BMC Systems Biology

CrossMark

Construction and analysis of gene-gene dynamics influence networks based on a Boolean model

Maulida Mazaya, Hung-Cuong Trinh and Yung-Keun Kwon[*]

Abstract

Background: Identification of novel gene-gene relations is a crucial issue to understand system-level biological phenomena. To this end, many methods based on a correlation analysis of gene expressions or structural analysis of molecular interaction networks have been proposed. They have a limitation in identifying more complicated gene-gene dynamical relations, though.

Results: To overcome this limitation, we proposed a measure to quantify a gene-gene dynamical influence (GDI) using a Boolean network model and constructed a GDI network to indicate existence of a dynamical influence for every ordered pair of genes. It represents how much a state trajectory of a target gene is changed by a knockout mutation subject to a source gene in a gene-gene molecular interaction (GMI) network. Through a topological comparison between GDI and GMI networks, we observed that the former network is denser than the latter network, which implies that there exist many gene pairs of dynamically influencing but molecularly non-interacting relations. In addition, a larger number of hub genes were generated in the GDI network. On the other hand, there was a correlation between these networks such that the degree value of a node was positively correlated to each other. We further investigated the relationships of the GDI value with structural properties and found that there are negative and positive correlations with the length of a shortest path and the number of paths, respectively. In addition, a GDI network could predict a set of genes whose steady-state expression is affected in *E. coli* gene-knockout experiments. More interestingly, we found that the drug-targets with side-effects have a larger number of outgoing links than the other genes in the GDI network, which implies that they are more likely to influence the dynamics of other genes. Finally, we found biological evidences showing that the gene pairs which are not molecularly interacting but dynamically influential can be considered for novel gene-gene relationships.

Conclusion: Taken together, construction and analysis of the GDI network can be a useful approach to identify novel gene-gene relationships in terms of the dynamical influence.

Keywords: Boolean dynamics, Gene-gene dynamics influence (GDI), Gene-gene molecular interaction (GMI) networks, Knockout mutation, Structural characteristics

---

* Correspondence: kwonyk@ulsan.ac.kr
Department of Electrical/Electronic and Computer Engineering, University of Ulsan, 93 Daehak-ro, Nam-gu, Ulsan 44610, Republic of Korea

Mazaya *et al. BMC Systems Biology* 2017, **11**(Suppl 7):133

Page 102 of 174

Background

Gene-gene relationships have been investigated for a long time in many previous studies [1–5]. In particular, most attention was focused on the functional properties of protein-protein interaction networks [6–8] or an epistasis which means that masking a particular allele prevents effects of another gene at a different locus [9, 10], and many methods based on statistical correlation analysis over gene expression datasets were developed to reveal a new epistasis [11–16]. However, they have a limitation in identifying more complicated gene-gene relationships because a state of a gene can be affected by many other genes along various signaling pathways [10]. In this regard, network-based approaches have been proposed [2, 11, 17, 18] and found more complicated forms of gene-gene relationships such as feedback and feedforward loops. These approaches often inferred false gene-gene relationships, though, because they were based on only the analysis of a network structure without considering network dynamics [19]. Accordingly, an investigation of novel gene-gene relationships in terms of the network dynamics was necessarily needed.

To this end, we proposed a method to quantify gene-gene dynamics influence (GDI) in this study. We computed how much a state trajectory of a gene is changed by a knockout mutation subject to another gene in a gene-gene molecular interaction (GMI) network using a Boolean network model [20–24]. By examining the GDI values of every ordered pair of genes, we can construct a GDI network where each directed edge indicates a positive dynamical influence from the source gene to the target gene of the edge. This notion can be regarded as an extension of previous studies about effects of genetic mutations on network dynamics. For example, it was shown that a single gene mutation can change a communication pattern between genes [25], which can lead to human diseases in gene regulatory networks [26, 27] or dysfunctional mechanism in T-cell survival signaling network [28, 29]. In our study, we analyzed properties of the GDI networks induced from real GMI networks. Through a comparison of the topologies between the GMI and the GDI networks, we found that the latter network was denser than the former, which implies that there exist a lot of gene pairs with a dynamically influencing but molecularly non-interacting relation. We further analyzed the degree distributions of large-scale GMI and GDI networks. We found that they are considerably different from each other because a lot of hub genes were generated in the latter network. Despite this difference with respect to connectivity, it was interesting to observe that the degree of a node in the GDI network is positively correlated to that in the GMI network. To deepen our understating about the structure of the GDI networks, we examined the relations of well-known structural properties

