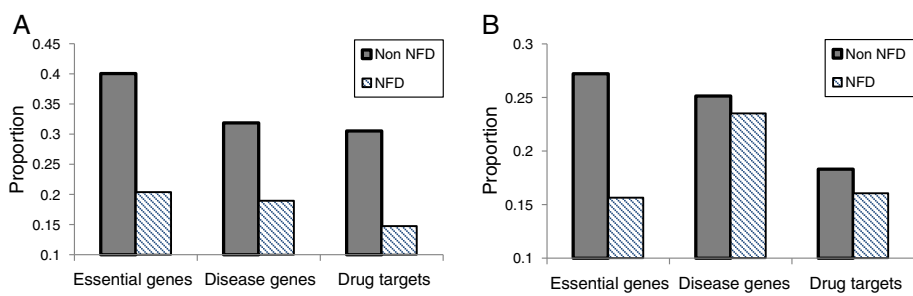


Fig. 2 Comparison between groups of NFD and non-NFD genes in signaling networks. **a** Result in KEGG network with 975 NFD genes and 684 non-NFD genes. The proportions of essential genes in the NFD and the non-NFD groups were 0.204 and 0.401, respectively. The proportions of disease genes in the NFD and the non-NFD groups were 0.190 and 0.319, respectively. The proportions of drug-targets in the NFD and the non-NFD groups were 0.148 and 0.306, respectively. **b** Result in WANG network with 1706 NFD genes and 4600 non-NFD genes. The proportions of essential genes in the NFD and the non-NFD groups were 0.157 and 0.272, respectively. The proportions of disease genes in the NFD and the non-NFD groups were 0.235 and 0.251, respectively. The proportions of drug-targets in the NFD and the non-NFD groups were 0.161 and 0.183, respectively. In both networks, all the proportions for the NFD group were significantly smaller than those for the non-NFD group (all p -values $< 10^{-10}$)

and thus $\Pr(Y|X_k) = 1/d_i$ ($\forall k \in \{1, \dots, d_i\}$). Note that $u_{i-1} \in W$ because u_{i-1} is one of the input nodes of u_i . With this result, the probability with which the perturbation subject to v_x is sustained through a path P can be derived as follows:

$$\Pr(P) = \begin{cases} \prod_{i=1}^L 1/d_i, & \text{if an FBL is involved,} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Let P_1, P_2, \dots, P_M be the set of all downstream paths of v_x . Then I define the perturbation-sustainable probability $\gamma(v_x)$, the probability that the perturbation subject to v_x will be sustained through at least one FBL, as follows:

$$\gamma(v_x) = \max_{i \in \{1, \dots, M\}} \Pr(P_i). \quad (2)$$

If a gene with a relatively high $\gamma(v_x)$ value is subject to a perturbation, the network is likely to induce an abnormal

dynamics due to the well conserved perturbation effect. In this regard, the perturbation-sustainable probability can indicate how much a gene is functionally or dynamically important in a signaling network. To show the usefulness of this probability, I examined the relationship between $\gamma(v_x)$ and the proportions of putatively susceptible genes in human signaling networks (Fig. 4; see Additional file 1: Tables S3 and S4 for details). Given a threshold value β , the proportions of essential genes, disease genes, and drug targets among the set of genes such that $\{v_x | \gamma(v_x) \geq \beta\}$ are plotted against the threshold value in the KEGG (Fig. 4a) and WANG (Fig. 4b) networks. As shown in the figure, the genes with a high perturbation-sustainable probability are more likely to be essential genes, disease genes, and drug targets in both networks. This implies that the perturbation-sustainable probability can adequately identify the functionally important genes in human signaling networks. In addition, it is notable that the relation of the

