BIOINFORMATICS ORIGINAL PAPER
Vol. 27 no. 8 2011, pages 1113–1120
doi:10.1093/bioinformatics/btr082

Systems biology

Advance Access publication February 15, 2011

The effects of feedback loops on disease comorbidity in human signaling networks

Duc-Hau Le and Yung-Keun Kwon*
School of Computer Engineering and Information Technology, University of Ulsan, 93 Daehak-ro, Nam-gu, Ulsan 680-749, Korea

Associate Editor: Olga Troyanskaya

ABSTRACT

Motivation: In general, diseases are more likely to be comorbid if they share associated genes or molecular interactions in a cellular process. However, there are still a number of pairs of diseases which show relatively high comorbidity but do not share any associated genes or interactions. This observation raises the need for a novel factor which can explain the underlying mechanism of comorbidity. We here consider a feedback loop (FBL) structure ubiquitously found in the human cell signaling network as a key motif to explain the comorbidity phenomenon, since it is well known to have effects on network dynamics.

Results: For every pair of diseases, we examined its comorbidity and length of all FBLs involved by the disease-associated genes in the human cell signaling network. We found that there is a negative relationship between comorbidity and length of involved FBLs. This indicates that a disease pair is more likely to comorbid if they are connected with FBLs of shorter length. We additionally showed that such a negative relationship is more obvious when the number of positive involved FBLs is larger than that of negative involved FBLs. Moreover, we observed that the negative relationship between comorbidity and length of involved FBLs holds especially for disease pairs that do not share any disease-associated genes. Finally, we proved all these results through intensive simulations, based on a Boolean network model.

Contact: kwonyk@ulsan.ac.kr

Supplementary information: Supplementary data are available at *BioInformatics* online.

Received on August 5, 2010; revised on January 18, 2011; accepted on February 2, 2011

1 INTRODUCTION

When two disorders or illnesses occur in the same person, simultaneously or one after the other, they are called comorbid. It is reported that 80% of the elderly population has three or more chronic conditions (Caughey *et al.*, 2008), which explains the importance of understanding disease comorbidity. There have been many studies observing the comorbidity phenomenon. Some studies have found high comorbidity rates of certain pairs of diseases (Gabriel and Michaud, 2009; Raja and Azzoni *et al.*, 2008; Ye *et al.*, 2005) and addressed statistical results of prevalence and mortality of specified regions (Caughey *et al.*, 2008; Tetsche *et al.*, 2008). Other studies attempted to quantify the effect of a disease on other diseases by introducing comorbidity measures (Kelli *et al.*, 2005; Tang *et al.*, 2008). In a recent study of illness progression, a database was also constructed to summarize statistical correlations between phenotypic diseases from histories of more than 30 million patients in a phenotypic disease network (Hidalgo *et al.*, 2009).

Another class of previous studies investigated factors to explain the cause of comorbidity. For example, Goh *et al.* (2007) constructed a human disease network in which a pair of diseases are linked when they share common disease-causing genes, and showed a common genetic origin of many diseases from this constructed network. In another study, Park *et al.* (2009) found statistically significant correlations between an underlying structure of cellular networks and comorbidity patterns in the human population by combining information on protein interactions, disease-gene associations and population-level disease patterns extracted from Medicare data. In that paper, the authors showed positive correlations between the degree of comorbidity, the number of shared genes and the number of shared protein interactions. A bipartite graph was also constructed in which nodes represent diseases and two diseases are linked if mutated enzymes associated with them catalyze adjacent metabolic reactions (Lee *et al.*, 2008). Based on this graph, it was shown that two connected metabolic diseases sharing some pathways tend to show significant comorbidity. This result is consistent with a general proteomic notion that diseases may be related if they share protein interactions (Park *et al.*, 2009) or proteins acting on the same pathway (Calvano *et al.*, 2005; Goehler *et al.*, 2004; Lim *et al.*, 2006; Oldham *et al.*, 2006; Pujana *et al.*, 2007; Rual *et al.*, 2005; Stelzl *et al.*, 2005). Taken together, it can be generally accepted that diseases that share associated genes or molecular interactions in a cellular process are more likely to be comorbid. However, it is interesting that many pairs of diseases show high comorbidity even though they do not share any associated genes or interactions (Park *et al.*, 2009). Therefore, there is still a pressing need to find other factors which is related to comorbidity.

In this study, we consider a FBL (FBL) structure as a novel factor to explain the comorbidity phenomenon. FBLs are a well-known critical motif to affect dynamics in biological networks (Mendoza *et al.*, 1999; Milo *et al.*, 2002; Prill *et al.*, 2005; Snoussi, 1998; Yeger-Lotem *et al.*, 2004). In particular, FBLs were shown to play an important role in robustly sustaining steady state of networks against perturbations (Kwon and Cho, 2008; Kwon *et al.*, 2007). In this regard, we investigated the relationship between comorbidity and FBLs involved with disease-associated genes in a human signaling network. Using integrated data from a disease–gene association database and a human cell-signaling network, we show that there is a negative relationship between comorbidity and the length of

*To whom correspondence should be addressed.

involved FBLs. In other words, a pair of diseases is highly comorbid if their associated genes are connected with FBLs of relatively short length. It is interesting to note that this relationship is valid especially for disease pairs that do not share any associated genes. Moreover, such a negative relationship is more clearly observed when positive FBLs are more abundant than negative FBLs between a disease pair. We also show the negative relationship between comorbidity and the length of FBLs through intensive simulations, based on random Boolean network models.

2 METHODS

2.1 Datasets

To obtain comorbidity information between pairs of diseases, we used the published dataset (Park *et al.*, 2009) containing two quantified comorbidity values, relative risk (RR) and ϕ-correlation coefficient (PHI), for a total of 83 924 disease pairs. The two comorbidity measures were defined as $RR = C_{ij}/C^*_{ij}$ and $PHI = (NC_{ij} - I_iI_j)/\sqrt{I_iI_j(N-I_i)(N-I_j)}$, respectively, where N is the number of patients, I_i denotes incidence of disease i, C_{ij} denotes the number of patients who were simultaneously diagnosed with diseases i and j, respectively and $C^*_{ij} = I_iI_j/N$. In that study, the Medicare database which includes the clinical history of 13 039 018 patients was used for comorbidity evaluation. When two diseases cooccur more frequently than expected by chance, we have $RR > 1$ and $PHI > 0$. Each measure was reported to carry unique biases that are complementary (Hidalgo *et al.*, 2009). Additionally, we defined the morbidity of a disease i as I_i/N(i.e. the prevalence of a disease).