to the GDI value and found that the length of a shortest path and the number of paths of a gene pair have negative and positive correlations, respectively, to the GDI value whereas the number of feedback loop showed no significant relation. In addition, we observed that a GDI network can predict a high proportion of genes of which the steady-state expression was changed in *E. coli* gene-knockout experiments. More interestingly, we observed that the drug-targets with side-effects in the GDI network have a larger number of outgoing links and a smaller number incoming links than the rest of genes. This implies that the drug-targets with side-effects are more likely to influence the dynamics of other genes, but unlikely to be influenced by other genes. Finally, we found biological evidences supporting that the gene pairs which are not molecularly interacting but dynamically influential can be considered for novel gene-gene relationships which were not identified by traditional approaches yet. Taken together, construction of a GDI network can be a useful approach to explain various dynamical behavior induced by complex gene-gene relations in large-scale GMI networks.

Methods

Datasets

To investigate the GDI networks, we used three datasets about the GMI networks such as an *Arabidopsis* morphogenesis regulatory network (AMRN) with 10 nodes and 20 interactions [30], a *guard* cell abscisic acid signaling network (ABAN) with 44 nodes and 78 interactions [31], and a human signaling network (HSN) with 1609 nodes and 5063 interactions [32, 33] after removing self-loop interactions from the original datasets (see Additional file 1: Tables S1-S3). Moreover, we classified all genes in HSN into non-drug targets, and drug targets with and without side effects by using a drug target database of DrugBank [34] and a side-effect information database of SIDER [35, 36] (see Additional file 1: Table S4).

A Boolean network model

In this work, we employed a Boolean network model to compute the GDI value. A Boolean network is one of the simplest computational models to describe network dynamics [37, 38], and has been generally used to investigate complicated behaviors of GMI networks [31, 39–41], which is represented by a directed graph $G = (V, A)$ where $V = \{v_1, v_2, ..., v_N\}$ is a set of nodes and A is a set of ordered pairs of the nodes called directed links ($|V|$ and $|A|$ denote the number of nodes and links, respectively). A directed link $(v_i, v_j) \in A$ represents a positive (activating) or a negative (inhibiting) regulation from v_i to v_j. Every $v_i \in V$ has a state value with 1 (on) or 0 (off). The state of v_i at time $t + 1$ denoted by $v_i(t + 1)$ is established by the values of k_i other nodes $v_{i_1}, v_{i_2}, ..., v_{i_{k_i}}$ with a link to v_i at time t by a

Mazaya *et al. BMC Systems Biology* 2017, **11**(Suppl 7):133

Page 103 of 174

Boolean function $f_i : \{0,1\}^{k_i} \rightarrow \{0,1\}$ and the states of all nodes are synchronously updated. Here, we implemented a nested canalyzing function (NCF) model [20, 42] to describe an update rule as follows:

$$
f_i(v_{i1}(t), v_{i2}(t), ..., v_{ik_i}(t)) = \begin{cases} O_1 & \text{if } v_{i_1}(t) = I_1 \\ O_2 & \text{if } v_{i_1}(t) \neq I_1 \text{ and } v_{i_2}(t) = I_2 \\ O_3 & \text{if } v_{i_1}(t) \neq I_1 \text{ and } v_{i_2}(t) \neq I_2 \text{ and } v_{i_3}(t) = I_3 \\ \vdots \\ O_{k_i} & \text{if } v_{i_1}(t) \neq I_1 \text{ and} \cdots \text{and } v_{i_{k_i-1}}(t) \neq I_{k_i-1} \text{ and } v_{ik_i}(t) = I_{k_i} \\ O_{def} & \text{otherwise} \end{cases}
$$

where all I_m and O_m ($m = 1, 2, \cdots, k_i$) denote the canalyzing and canalyzed Boolean values, respectively, and O_{def} is set to $1-O_{k_i}$ in general. In this paper, each NCF is randomized by specifying every I_m and O_m between 0 and 1 uniformly at random. We note that many molecular interactions were successfully represented by NCFs [43–45].

