When two disorders or illnesses occur in the same person, 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 strength 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 be 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 (Caughley, 2005) and addressed statistical results of prevalence and mortality. Other studies attempted to quantify the effect of a disease on other diseases by introducing comorbidity measures (Kellis 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 they 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; Pajana 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
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 (PHE), for a total of 81,924 disease pairs. The two comorbidity measures were defined as RR = Cij/Cij and PHE = (NCij − L)· √ 2N(N − L), respectively, where N is the number of patients, Id denotes occurrence of disease i, Cij denotes the number of patients who were simultaneously diagnosed with diseases i and j, respectively and Cij = N/i, N. In that study, the Medicare database which includes the clinical history of 130,099,018 patients was used for comorbidity evaluation. When two diseases cooccur more frequently than expected by chance, we have RR > 1 and PHE > 0. Each measure was reported to carry evaluation. When two diseases cooccur more frequently than expected by chance, we have RR > 1 and PHE > 0. Each measure was reported to carry

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 (PHE), for a total of 81,924 disease pairs. The two comorbidity measures were defined as RR = Cij/Cij and PHE = (NCij − L)· √ 2N(N − L), respectively, where N is the number of patients, Id denotes occurrence of disease i, Cij denotes the number of patients who were simultaneously diagnosed with diseases i and j, respectively and Cij = N/i, N. In that study, the Medicare database which includes the clinical history of 130,099,018 patients was used for comorbidity evaluation. When two diseases cooccur more frequently than expected by chance, we have RR > 1 and PHE > 0. Each measure was reported to carry evaluation. When two diseases cooccur more frequently than expected by chance, we have RR > 1 and PHE > 0. Each measure was reported to carry

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 (vi, vj) is assigned with either a positive (‘activating’) or negative (‘inhbiting’) relationship from vi ∈ V to vj ∈ 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) ⊆ 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), v0 → v1 → v2 → ・・・ → v2−1 → v2 is a FBL of length L ≥ 2 if there are links from v2−1 to v1 for all i = 1, 2, ・・・, n with v1 = vj and v2 ≠ vj for j, j ∈ [0, 1, ・・・, 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 ∈ 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 ∈ V and v’ ∈ V, we call v and v connected with a FBL of a maximal length L if there exists at least one FBL of length L involved both with v and v’. In a similar way, when G(V, A) represents the human signaling network and a pair of diseases, D and D’, 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 ∈ V(D) and v’ ∈ 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+ or NuFBL-, 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., 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, 2012), a central intrinsic and extrinsic apoptosis pathway (Ma and Liu, 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 vi ∈ V has a value of 1 (‘on’) or 0 (‘off’), which represents the possible states of the corresponding elements. The value of each variable at time t+1 is determined by the values of k1 other variables v1,v2,...,vk1 with a link to vi at time t by the Boolean function fi ∈ {0, 1}k1 → {0, 1}. Hence, we can write the update rule as x(t + 1) = f(vi(v1(t), v2(t), ・・・, vi−1(t), vi+1(t), ・・・, vn(t))) where we randomly select either a logical conjunction or disjunction for all signed relationships in fi with a uniform probability distribution. For example, if a Boolean variable vi has a positive relationship from v1, a negative relationship from v2 and a positive relationship from v3, then the conjunction and disjunction update rules are x(t + 1) = v1(t) v2(t) v3(t) and x(t + 1) = v1(t) v2(t) v3(t), respectively. In the case of a conjunction, the value of vi is at time t+1 is 1 only if the values of v1, v2, and v3 are 1 at time t are 1, 0, and 1 respectively whereas, in the case of a disjunction, the value of vi is at time t+1 is 1 if at least one of the states of the clauses, v1(t), v2(t), and v3(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., 1999; Pomerening et al., 2003).

In addition, we define a transient sequence of values of a node \( v \) as follows: When a Boolean network \( G(V,A) \) is initialized with \( v_1(0), v_2(0), \ldots, v_N(0) \), the initial-state perturbation at a node \( v_1 \) (i.e., \( v_1(0) \) to \( 0 \)). On the other hand, the updating-rule perturbation at a node \( v_1 \) means switching the updating-rule at \( v_1 \) from a conjunctive function to a disjunctive 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 \( T_i \), the valid convergent time of \( v_i \), defined as \( T_i = \max(T_T, T_T') \) where \( T_T \) or \( T_T' \) represents 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 \( v_i \), \( v_i(0), T, v_i(T) \) and \( v_i(0), T', v_i(T') \), when \( v_i \) was subject to the perturbation or not, respectively.
- (iii) Then, we compute \( \mu(v_i,v_j) = d_{(0,T)}(v_i, v_i(T)) / T \) where \( d_{(0,T)} \) means the Hamming distance (i.e. the number of bits having different values) between two sequences. Thus, \( \mu(v_i,v_j) \) expresses how largely the trajectory with respect to \( v_j \) was affected by the perturbation at \( v_i \). Since \( \mu \) is not commutative, we derive the mutual-effectiveness for a pair of nodes \( v_i \) and \( v_j \), \( \rho(v_i,v_j) \), 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 \( v_5 \). To compute \( \rho(v_4,v_5) \), we get two transient sequences of \( v_4 \), \( v_4(0), T, v_4(T) \) and \( v_4(0), T', v_4(T') \), where \( v_4 \) was subject to a perturbation or not, respectively. In the same way, \( \mu(v_2,v_4) \) is computed and finally \( \rho(v_2,v_4) \) are obtained by averaging \( \mu(v_2,v_4) \) and \( \mu(v_4,v_2) \).