In this aarticle, we analyze a large-scale human signaling network to find a topological characteristic related to the comorbidity phenomenon. To this end, we first obtained the human signaling network (Cui *et al.*, 2007) which had been built up by integrating one cancer-related signaling network obtained from Cancer Cell Map (http://cancer.cellmap.org/cellmap/), and three general signaling networks, which are not cancer-specific, obtained from BioCarta (http://www.biocarta.com) and two previous results of Ma'ayan *et al*. (2005) and Awan *et al.* (2007), respectively. We removed 62 small molecules such as Ca^{++}, glucocorticoid, and H_2O_2 and their interactions from the network. As a result, the network consisting of 1517 nodes and 4761 links was constructed for this investigation (Table S1 in Supplementary Material). In addition, we also obtained the entire list of disease-gene associations from the previous study (Park *et al.*, 2009). After matching these databases (See Figure S1 in Supplementary Material for more details of database relationship diagram), we finally constructed a list of 334 diseases, their associated genes (Table S2 in Supplementary Material) and their corresponding names (Table S3 in Supplementary Material). We also collected comorbidity values of 18 896 pairs of those diseases (Table S4 in Supplementary Material).

2.2 Definitions of topological properties in a network

We defined some topological properties with respect to either a single gene or a set of genes. In this article, we consider a network represented by a directed graph $G = (V, A)$, where V is a set of nodes and A is the set of ordered pairs of the nodes called directed links. A directed link (v_i, v_j) is assigned with either a positive ('activating') or negative ('inhibiting') relationship from $v_i \in V$ to $v_j \in V$. When the human signaling network is represented by $G(V, A)$, we denote the set of genes associated with a disease D by $V(D) \subseteq V$ and then $|V(D)|$ represents the number of genes associated with D. For a gene v we consider the connectivity of v which is defined as the number of links involving v. Moreover, connectivity of D is defined as the average connectivity over the set of genes in $V(D)$.

In this article, we consider FBLs as an important topological property. FBLs are ubiquitously found and play an important role in dynamical behaviors of cellular signaling networks (Milo *et al.*, 2002; Prill *et al.*, 2005; Yeger-Lotem *et al.*, 2004). A FBL is defined as a circular chain of relationships. For example, given a network $G(V, A)$, $v_0 \to v_1 \to v_2 \to \ldots \to v_{L-1} \to v_L$ is a FBL of length $L(\geq 2)$ if there are links from v_{i-1} to v_i for all $i = 1, 2, \ldots, L$ with $v_0 = v_L$ and $v_j \neq v_k$ for $j, k \in \{0, 1, \ldots, L-1\}$, then the number of FBLs involved with a node v, denoted by *NuFBL*(v), is defined as the number of different FBLs involved with v. Similarly, the number of FBLs involved with a disease D, denoted by *NuFBL*(D), is defined as the average *NuFBL*(v) over $\{v | v \in V(D)\}$. In addition, the sign of a FBL is easily determined by the parity of the number of negative relationships involved. If the parity number is even or zero, the sign is positive; otherwise, it is negative. We denote the number of positive and negative FBLs by $NuFBL_+$ and $NuFBL_-$, respectively.

To analyze topological properties between a pair of nodes or diseases, we extend the definitions with respect to FBLs, as follows. Given a network $G(V, A)$ and a pair of nodes $v \in V$ and $v' \in V$, we call v and v' connected with a FBL of a maximal length L if there exists at least one FBL of length $\leq L$ involved with both v and v'. In a similar way, when $G(V, A)$ represents the human signaling network and a pair of diseases, D and $D\prime$, are given, D and D' are called connected with a FBL of a maximal length L if there exists at least one pair of genes, $v \in V(D)$ and $v' \in V(D')$,such that v and v' are connected with a FBL of a maximal length L. Additionally, we denote the number of FBLs between a pair of nodes or diseases by *NuFBL*(v, v') or *NuFBL*(D, D'), respectively.

2.3 Definitions of dynamical properties in a network

To prove our hypothesis, we employed a Boolean network model, which has been widely used to represent biological networks and successfully captured some biological characteristics (Kauffman *et al.*, 2003; Kauffman *et al.*, 2004; Kwon and Cho, 2007; Shmulevich *et al.*, 2003, 2005). In particular, it has been also frequently used in simulating the dynamics of various signaling networks such as a guard cell abscisic acid signaling (Saadatpour *et al.*, 2010), a central intrinsic and extrinsic apoptosis pathway (Mai and Liu, 2009; Schlatter *et al.*, 2009), a mammalian Epidermal Growth Factor Receptor (EGFR) signaling pathway (Sahin *et al.*, 2009), a T-cell receptor signaling (Saez-Rodriguez *et al.*, 2007), a neurotransmitter signaling pathway (Saez-Rodriguez *et al.*, 2007) and so on.

2.3.1 A random Boolean network When a Boolean network is represented by a directed graph G(V, A), each $v_i \in$V has a value of 1 ('on') or 0 ('off'), which represents the possible states of the corresponding elements. The value of each variable v_i at time $t+1$ is determined by the values of k_i other variables $v_{i_1}, v_{i_2}, \ldots, v_{i_{k_i}}$, with a link to v_i at time t by the Boolean function $f_i : \{0, 1\}^{k_i} \to \{0, 1\}$. Hence, we can write the update rule as $v_i(t+1) = f_i(v_{i_1}, v_{i_2}, \ldots, v_{i_{k_i}})$ where we randomly select either a logical conjunction or disjunction for all signed relationships in f_i with a uniform probability distribution. For example, if a Boolean variable v has a positive relationship from v_1, a negative relationship from v_2 and a positive relationship from v_3, then the conjunction and disjunction update rules are $v(t+1) = v_1(t) \wedge \bar{v}_2 \wedge v_3(t)$ and $v(t+1) = v_1(t) \vee \bar{v}_2(t) \vee v_3(t)$, respectively. In the case of a conjunction, the value of v at time $t+1$ is 1 only if the values of v_1, v_2 and v_3 at time t are 1, 0 and 1, respectively whereas, in the case of a disjunction, the value of v at time $t+1$ is 1 if at least one of the states of the clauses, $v_1(t)$, $\bar{v}_2(t)$, and $v_3(t)$ is 1. Although there can be many other logical functions in addition to conjunction and disjunction functions, biological networks were successfully described by Boolean models using only those two functions in many previous studies (Albert, 2004; Faure *et al.*, 2006; Helikar *et al.*, 2008; Huang and Ingber, 2000; Kwon and Cho, 2007). In addition, the sign of each link is determined between positive and negative ones uniformly at random.