A *network state* at time t can be denoted by an ordered list of state values of all nodes, $\mathbf{v}(t) = [v_1(t), v_2(t), ..., v_N(t)] \in \{0,1\}^N$. Every network state transits to another network state through a set of Boolean update functions $F = \{f_1, f_2, ..., f_N\}$. Hence, a network state trajectory starting from an initial network state eventually converges to either a fixed-point or a limit-cycle attractor. We define the attractor more rigorously as follows.

Definition
Let $\mathbf{v}(0), \mathbf{v}(1), \cdots$, be a network state trajectory starting at $\mathbf{v}(0)$. The *attractor* is defined as an ordered list of network states $\langle G, F, \mathbf{v}(0) \rangle = [\mathbf{v}(\tau), \mathbf{v}(\tau + 1), ..., \mathbf{v}(\tau + p - 1)]$ where τ is the smallest time step such that $\mathbf{v}(t) = \mathbf{v}(t + p)$ for $\forall t \geq \tau$ with $\mathbf{v}(i) \neq \mathbf{v}(j)$ for $\forall i \neq j \in \{\tau, \tau + 1, ..., \tau + p - 1\}$ (herein, p is called a length of the attractor). In addition, the state sequences of v_i in $\langle G, F, \mathbf{v}(0) \rangle$ is denoted by $\langle G, F, \mathbf{v}(0) \rangle_i = [v_i(\tau), v_i(\tau + 1), ..., v_i(\tau + p - 1)]$.

Examination of attractors is required to compute the gene-gene dynamics influence in a network. To implement this, we specified a set of initial states (S) and computed a state trajectory starting at every $\mathbf{v}(0) \in S$ until an attractor is found. In the case of AMRN with a small number of nodes ($|N| = 10$), we could consider all 2^N possible states for S. Unfortunately, this exhaustive examination is not feasible to analyze a huge network. Therefore, we generated 2000 and 4000 random initial states to construct S in the case of ABAN ($|N| = 44$) and HSN ($|N| = 1609$), respectively.

Construction of a GDI network
In this study, the dynamics influence of v_i on v_j for an ordered pair of genes (v_i, v_j) represents how much the states sequence of v_j is changed by a mutation subject to v_i in a Boolean network model. Specifically, we considered a knockout mutation [31, 46, 47] which describes a condition that the state of the mutated gene is frozen to 0 (off) state. This mutation can be implemented by changing F into F' which is defined as follows:

$$
F' = \begin{cases} \{f_1, ..., f_{i-1}, 0, f_{i+1}, ..., f_N\} & , \forall t \leq T \\ F & , \forall t > T \end{cases}
$$

where T is a parameter to denote the mutation duration time. In other words, the knockout mutation lasts for only $\forall t \leq T$, and the update-rule of v_i is restored to f_i right after time step T. It was also shown that the mutation duration parameter can significantly affect the mutation process in complex GMI networks [48–50]. In the following, we explain how to compute the dynamics influence value from v_i to v_j denoted by $\mu(v_i, v_j)$ in detail.

(1) Generate a set of random initial states S. For each initial state $\mathbf{v}(0) \in S$, obtain two attractors $\langle G, F, \mathbf{v}(0) \rangle$ and $\langle G, F', \mathbf{v}(0) \rangle$ in the wild-type and the v_i-mutant networks, respectively. For convenience, let $\langle G, F, \mathbf{v}(0) \rangle = [\mathbf{v}(\tau), \mathbf{v}(\tau + 1), ..., \mathbf{v}(\tau + p - 1)]$ and $\langle G, F', \mathbf{v}(0) \rangle = [\mathbf{v}'(\tau'), \mathbf{v}'(\tau' + 1), ..., \mathbf{v}'(\tau' + p - 1)]$.