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 \( \geq 0.05 \).

### 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 $\approx -0.40655$) (b) Result of PHI comorbidity (slope of linear regression $\approx -0.0003$).

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 nodes.
are Boolean networks generated by Model-C (slopes of solid and dashed lines are random Boolean networks generated by Model-B (slopes of solid and dashed lines are random Boolean networks generated by Model-A (slopes of solid 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) Results of random Boolean networks generated by Model-B (slopes of solid and dashed lines are $-0.032934$ and $-0.019254$, respectively). (b) Results of random Boolean networks generated by Model-C (slopes of solid and dashed lines are $-0.01666$ and $-0.00412$, respectively).

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). We defined the ‘+’ group as the group with a FBL length of at least one and the ‘−’ group as the group with a FBL length of less than one. For both RR and PHI, we observed that disease pairs belonging to the ‘+’ group showed a more outstanding negative relationship between average mutual-effectiveness and maximal FBL length than those belonging to the ‘−’ group. 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 ‘+’ group showed a clear negative relationship between average comorbidity and maximal FBL length compared to those belonging to the ‘−’ group. 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 ‘+’ group showed a clear negative relationship between average mutual-effectiveness and maximal FBL length compared to those belonging to the ‘−’ group. For cases of initial-state perturbations, $P$-values are $0.00535$, $0.02164$ and $0.01898$ in Fig. 6a, b and c, respectively.)

Results of 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 ‘+’ group showed a clear negative relationship between average mutual-effectiveness and maximal FBL length compared to those belonging to the ‘−’ group. For cases of initial-state perturbations, $P$-values are $0.00535$, $0.02164$ and $0.01898$ in Fig. 6a, b and c, respectively.)

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). We defined the ‘+’ group as the group with a FBL length of at least one and the ‘−’ group as the group with a FBL length of less than one. For both RR and PHI, we observed that disease pairs belonging to the ‘+’ group showed a more outstanding negative relationship between average mutual-effectiveness and maximal FBL length than those belonging to the ‘−’ group. 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 ‘+’ group showed a clear negative relationship between average comorbidity and maximal FBL length compared to those belonging to the ‘−’ group. 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 ‘+’ group showed a clear negative relationship between average mutual-effectiveness and maximal FBL length compared to those belonging to the ‘−’ group. For cases of initial-state perturbations, $P$-values are $0.00535$, $0.02164$ and $0.01898$ in Fig. 6a, b and c, respectively.)
D.-H. Le and Y.-K. Kwon

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: $\text{NuFBL}_+ \geq \text{NuFBL}_-$ if the number of involved positive FBLs is larger than or equal to that of involved negative FBLs, and $\text{NuFBL}_+ < \text{NuFBL}_-$ 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 $\text{NuFBL}_+ \geq \text{NuFBL}_-$ and $\text{NuFBL}_+ < \text{NuFBL}_-$ 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 $-0.04607$, $-0.02371$, $-0.02687$ and $-0.00755$, 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). For cases of updating-rule perturbations, $P$-values are 0.04647, 0.00928 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, 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 = $3.046 \times 10^{-5}$ 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

<table>
<thead>
<tr>
<th>Disease 1 (D₁)</th>
<th>Disease 2 (D₂)</th>
<th>RR</th>
<th>PHI</th>
<th>The number of FBLs between D₁ and D₂ of length l</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autoimmune</td>
<td>Systemic lupus erythematosus</td>
<td>41.633</td>
<td>0.0262657</td>
<td>1 2 12 119 841</td>
</tr>
<tr>
<td>Chondrosarcoma</td>
<td>Lymphoma</td>
<td>6.8114</td>
<td>0.0026211</td>
<td>1 2 7 9 52 412</td>
</tr>
<tr>
<td>Nasopharyngeal carcinoma</td>
<td>Lymphoma</td>
<td>5.7397</td>
<td>0.0021259</td>
<td>1 3 6 9 52 395</td>
</tr>
<tr>
<td>Adrenal cortical carcinoma</td>
<td>Lymphoma</td>
<td>4.0689</td>
<td>0.0021106</td>
<td>1 3 6 9 53 400</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>Lymphoma</td>
<td>3.6934</td>
<td>0.0075878</td>
<td>1 3 6 16 88 621</td>
</tr>
</tbody>
</table>

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

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 mutual-effectiveness 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 pathways at the system level is needed to understand the comorbidity mechanism. In particular, we considered FBLs as an important

...


Yeh, Y. et al. (2005) A genome-wide tree- and forest-based association analysis of comorbidity of alcoholism and smoking. BMC Genet., 6, 5135.