To generate a large number of random Boolean networks, we considered three models. The first model generates random Boolean networks in a way that the connectivity of every node is ≥ 1. On the other hand, the second model generates random Boolean networks in a way that every node has at least one incoming link and at least one outgoing link. These two models are denoted as Model-A and Model-B, respectively. The difference between the

two models is that the first model permits presence of input and output nodes, while the second one does not. (An input or output node means one which has no incoming or outgoing link, respectively, and there is no employed constraint on the number of input and output nodes in Model-A.) Thus, the reason why we considered those different models is to show that our simulation results are not dependent on the presence of input and output nodes in the Boolean networks. Two visualization examples of random Boolean networks generated by Model-A and Model-B are shown in Figure S2 in Supplementary Material. In addition to the two models, we considered the other model proposed by Barabasi and Albert (1999) which can generate random networks with a scale-free property, namely a power-law degree distribution (see Figure S3 in Supplementary Material for a more detailed generation process). We employed this model, which is denoted by Model-C, to examine whether our simulation results also hold for scale-free networks.

Given a Boolean network with N Boolean variables, $v_1, v_2, \ldots, v_N$, we define a network *state* as a vector consisting of values of the Boolean variables: there are 2^N states in total. Each state transits to another state through a set of N Boolean update functions, $f_1, f_2, \ldots, f_N$. We can construct a *state transition diagram* that represents the transition of each state. A state trajectory starts from an initial state and eventually converges to either a fixed-point or a limit-cycle attractor. Attractors can represent diverse behaviors of biological networks, such as multi-stability, homeostasis and oscillation (Bhalla *et al.*, 2002; Ferrell *et al.*, 1998; Pomerening *et al.*, 2003). In addition, we define a transient sequence of values of a node v as follows: When a Boolean network $G(V,A)$ was initialized with $v_1(0), v_2(0), \ldots$, and $v_N(0)$ at the starting time 0, $v_i(t_0, t_1)$ represents a sequence of the transient values of a node v_i during the time interval from t_0 to t_1.

2.3.2 Mutual-effectiveness in a random Boolean network In Boolean networks, we propose a novel measure, mutual-effectiveness, to quantify the mutual influence between a pair of nodes or node groups in terms of the network dynamics. To define it, we first introduce two types of perturbations, an initial-state perturbation and an updating-rule perturbation. Given a Boolean network initialized with $v_1(0), v_2(0), \ldots$, and $v_N(0)$, the initial-state perturbation at a node $v_i \in V$ means flipping $v_i(0)$ to $\bar{v}_i(0)$. On the other hand, the updating-rule perturbation at a node $v_i \in V$ means switching the updating-rule at v_i from a conjunction function to a disjunction function or vice versa, depending on the current function type. Assuming a perturbation at v_i, we define the effectiveness from v_i to another node v_j, $\mu(v_i, v_j)$, as follows:

(i) Let τ_i, the valid convergent time of v_i, defined as $\tau_i = \max\{T_i, T_i'\}$ where T_i or T_i' represent the time steps for the network to converge to an attractor when v_i was subject to the perturbation or not, respectively.
(ii) We obtain two different transient sequences of v_j, $v_j(0, \tau_i)$ and $v_j'(0, \tau_i)$, when v_i was subject to the perturbation or not, respectively.
(iii) Then, we compute $\mu(v_i, v_j) = d(v_j(0, \tau_i), v_j'(0, \tau_i))/\tau_i$ where $d(\bullet)$ means the Hamming distance (i.e. the number of bits having different values) between *two sequences*. Thus, $\mu(v_i, v_j)$ represents how largely the trajectory with respect to v_j was affected by the perturbation at v_i.

Since μ is not commutative, we derive the mutual-effectiveness for a pair of nodes v_i and v_j , ρ (v_i, v_j), as follows:

$$\rho(v_i, v_j) = \frac{(\mu(v_i, v_j) + \mu(v_j, v_i))}{2}$$

Therefore, mutual-effectiveness is a measure about how largely each node is mutually affected by perturbation at the other node in terms of dynamics. In this regard, mutual-effectiveness in Boolean networks can be used to represent the comorbidity phenomenon in signaling networks. Figure 1 shows an example of the calculation of mutual-effectiveness of a node pair, v_4 and v_7. To compute $\mu(v_4, v_7)$, we get two transient sequences of v_7, $v_7(0,\tau_4)$ and $v_7'(0,\tau_4)$, *when* v_4 was subject to a perturbation or not, respectively. In the same way, $\mu(v_7, v_4)$ is computed and finally $\rho(v_4, v_7)$ are obtained by averaging $\mu(v_4, v_7)$ and $\mu(v_7, v_4)$.

In a Boolean network, a node is called a functional important node if a perturbation at the node makes the network converge to another attractor, which is different from the original attractor to which the network converged when the node was not subject to the perturbation. In this article, we focus on the mutual effectiveness of only functional important nodes since disease genes can be considered as a kind which affect the cellular dynamical behavior. In all simulations of this study, we generated random Boolean networks, such that the ratio of functionally important nodes over the total number of nodes is ≥ 0.05.

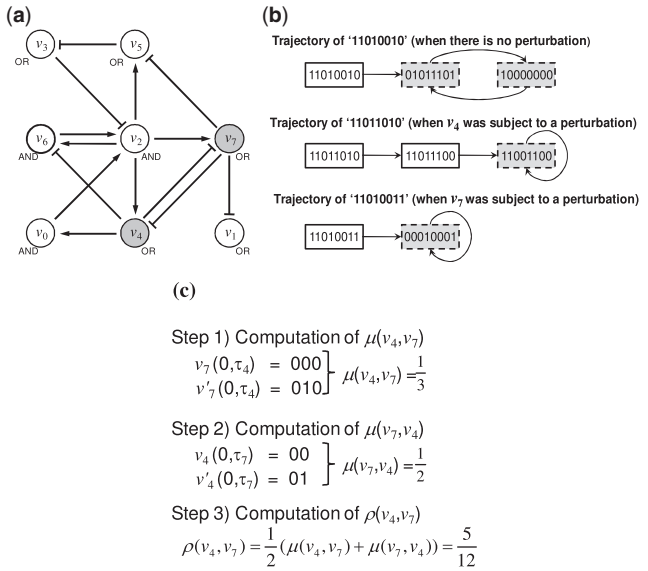

Fig. 1. An illustrative example of calculating mutual-effectiveness in a Boolean network. (**a**) A Boolean network with eight nodes and 14 links where arrows and bar-headed lines represent positive and negative interactions, respectively. 'AND' and 'OR' denote conjunction and disjunction update functions, respectively. (**b**) Trajectories starting from an initial state (11010010) and two other states (11011010 and 11010011) where v_4 or v_7 is subject to an initial-state perturbation, respectively. Eight-bit strings in rectangles represent values of v_0 through v_7 in sequence and grayed rectangles with dashed lines mean attractors. They are calculated from the network in (a). (**c**) Process of computing the mutual-effectiveness between v_4 and $v_7$with respect to the trajectories in (b).