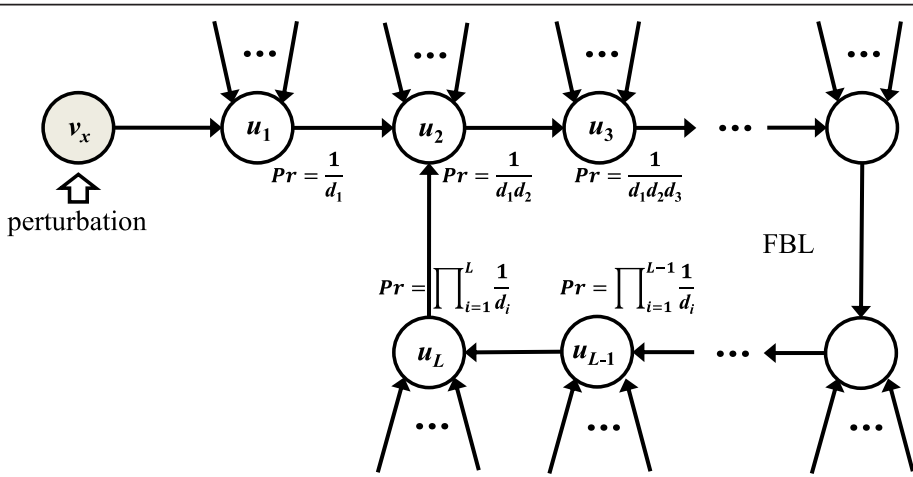
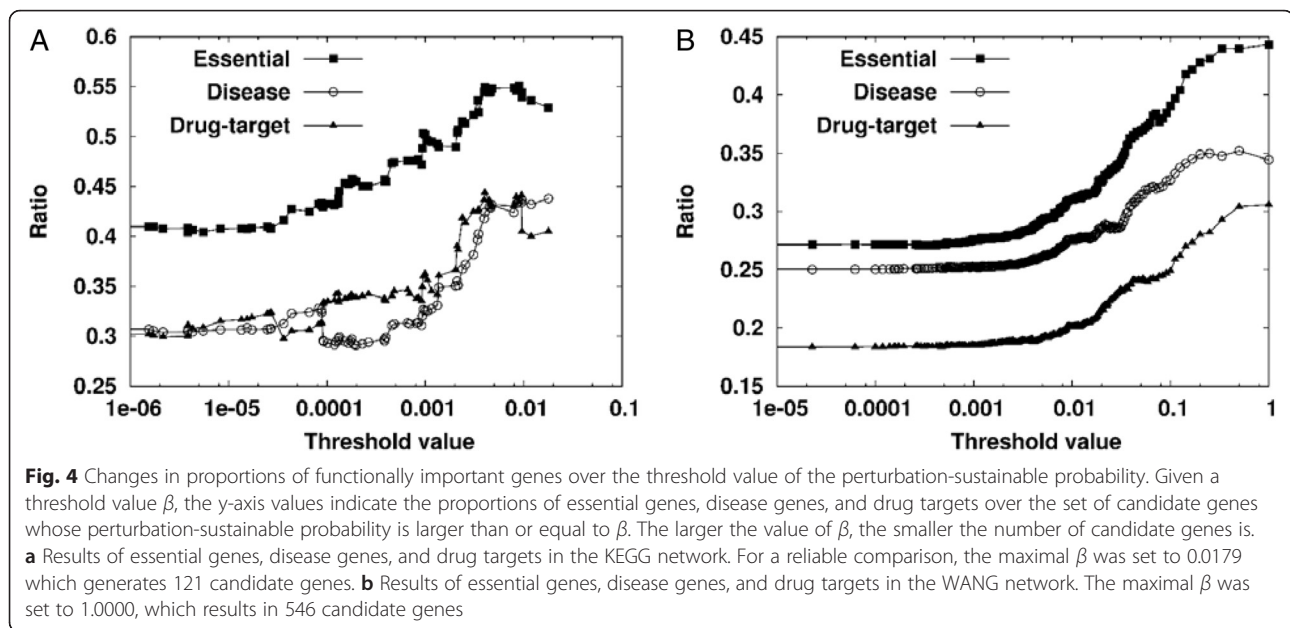


Fig. 3 An example of a downstream path with a feedback loop. A node v_x is subject to a state perturbation, and the perturbation effect can be sustained at u_1 with a probability of the inverse of the in-degree of u_1 . This propagation is continued along the path $P = v_x u_1 u_2 \dots u_L$ involved with a FBL



perturbation-sustainable probability to the functionally important genes was not observed in the random networks created by rewiring the interactions of the signaling networks (see Additional file 2: Figure S2). As in the results of Additional file 2: Figure S1, this also implies that the functionally important genes in the real signaling networks are not randomly distributed in terms of NFD classification. Taken together, it is interesting that such a simple topological measurement of genes based only on FBLs can efficiently predict the functionally important genes in human signaling networks.

Discussion

In this study, I did not address the dynamics of non-NFU and non-NFD nodes, *i.e.*, nodes that are involved in FBLs, and the analysis of their dynamics remains an open problem. In addition, the update-rule perturbation, another well-known type of perturbations, was not considered in this study because it influences the network robustness in a different way than the initial-state perturbation by changing the state transition diagram. Therefore, a future study should include analyses of genes that are neither NFU nor NFD nodes, and analysis of robustness against update-rule perturbations. Finally, it should be noted that the analyses in this study might not be effective for other types of biological networks than the signaling networks. For example, NFU/NFD classification was not meaningful in the large-scale gene regulatory networks [34, 35] because most genes were classified to NFD nodes. This implies that another method to further classify NFD nodes is required for analysis of those networks.