(2) Compute a distance between $\langle G, F, \mathbf{v}(0) \rangle_j$ and $\langle G, F', \mathbf{v}(0) \rangle_j$ defined as follows

$$
d(\mathbf{v}(0), v_i, v_j) = \min_{m \in [0, d-1]} \frac{\sum_{l=0}^{c-1} I\left(v_j(\tau + l + m) \neq v_j'(\tau' + l)\right)}{c}
$$

where c and d are the least common multiple and the greatest common divisor, respectively, of p and p', and $I(condition)$ is a function which outputs 1 if $condition$ is true, and 0 otherwise. As a result, $d(\mathbf{v}(0), v_i, v_j)$ represents the minimum ratio of a bitwise difference between the states sequence of v_j in the wild-type and the v_i-mutant attractors over the least common period (c) of the two attractors.

(3) Lastly, compute the dynamics influence of v_i on v_j denoted by $\mu(v_i, v_j)$ by averaging out $d(\mathbf{v}(0), v_i, v_j)$ over all initial states in S as follows:

$$
\mu(v_i, v_j) = \frac{\sum_{\mathbf{v}(0) \in S} d(\mathbf{v}(0), v_i, v_j)}{|S|}
$$

Figure 1 shows an illustrative example to compute the GDI value in a network where node v_3 out of four nodes is subject to the knockout mutation. The set of update rules F is modified into F' where state value of v_3 is frozen to 0 (Fig. 1a) during a mutation duration time T, and we can obtain the wild-type and v_3-mutant attractors (Fig. 1b). To compute the dynamical influence on node v_1, we compute the minimum bitwise difference between the state sequence of v_1 in two attractors considering all possible alignments of the sequences (Fig. 1c), and

Mazaya *et al. BMC Systems Biology* 2017, **11**(Suppl 7):133
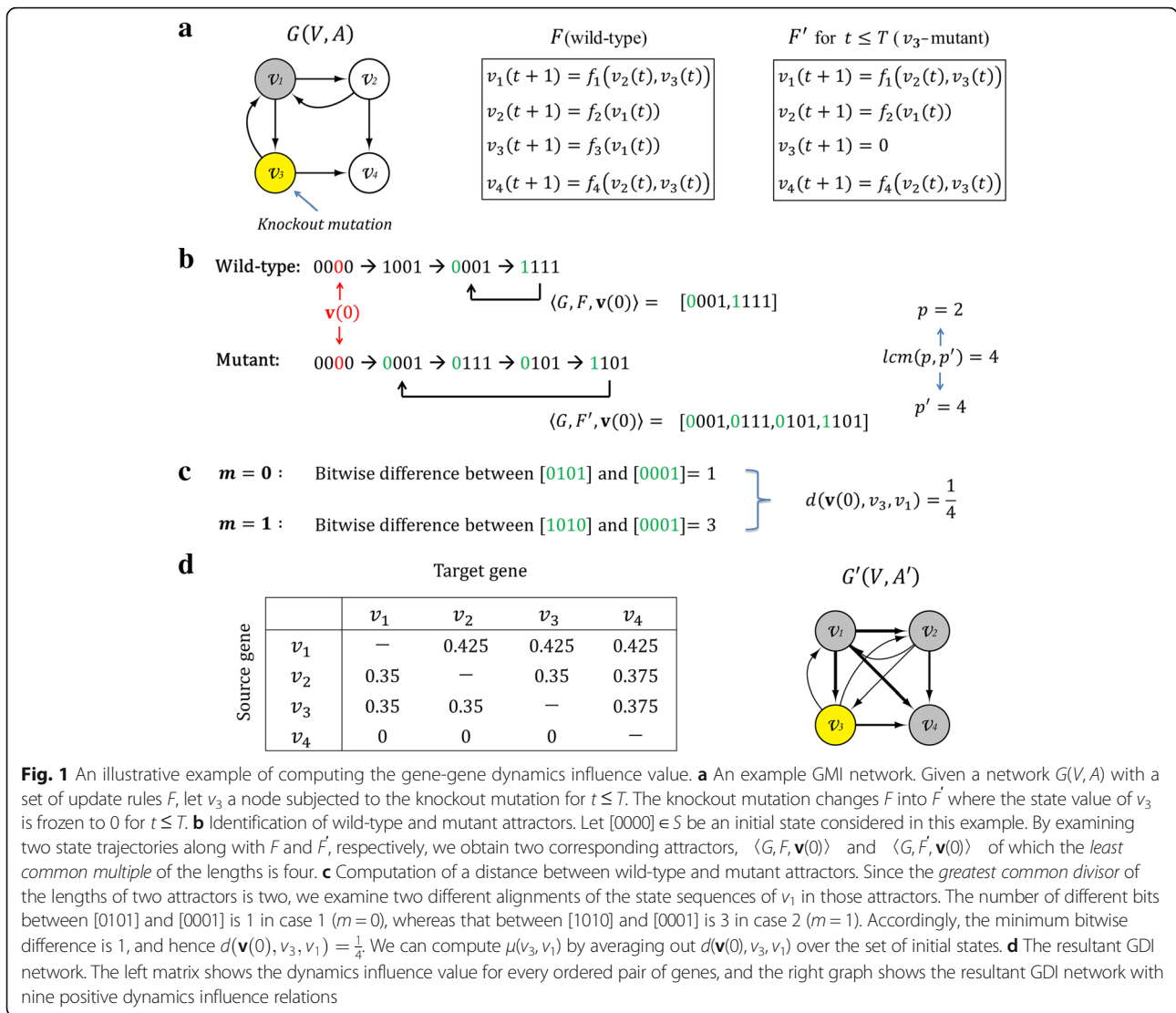
Page 104 of 174