3 RESULTS

3.1 Analysis of morbidity in terms of the number of disease-associated genes, connectivity and number of FBLs

Before we investigated comorbidity, we first addressed how well morbidity could be explained in terms of some topological characteristics in the human signaling network. Morbidity of a disease was defined as the prevalence of a disease (see Section 2 for the definition) and three topological properties were considered for analysis: the number of associated genes of a disease, the connectivity of a disease and the number of FBLs involved with a disease (see Section 2 for the definitions). We plotted the relation of disease prevalence to each topological property (Fig. 2). This result shows that the correlation between disease prevalence and the three topological properties is very small. In other words, morbidity of a disease is not easy to simply understand in terms of topological properties in the human signaling network.

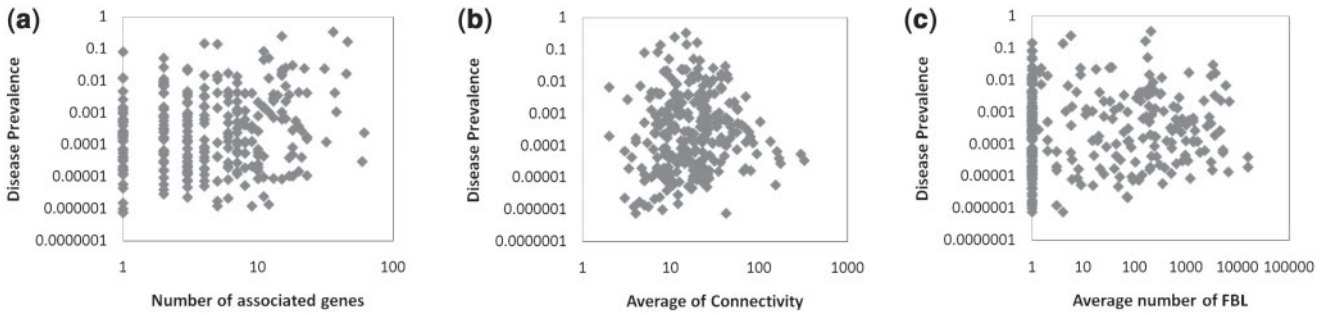

Fig. 2. Correlations between disease prevalence and topological properties over 334 diseases. Considered topological properties are (**a**) the number of genes associated with a disease, (**b**) the connectivity of a disease and (**c**) the number of FBLs involved with a disease. Pearson correlation coefficients are 0.260, −0.065 and −0.041, respectively. All axes are logarithmic in scale.

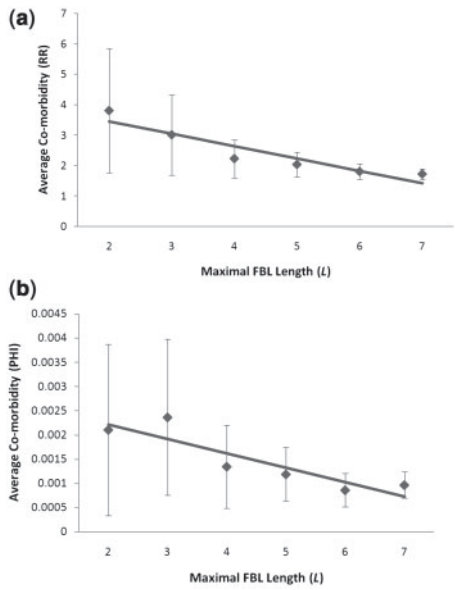

Fig. 3. The relationship between average comorbidity and maximal FBL length in the human signaling network. All *y*-axis values represent the average comorbidity with a 95% confidence level. Blue lines represent linear regression of the average comorbidity values. (**a**) Result of RR comorbidity (slope of linear regression ≈ −0.40655) (**b**) Result of PHI comorbidity (slope of linear regression ≈ −0.0003).

3.2 Analysis of comorbidity in terms of length of FBLs in the human signaling network

We investigated the relationship of FBLs to comorbidity in the human signaling network as follows: when a length L was specified, we collected a set of pairs of diseases which are connected to a FBL of length $\leq L$ (see Section 2 for the definition) and computed the average comorbidity over the set of collected pairs of diseases (Fig. 3). Varying L from 2 to 7, we examined two kinds of comorbidity values, i.e. RR (Fig. 3a) and PHI (Fig. 3b). RR is maximal when L is 2 while PHI is so when L is 3. Although there is such a difference between the peak points of RR and PHI, the relationship trend between each comorbidity value and the maximal FBL length is interestingly negative (P-value = 0.00324 in Fig. 3a and P-value = 0.00824 in Fig. 3b). In other words, the comorbidity value of a pair of diseases is more likely to be high as their associated genes are involved with FBLs of shorter lengths Considering it was not easy to find any topological property correlated to morbidity in Figure 2, this finding is intriguing. Moreover, other topological properties such as the length of the shortest path between a disease pair and the number of FBLs connecting a pair of diseases did not show any obvious relationship to comorbidity values (see Figure S4a and b in Supplementary Material).