Conclusions

It is well known that biological networks can keep their regulatory functions robust against external or internal perturbations. More interestingly, the network robustness is highly related to the network's structural characteristics, including FBLs. However, previous results [2, 9, 10] have been presented mainly through simulation and experiment studies because of the complexity of real biological networks. That raised a pressing need to develop various analytical approaches to validate the promising conjectures. In this paper, I used a synchronous Boolean network model in which a node state is represented by a Boolean value and updated by a logical rule. A network is considered robust if the attractor does not change against a state perturbation. Based on that assumption, I created a novel concept to characterize the nodes with respect to FBL structures: *no-FBL-in-upstream* (NFU) and *no-FBL-in-downstream* (NFD). This FBL-based characterization is different from other FBL-based measures [10, 36] in that it focuses on involvement with FBLs in the upstream or downstream paths, not with the node itself. Based on that notion, I proved two simple but useful theorems. One is that an NFU node is always frozen irrespective of the initial states of other nodes. Thus, the converging dynamics of an NFU node can be simply determined. The other is that a network is robust against an arbitrary perturbation subject to a non-source NFD node. This result shows that a network state eventually converges to the same attractor despite a perturbation subject to non-source NFD nodes. Note that the two theorems hold for arbitrary update functions as well as initial states. In addition, the second theorem led me to develop a function to approximately compute the perturbation-sustainable

probability. I verified its effectiveness by showing that the higher the probability, the larger the proportion of essential, disease-associated, and drug-target genes in human signaling networks. I believe these results will promote understanding of the effects of FBLs on network dynamics and reduce the cost of computing robustness in existing tools which simulate a network state trajectory [13–17].

Additional files

Additional file 1: Table S1. Dataset of KEGG network. It consists of 1659 genes and 7964 interactions constructed by integrating all the human signaling pathways in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. **Table S2.** Dataset of WANG network. It consists of 6306 genes and 62,937 interactions downloaded from <http://www.bri.nrc.ca/wang>. **Table S3.** Detailed analysis result of the KEGG network. It includes information about essential genes, disease genes, drug targets, NRU/NRD genes, and perturbation-sustainable probabilities in the KEGG network. **Table S4.** Detailed analysis result of the WANG network. It includes information about essential genes, disease genes, drug targets, NRU/NRD genes, and perturbation-sustainable probabilities in the WANG network. (ZIP 352 kb)

Additional file 2: Figure S1. Comparison between groups of NFD and non-NFD genes in random networks with respect to the proportions of essential genes, disease genes, and drug targets. In each subfigure, a set of 100 random networks were generated by rewiring the interactions of the signaling networks such that the in-degree and the out-degree of all nodes are conserved. (A) Result of random networks shuffled from KEGG network. The average numbers of NFD and non-NFD genes were 509 and 1150, respectively. The average proportions of essential genes in the NFD and the non-NFD groups were 0.2849 and 0.2852, respectively. The proportions of disease genes in the NFD and the non-NFD groups were 0.2417 and 0.2435, respectively. The proportions of drug-targets in the NFD and the non-NFD groups were 0.2141 and 0.2122, respectively. (B) Result of random networks shuffled from WANG network. The average numbers of NFD and non-NFD genes were 1544 and 4761, respectively. The proportions of essential genes in the NFD and the non-NFD groups were 0.2390 and 0.2413, respectively. The proportions of disease genes in the NFD and the non-NFD groups were 0.2455 and 0.2472, respectively. The proportions of drug-targets in the NFD and the non-NFD groups were 0.1768 and 0.1769, respectively. **Figure S2.** Changes in the proportion of functionally important genes over the threshold value of the perturbation-sustainable probability in random networks. In each subfigure, a set of 100 random networks were generated by rewiring the interactions of the signaling network such that the in-degree and the out-degree of all nodes are conserved. Given a threshold value β , the y-axis values indicate the average proportions of essential genes, disease genes, and drug targets over the set of candidate genes whose perturbation-sustainable probability is larger than or equal to β in random networks. (A) Results in random networks shuffled from KEGG network. For a reliable comparison, the maximal β was set to 0.0266 which generates 131 candidate genes on average. (B) Results in random networks shuffled from WANG network. The maximal β was set to 0.0908, which results in 121 candidate genes on average. (ZIP 362 kb)

Acknowledgements

Not applicable.

Funding

This work was supported by the 2015 Research Fund of University of Ulsan.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Authors' contributions

YKK conceived the study, designed the analyses, wrote and revised the manuscript.