Fig. 1 An illustrative example of computing the gene-gene dynamics influence value. **a** An example GMI network. Given a network $G(V, A)$ with a set of update rules F, let v_3 a node subjected to the knockout mutation for $t \leq T$. The knockout mutation changes F into F' where the state value of v_3 is frozen to 0 for $t \leq T$. **b** Identification of wild-type and mutant attractors. Let $[0000] \in S$ be an initial state considered in this example. By examining two state trajectories along with F and F', respectively, we obtain two corresponding attractors, $\langle G, F, \mathbf{v}(0) \rangle$ and $\langle G, F', \mathbf{v}(0) \rangle$ of which the *least common multiple* of the lengths is four. **c** Computation of a distance between wild-type and mutant attractors. Since the *greatest common divisor* of the lengths of two attractors is two, we examine two different alignments of the state sequences of v_1 in those attractors. The number of different bits between [0101] and [0001] is 1 in case 1 ($m = 0$), whereas that between [1010] and [0001] is 3 in case 2 ($m = 1$). Accordingly, the minimum bitwise difference is 1, and hence $d(\mathbf{v}(0), v_3, v_1) = \frac{1}{4}$. We can compute $\mu(v_3, v_1)$ by averaging out $d(\mathbf{v}(0), v_3, v_1)$ over the set of initial states. **d** The resultant GDI network. The left matrix shows the dynamics influence value for every ordered pair of genes, and the right graph shows the resultant GDI network with nine positive dynamics influence relations

eventually obtain the distance between $\langle G, F, \mathbf{v}(0) \rangle_1$ and $\langle G, F', \mathbf{v}(0) \rangle_1$. Then $\mu(v_3, v_1)$ is the average distance over the set of different initial states. Based on this measure, we can construct a GDI network $G'(V, A')$ from a GMI network $G(V, A)$ by calculating $\mu(v_i, v_j)$ for every ordered gene pair (v_i, v_j). More specifically, a GDI network is a directed graph where $(v_i, v_j) \in A'$ if and only if $\mu(v_i, v_j) > 0$. In other words, a directed edge in a GDI network means that the state sequences of the target node are changed by the knockout mutation subject to the source node of the edge. Figure 1d shows a matrix of $\mu(v_i, v_j)$ values for every ordered gene pair (v_i, v_j) and a resultant GDI network with nine positive influence relations.