3.3 The effect of FBL length on mutual-effectiveness in random Boolean networks

To understand why the comorbidity trend is negatively related to FBL length, we performed extensive simulations based on random Boolean networks models. We generated 100 random Boolean networks with $|V| = 50$ and $|A| = 75$, collected a group of functional important node pairs, which are connected with a FBL of length $\leq L$ by varying L from 2 to 10 and examined the mutual-effectiveness of each group (Fig. 4). We used three kinds of random Boolean networks models: one that permits the presence of input/output nodes (Model-A; Fig. 4a), another does not (Model-B; Fig. 4b) and the other generates scale-free networks (Model-C; Fig. 4c) (see Section 2 for the definitions). In addition, we considered two types of perturbations: an initial-state perturbation and an update-rule perturbation (see Section 2 for the definitions). We observed a strong negative relationship between the maximal FBL length and mutual-effectiveness (all P-values <0.001), irrespective of the generation models and types of perturbations. Moreover, we also performed the same simulations with random Boolean networks of different network sizes ($|V|$) and different network densities (ratio of $|A|$ over $|V|$). We found that the relationship between the maximal FBL length and mutual-effectiveness is consistently negative, irrespective of network size and density (See Figure S5 in Supplementary Material). From these observations, we can conclude that the shorter the involved FBL length is, the greater the mutual-effectiveness between two nodes is. Considering that mutual-effectiveness between a pair of nodes can represent the potential degree of comorbidity, the simulation result in the random Boolean networks is consistent with the observation in the human signaling network in Figure 3. In addition, we conclude that the reason why mutual-effectiveness is affected by the length of involved FBLs is as follows. A FBL of longer length involves a large number of other nodes, and thus a perturbation effect at a point cannot be well transferred to other

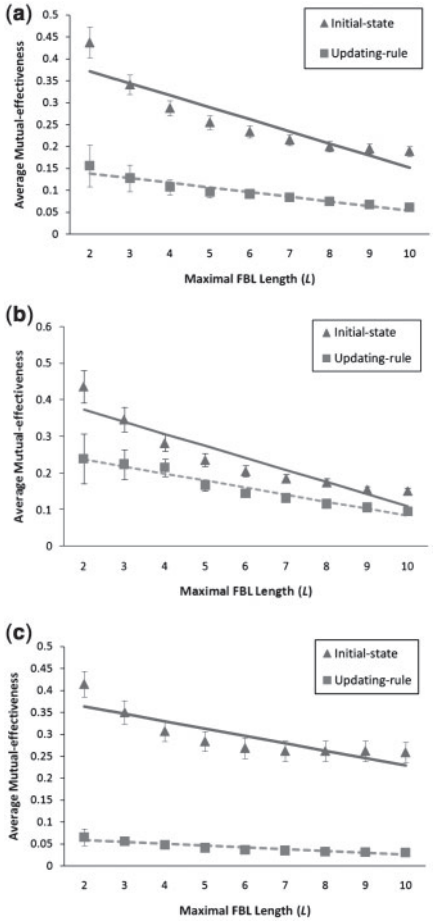

Fig. 4. The relationship between average mutual-effectiveness of functional important nodes and maximal FBL length in random Boolean networks with $|V| = 50$ and $|A| = 75$. All the *y*-axis values represent the average mutual-effectiveness of with 95% confidence level. Blue solid line and green dashed line represent a linear regression of the average mutual-effectiveness values against an initial-state and an updating-rule perturbation, respectively. (**a**) Results of random Boolean networks generated by Model-A (slopes of solid and dashed lines are −0.02746 and −0.01063, respectively) (**b**) Results of random Boolean networks generated by Model-B (slopes of solid and dashed lines are −0.032934 and −0.019254, respectively). (**c**) Results of random Boolean networks generated by Model-C (slopes of solid and dashed lines are −0.01666 and −0.00412, respectively).

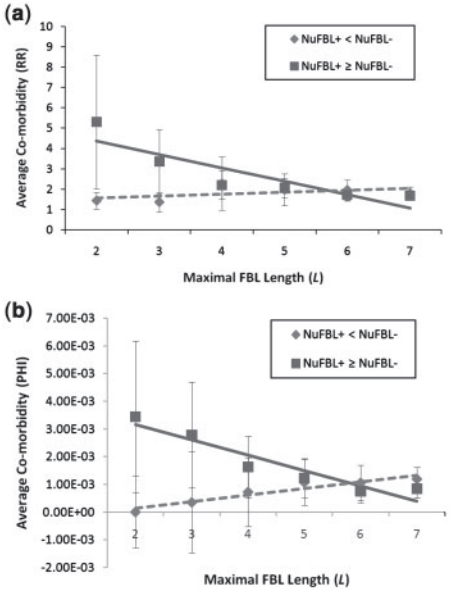

Fig. 5. Comparisons of the relationship between average comorbidity and maximal FBL length, according to the majority sign of the involved FBLs in the human signaling network. Each disease pair is classified into two categories: $NuFBL_+ \geq NuFBL_-$ if the number of involved positive FBLs is larger than or equal to that of involved negative FBLs, and $NuFBL_+ < NuFBL_-$ if otherwise. All the *y*-axis values represent the average comorbidity values with a 95% confidence level. Solid and dashed lines represent a linear regression of '$NuFBL_+ \geq NuFBL$' and '$NuFBL_+ < NuFBL_-$' categories, respectively. (**a**) Result of RR comorbidity (slopes of solid and dashed lines are −0.66119 and 0.09495, respectively.) (**b**) Result of PHI comorbidity (slopes of solid and dashed lines are −0.00055 and 0.00024, respectively).

points since a number of other nodes are also involved in that transference. Moreover, other topological properties such as the length of the shortest path and the number of FBLs between a pair of nodes did not show any obvious relationship to mutual-effectiveness values as in the case of comorbidity in Section 3.2 (see Figure S4c and d in Supplementary Material).

3.4 The effect of sign of FBL on comorbidity and mutual-effectiveness

It was also reported that the dynamic behavior of networks depends on the sign of FBLs. In terms of converging dynamics, networks with a relatively large number of positive FBLs are more likely to induce fixed-point attractors; on the other hand, networks with a relatively large number of negative FBLs are more likely to induce limit-cycle attractors (Kwon and Cho, 2007). Therefore, we further examined the effect of the sign of FBLs on comorbidity in the signaling network (Fig. 5) and mutual-effectiveness in random Boolean networks (Fig. 6). When a maximal FBL length L was specified, we collected a set of disease pairs, D and D', which are connected with a FBL of length $\leq L$ and classified them into two categories, '$NuFBL_+ \geq NuFBL_-$' and '$NuFBL_+ < NuFBL_-$', according to difference of $NuFBL_+(D, D')$ and $NuFBL_-(D, D')$. We compared comorbidity between those two categories in the human signaling network (Fig. 5). For both RR and PHI, we observed that disease pairs belonging to the '$NuFBL_+ \geq NuFBL_-$' group showed a clear negative relationship between average comorbidity and maximal FBL length compared to those belonging to the '$NuFBL_+ < NuFBL_-$' group (P-value = 0.00229 in Fig. 5a and P-value = 2.277×10^{-5} in Fig. 5b). The slope difference between the two groups was larger in the case of PHI than RR. In a similar way, we examined mutual-effectiveness in random Boolean networks generated by three models, Model-A (Fig. 6a), Model-B (Fig. 6b) and Model-C (Fig. 6c). For all models, we also observed that pairs of nodes belonging to the '$NuFBL_+ \geq NuFBL_-$' group showed a more outstanding negative relationship between average of mutual-effective and maximal FBL length than those belonging to '$NuFBL_+ < NuFBL_-$' (For cases of initial-state perturbations, P-values are 0.00535, 0.02164 and 0.01898 in Fig. 6a, b and c,