Competing interests

The author declares that he has no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Received: 20 January 2016 Accepted: 14 July 2016

Published online: 24 August 2016

References

- Little JW, Shepley DP, Wert DW. Robustness of a gene regulatory circuit. *EMBO J.* 1999;18(15):4299–307.
- Yi TM, Huang Y, Simon MI, Doyle J. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci U S A.* 2000;97(9):4649–53.
- Ingolia NT. Topology and robustness in the Drosophila segment polarity network. *PLoS Biol.* 2004;2(6):e123.
- Klein C, Marino A, Sagot MF, Vieira Milreu P, Brilli M. Structural and dynamical analysis of biological networks. *Brief Funct Genomics.* 2012;11(6):420–33.
- Kitano H. Biological robustness. *Nat Rev Genet.* 2004;5(11):826–37.
- Kitano H. Towards a theory of biological robustness. *Mol Syst Biol.* 2007;3:137.
- Lev Bar-Or R, Maya R, Segel LA, Alon U, Levine AJ, Oren M. Generation of oscillations by the p53-Mdm2 feedback loop: a theoretical and experimental study. *Proc Natl Acad Sci U S A.* 2000;97(21):11250–5.
- Morohashi M, Winn AE, Borisuk MT, Bolouri H, Doyle J, Kitano H. Robustness as a measure of plausibility in models of biochemical networks. *J Theor Biol.* 2002;216(1):19–30.
- Kwon YK, Cho KH. Coherent coupling of feedback loops: a design principle of cell signaling networks. *Bioinformatics.* 2008;24(17):1926–32.
- Kwon YK, Choi SS, Cho KH. Investigations into the relationship between feedback loops and functional importance of a signal transduction network based on Boolean network modeling. *BMC Bioinformatics.* 2007;8:384.
- Kauffman S, Peterson C, Samuelsson B, Troein C. Random Boolean network models and the yeast transcriptional network. *Proc Natl Acad Sci U S A.* 2003;100(25):14796–9.
- Le DH, Kwon YK. A coherent feedforward loop design principle to sustain robustness of biological networks. *Bioinformatics.* 2013;29(5):630–7.
- Gonzalez AG, Naldi A, Sanchez L, Thieffry D, Chaouiya C. GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. *BioSystems.* 2006;84(2):91–100.
- Mussel C, Hopfensitz M, Kestler HA. BoolNet—an R package for generation, reconstruction and analysis of Boolean networks. *Bioinformatics.* 2010;26(10):1378–80.
- Zheng J, Zhang D, Przytycki PF, Zielinski R, Capala J, Przytycka TM. SimBoolNet—a Cytoscape plugin for dynamic simulation of signaling networks. *Bioinformatics.* 2010;26(1):141–2.
- Trinh HC, Le DH, Kwon YK. PANET: a GPU-based tool for fast parallel analysis of robustness dynamics and feed-forward/feedback loop structures in large-scale biological networks. *PLoS One.* 2014;9(7):e103010.
- Bock M, Scharp T, Talnikar C, Klipp E. BooleSim: an interactive Boolean network simulator. *Bioinformatics.* 2014;30(1):131–2.
- Bhalla US, Ram PT, Iyengar R. MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science.* 2002;297(5583):1018–23.
- Ferrell Jr JE, Machleder EM. The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. *Science.* 1998;280(5365):895–8.
- Pomerening JR, Sontag ED, Ferrell Jr JE. Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nat Cell Biol.* 2003;5(4):346–51.
- Ciliberti S, Martin OC, Wagner A. Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS Comput Biol.* 2007;3(2):e15.

22. Huang S, Eichler G, Bar-Yam Y, Ingber DE. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett*. 2005; 94(12):128701.
23. Kitano H. Cancer as a robust system: implications for anticancer therapy. *Nat Rev Cancer*. 2004;4(3):227–35.
24. 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*. 2004;101(14):4781–6.
25. Kim JR, Kim J, Kwon YK, Lee HY, Heslop-Harrison P, Cho KH. Reduction of complex signaling networks to a representative kernel. *Sci Signal*. 2011; 4(175):ra35.
26. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
27. Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, Yang S, Zhang S, Liu L, Lu M, O'Connor-McCourt M, Purisima EO, Wang E. A map of human cancer signaling. *Mol Syst Biol*. 2007;3:152.
28. Zhang R, Lin Y. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res*. 2009;37(Database issue):D455–8.
29. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res*. 2009;37(Database issue):D793–6.
30. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, Wishart DS. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res*. 2011;39(Database issue):D1035–41.
31. Robert F. *Discrete iterations: a metric study*. Berlin: Springer Verlag; 1986.
32. Thomas R, Thieffry D, Kaufman M. Dynamical behaviour of biological regulatory networks—I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull Math Biol*. 1995;57(2):247–76.
33. Kwon YK, Cho KH. Boolean dynamics of biological networks with multiple coupled feedback loops. *Biophys J*. 2007;92(8):2975–81.
34. Bovolenta LA, Acencio ML, Lemke N. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics*. 2012;13:405–2164-13-405.
35. Liu ZP, Wu C, Miao H, Wu H. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database (Oxford)*. 2015. doi:10.1093/database/bav095. Print 2015.
36. Kwon YK, Cho KH. Quantitative analysis of robustness and fragility in biological networks based on feedback dynamics. *Bioinformatics*. 2008;24(7):987–94.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