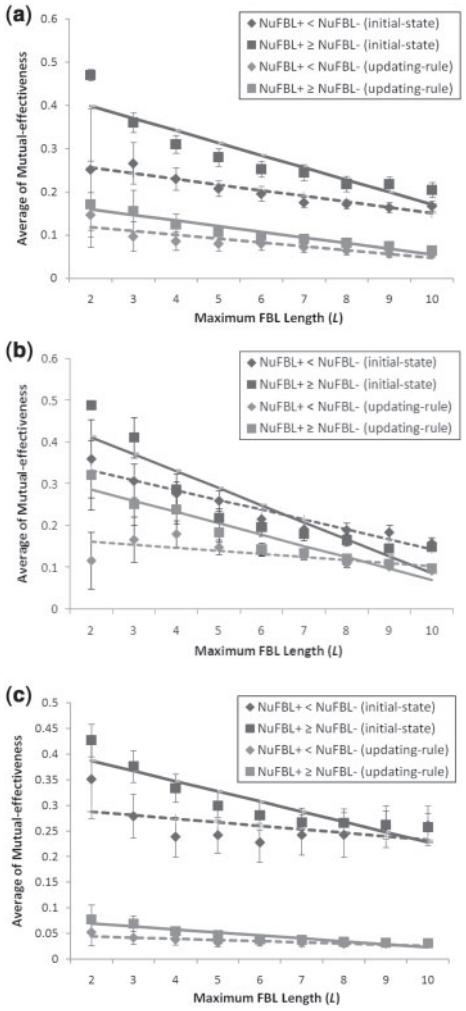

Fig. 6. Comparisons of the relationship between average mutual-effectiveness of functional important nodes and maximal FBL length according to the majority sign of the involved FBLs in random Boolean networks with $|V| = 50$ and $|A| = 75$. Each pair of nodes is classified into two categories: $NuFBL_+ \geq NuFBL_-$ if the number of involved positive FBLs is larger than or equal to that of involved negative FBLs, and $NuFBL_+ < NuFBL_-$ if otherwise. Two types of perturbations, initial-state perturbation (blue) and updating-rule perturbation (green), were considered. All the *y*-axis values represent the average mutual-effectiveness values with a 95% confidence level. Solid and dashed lines represent a linear regression of average mutual-effectiveness values of '$NuFBL_+ \geq NuFBL_-$' and '$NuFBL_+ < NuFBL_-$' categories, respectively. (**a**) Results of random Boolean networks generated by Model-A (slopes of blue solid, blue dashed, green solid and green dashed lines are −0.02846, −0.01325, −0.01295 and −0.00888, respectively) (**b**) Results of random Boolean networks generated by Model-B (slopes of blue solid, blue dashed, green solid and green dashed lines are −0.04067, −0.02371, −0.02687 and −0.00753, respectively) (**c**) Results of random Boolean networks generated by Model-C (slopes of blue solid, blue dashed, green solid and green dashed lines are −0.01978, −0.00692, −0.00581 and −0.00233, respectively.)

respectively; For cases of updating-rule perturbations, *P*-values are 0.04647, 0.00028 and 0.00151 in Fig. 6a, b and c, respectively). As shown in Figure 6, this observation was consistent irrespective of the type of perturbation and network generation model. Also,

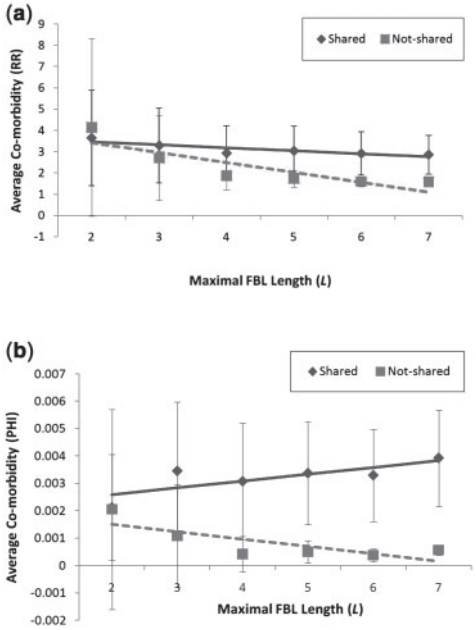

Fig. 7. Comparisons of the relationship between average comorbidity and maximal FBL length according to whether the shared genes exist or not. All pairs of diseases are classified into two groups: 'Shared'(blue solid line) or 'Not-shared'(green dashed line), which represent a set of disease pairs that share at least one gene or no genes, respectively. (**a**) Result of RR comorbidity (slopes of blue and green line are −0.14136 and −0.46398, respectively). (**b**) Result of PHI comorbidity (slopes of blue and green line are 0.00025 and −0.00027, respectively). All *y*-axis values represent the average comorbidity with a 95% confidence level.

it was independent of network size and density (See Figure S6 in Supplementary Material). Taken together, the negative relationship between comorbidity/mutual-effectiveness and the length of the involved feedback is more apparent in the case in which the number of involved positive FBLs is larger than that of the involved negative FBLs. This may be because positive FBLs are mainly related to amplifying signals, while negative FBLs play a role in inhibiting signals (Claire, 2004; Mendoza *et al.*, 1999).

3.5 The effect of FBLs on comorbidity when there is no common disease gene

A previous study showed that a disease pair becomes more comorbid as they share a larger number of disease genes (Park *et al.*, 2009). Inspired by the result, we further investigated the relationship between comorbidity and FBL length for the group of disease pairs that do not share any disease genes (Fig. 7). We classified every disease pair into two groups: 'Shared' (set of disease pairs having at least one shared gene) or 'Not-shared' (set of disease pairs having no shared genes). We compared the relationship between comorbidity and maximal length of FBLs between the two groups. We first observed that the average comorbidity of the 'Shared'group was larger than that of the 'Not-shared' group (*P*-value = 0.04746 in Fig. 7a and *P*-value = 3.046×10^{-5} in Fig. 7b). This means that the number of shared genes is an important indicator for comorbidity, as shown in the previous study (Park *et al.*, 2009). In addition, we observed that there is a negative relationship between comorbidity

Table 1. Examples of five disease pairs showing high comorbidity

Disease 1 (D_1)	Disease 2 (D_2)	RR	PHI	The number of FBLs between D_1 and D_2 of length l					
				$l=2$	3	4	5	6	7
Autoimmune disease	Systemic lupus erythematosus susceptibility	41.633	0.0262657	1	1	2	12	119	841
Chondrosarcoma	Lymphoma	6.8114	0.0026211	1	4	7	9	52	412
Nasopharyngeal carcinoma	Lymphoma	5.7397	0.0021259	1	3	6	9	52	395
Adrenal cortical carcinoma	Lymphoma	4.0689	0.0021106	1	3	6	9	53	400
Adenocarcinoma	Lymphoma	3.6934	0.0075878	1	3	6	16	88	621

Each pair of diseases is connected with FBLs of short lengths but have no shared gene in the human signaling network.

and length of FBLs in the 'Not-shared' group (P-value = 0.01352 in Fig. 7a and P-value = 0.03353 in Fig. 7b), while the relationship is ambiguous in the 'Shared' group (the relationship is even positive in the case of PHI). This implies that a pair of diseases having no shared disease gene can be highly comorbid if they are connected with FBLs of relatively short length. Considering most of pairs of diseases (18 404 disease pairs out of the total 18 896 disease pairs) have no shared genes, the FBL is also necessary to understand comorbidity. Table 1 presents examples of disease pairs that do not have any shared genes but still show high comorbidity. We find that there exist FBLs of relatively short lengths connecting them. Figure 8 shows a disease pair, adenocarcinoma and lymphoma, having very high comorbidity values (RR = 3.6934, PHI = 0.0075878), but sharing no disease gene. This pair of diseases has been often reported to occur simultaneously in same patients (Lee *et al.*, 2005; Neil *et al.*, 2009; Nishigami *et al.*, 2010; Nishino *et al.*, 1996).

4 DISCUSSION

High comorbidity between disease pairs, such as diabetes mellitus and obesity, or hypertension and spasms, have been observed. Previous studies explained that such a phenomenon was mostly due to mutational defects of a common disease gene shared by those diseases. However, many pairs of other diseases that do not share any associated gene have also been found, and this may be because complex molecular interaction networks spread the disorder of a gene to the other genes. Therefore, careful analysis of signaling pathways at the system level is needed to understand the comorbidity mechanism. In particular, we considered FBLs as an important structural feature for comorbidity, since many previous studies have revealed that they can make the network dynamics complicated and nonlinear.

In this regard, we investigated the effect of FBLs in the human cell signaling network on comorbidity, and showed that disease pairs connected with a FBL of shorter length make the comorbidity higher than those connected with an FBL of longer length. In other words, a pair of diseases is more likely to be comorbid when their associated genes are involved in FBLs of relatively short lengths. Moreover, we showed that such a relationship is apparent when the number of positive involved FBLs is larger than that of negative involved FBLs.

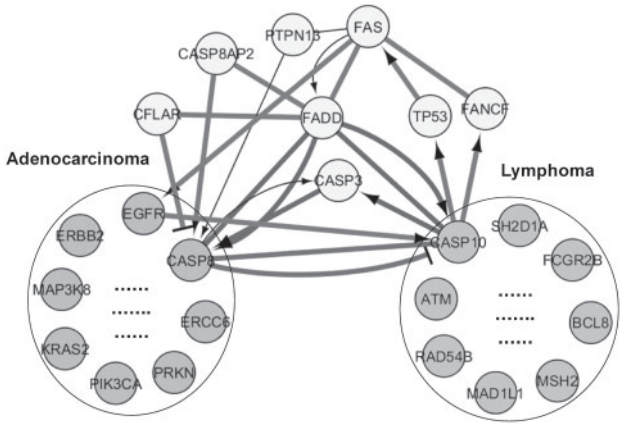

Fig. 8. A partial human signaling network with respect to adenocarcinoma and lymphoma which have 12 and 16 associated genes, respectively. RR and PHI values of this disease pair are 3.6934 and 0.0075878, respectively. We note that there is no shared gene between these two diseases but there exist FBLs of relatively short length. For example, a FBL of length 2 (CASP10 — CASP8 –| CASP10) and length 3 (CASP10 → CASP3 → CASP8 –| CASP10) are shown, where →, –|, and — denote activation, inhibitory, and either neutral or unknown interaction, respectively.

In addition, it was also shown that a pair of diseases connected with FBLs of relatively short length can be highly comorbid especially when they do not share any gene. Through intensive simulations based on Boolean networks, we proved the relationship between FBLs and comorbidity is a fundamental property with respect to network dynamics.

In synthetic biology, the findings of this article can be useful in order to control mutual-effectiveness between components of an artificial cell. Also, the findings of this study can be considered evidence showing the usefulness of a Boolean network model in studying the dynamics of biological networks. In addition, it will be a significant future study to find out a more realistic measure than mutual-effectiveness to simulate comorbidity phenomenon in a Boolean network model.

Funding: This work was supported by National Research Foundation of Korea Grant funded by the Korean Government (KRF-2008-313-D00995).

Conflict of Interest: none declared.

REFERENCES

Albert,R. (2004) Boolean modeling of genetic regulatory networks. *Lect. Notes Phys.*, **650**, 459–481.

Awan,A. *et al.* (2007) Regulatory network motifs and hotspots of cancer genes in a mammalian cellular signalling network. *IET Syst. Biol.*, **1**, 292–297.

Barabasi,A.L. and Albert,R. (1999) Emergence of scaling in random networks. *Science*, **286**, 509–512.

Bhalla,U.S. *et al.* (2002) MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science*, **297**, 1018–1023.

Calvano,S.E. *et al.* (2005) A network-based analysis of systemic inflammation in humans. *Nature*, **437**, 1032–1037.

Caughey,G. *et al.* (2008) Prevalence of comorbidity of chronic diseases in Australia. *BMC Public Health*, **8**, 221.

Claire,M.-E. (2004) Kinetic logic: a tool for describing the dynamics of infectious disease behavior. *J. Cell. Mol. Med.*, **8**, 269–281.

Cui,Q. *et al.* (2007) A map of human cancer signaling. *Mol. Syst. Biol.*, **3**, 152.

Faure,A. *et al.* (2006) Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics*, **22**, e124–e131.
Ferrell,J.E. Jr and Machleder,E.M. (1998) The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. *Science*, **280**, 895–898.
Gabriel,S. and Michaud, K. (2009) Epidemiological studies in incidence, prevalence, mortality, and comorbidity of the rheumatic diseases. *Arthritis Res. Ther.*, **11**, 229.
Goehler,H. *et al.* (2004) A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. *Mol. Cell*, **15**, 853–865.
Goh,K.-I. *et al.* (2007) The human disease network. *Proc. Natl Acad. Sci. USA*, **104**, 8685–8690.
Gupta,S. *et al.* (2007) Boolean network analysis of a neurotransmitter signaling pathway. *J. Theor. Biol.*, **244**, 463–469.
Helikar,T.Å. *et al.* (2008) Emergent decision-making in biological signal transduction networks. *Proc. Natl Acad. Sci. USA*, **105**, 1913–1918.
Hidalgo,C.S.A. *et al.* (2009) A dynamic network approach for the study of human phenotypes. *PLoS Comput. Biol.*, **5**, e1000353.
Huang,S. and Ingber,D.E. (2000) Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. *Exp. Cell Res.*, **261**, 91–103.
Kauffman,S. *et al.* (2003) Random Boolean network models and the yeast transcriptional network. *Proc. Natl Acad. Sci.USA*, **100**, 14796–14799.
Kauffman,S. *et al.* (2004) Genetic networks with canalyzing Boolean rules are always stable. *Proc. Natl Acad. Sci.USA*, **101**, 17102–17107.
Kelli,L.D. *et al.* (2005) Comparison of three comorbidity measures for predicting health service use in patients with osteoarthritis. *Arthritis Care Res.*, **53**, 666–672.
Kwon,Y.-K. and Cho,K.-H. (2007) Boolean dynamics of biological networks with multiple coupled feedback loops. *Biophys. J.*, **92**, 2975–2981.
Kwon,Y.-K. and Cho,K.-H. (2008) Quantitative analysis of robustness and fragility in biological networks based on feedback dynamics. *Bioinformatics*, **24**, 987–994.
Kwon,Y.-K. *et al.* (2007) Investigations into the relationship between feedback loops and functional importance of a signal transduction network based on Boolean network modeling. *BMC Bioinformatics*, **8**, 384.
Lee,D.S. *et al.* (2008) The implications of human metabolic network topology for disease comorbidity. *Proc. Natl Acad. Sci. USA*, **105**, 9880–9885.
Lee,S.-Y. *et al.* (2005) Synchronous adenocarcinoma and mucosa-associated lymphoid tissue (MALT) lymphoma in a single stomach. *Jpn J. Clin. Oncol.*, **35**, 591–594.
Lim,J. *et al.* (2006) A proteinprotein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell*, **125**, 801–814.
Ma'ayan,A. *et al.* (2005) Formation of regulatory patterns during signal propagation in a mammalian cellular network. *Science*, **309**, 1078–1083.
Mai,Z. and Liu,H. (2009) Boolean network-based analysis of the apoptosis network: irreversible apoptosis and stable surviving. *J. Theor. Biol.*, **259**, 760–769.
Mendoza,L. *et al.* (1999) Genetic control of flower morphogenesis in Arabidopsis thaliana: a logical analysis. *Bioinformatics*, **15**, 593–606.
Milo,R. *et al.* (2002) Network motifs: simple building blocks of complex networks. *Science*, **298**, 824–827.
Neil,S. *et al.* (2009) Synchronous adenocarcinoma and marginal zone B-cell lymphoma of the colon: a case report. **15**, 318–322.
Nishigami,T. *et al.* (2010) Concomitant adenocarcinoma and colonic non-Hodgkin's lymphoma in a patient with ulcerative colitis: a case report and molecular analysis. *Pathol.– Res. Pract.*, **206**, 846–850.
Nishino,N. *et al.* (1996) Synchronous lymphoma and adenocarcinoma occurring as a collision tumor in the stomach: report of a case. *Surg. Today*, **26**, 508–512.
Oldham,M.C. *et al.* (2006) Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc. Natl Acad. Sci. USA*, **103**, 17973–17978.
Park,J. *et al.* (2009) The impact of cellular networks on disease comorbidity. *Mol. Syst. Biol.*, **5**.
Pomerening,J.R. *et al.* (2003) Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nat. Cell Biol.*, **5**, 346–351.
Prill,R.J. *et al.* (2005) Dynamic properties of network motifs contribute to biological network organization. *PLoS Biol.*, **3**, e343.
Pujana,M.A. *et al.* (2007) Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat. Genet.*, **39**, 1338–1349.
Raja,M. and Azzoni, A. (2008) Comorbidity of Asperger's syndrome and bipolar disorder. *Clin. Pract. Epidemiol. Ment. Health*, **4**, 26.
Rual,J.-F. *et al.* (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, **437**, 1173–1178.
Saadatpour,A. *et al.* (2010) Attractor analysis of asynchronous Boolean models of signal transduction networks. *J. Theor. Biol.*, **266**, 641–656.
Saez-Rodriguez,J. *et al.* (2007) A logical model provides insights into T cell receptor signaling. *PLoS Comput. Biol.*, **3**, e163.
Sahin,O. *et al.* (2009) Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. *BMC Syst. Biol.*, **3**, 1.
Schlatter,R. *et al.* (2009) ON/OFF and beyond - a Boolean model of apoptosis. *PLoS Comput. Biol.*, **5**, e1000595.
Shmulevich,I. *et al.* (2003) The role of certain post classes in Boolean network models of genetic networks. *Proc. Natl Acad. Sci. USA*, **100**, 10734–10739.
Shmulevich,I. *et al.* (2005) Eukaryotic cells are dynamically ordered or critical but not chaotic. *Proc. Natl Acad. Sci. USA*, **102**, 13439–13444.
Snoussi,E.H. (1998) Necessary conditions for multistationarity and stable periodicity. *J. Biol. Syst.*, **6**, 3–9.
Stelzl,U. *et al.* (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, **122**, 957–968.
Tang,J. *et al.* (2008) Performance of comorbidity measures to predict stroke and death in a community-dwelling, hypertensive medicaid population. *Stroke*, **39**, 1938–1944.
Tetsche,M. *et al.* (2008) The impact of comorbidity and stage on ovarian cancer mortality: a nationwide Danish cohort study. *BMC Cancer*, **8**, 31.
Ye,Y. *et al.* (2005) A genome-wide tree- and forest-based association analysis of comorbidity of alcoholism and smoking. *BMC Genet.*, **6**, S135.
Yeger-Lotem,E. *et al.* (2004) Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction. *Proc. Natl Acad. Sci. USA*, **101**, 5934–5939.