Structural characteristics of networks

In real GMI networks, some structural characteristics of genes and interactions have been shown to be relevant to sustainability of network dynamics [51–53]. In this regard,

we employed the following well-known structural properties to investigate their relationships with the GDI value.

- The length of a shortest path for an ordered gene pair (v_i, v_j), denoted by $l(v_i, v_j)$, means the number of edges included in a shortest path from v_i to v_j [23, 40].
- The number of paths for an ordered gene pair (v_i, v_j), denoted by $n(v_i, v_j)$, means the number of different paths from v_i to v_j [40, 54].
- The number of feedback loops for an ordered gene pair (v_i, v_j), denoted by $f(v_i, v_j)$, means the number of feedback loops involving both v_i and v_j [22, 26]. A feedback loop is a circular chain of nodes where any node is not revisited except both end nodes. Specifically, $u_1 \rightarrow u_2 \rightarrow \ldots \rightarrow u_L$ is a feedback loop of length $L(>1)$ if there exists a link from u_i to u_{i+}

Mazaya *et al. BMC Systems Biology* 2017, **11**(Suppl 7):133

Page 105 of 174

$_1 (i \in \{1, \ldots, L-1\})$ such that $u_1 = u_L$ and $u_j \neq u_k$ for $\forall j \neq k \in \{1, \ldots, L-1\}$.

Construction of random networks

To verify that the results found in the real GMI networks hold in randomly structured networks, we generated random networks by using two models, the Barabási Albert (BA) model [55] and the shuffling model [41] (see Additional file 1: Figures S1 and S2, respectively, for the pseudo-codes), and analyzed their corresponding GDI networks. The BA model generates a random network using a preferential attachment scheme which is a network growth model. On the other hand, the shuffling model creates a random network by rewiring the links of a GMI network in a way that both in- and out-degrees of every node are preserved. Accordingly, the latter can generate a random network whose structure is more similar with the GMI network than the former.

Results

We generated the GDI networks from each of three real GMI networks, AMRN, ABAN, and HSN (see Methods). For convenience, we denote a GMI network and a corresponding GDI network by $G(V, A)$ and $G'(V, A')$, respectively.

Topological comparison between GMI and GDI networks

To investigate a topological difference between the GMI and the corresponding GDI networks, we first visualized them (Fig. 2 for the result of AMRN; see Additional file 1: Figures S3-S4 for the results of ABAN and HSN). For a further analysis, we classified every ordered pair of genes (v_i, v_j) into three groups as follows: a group of molecularly interacting and dynamically influential (MIDI) gene pairs (i.e., $(v_i, v_j) \in A$ and $(v_i, v_j) \in A'$), a group of molecularly

non-interacting but dynamically influential (MNDI) gene pairs (i.e., $(v_i, v_j) \notin A$ and $(v_i, v_j) \in A'$), and a group of molecularly interacting but dynamically non-influential (MIDN) gene pairs (i.e., $(v_i, v_j) \in A$ and $(v_i, v_j) \notin A'$). Table 1 shows the numbers of gene pairs belonging to MIDI, MNDI, and MIDN groups. We observed that the number of links of the GDI network was larger than that of the GMI network, which is because MNDI gene pairs are more frequently found than MIDN gene pairs (for example, we found 18 MNDI but no MIDN gene pairs in the case of AMRN). It is also interesting to observe a considerably large number of MIDN gene pairs in ABAN and HSN, because this implies that a molecularly interacting gene pair does not always induce a dynamically influencing relation. We note that the number of MIDN gene pairs in HSN was even larger than twice that of MIDI gene pairs.

We further compared the GMI and the GDI networks with respect to the degree distributions. Considering the network size, we investigated the case of HSN only (Fig. 3). We found that the degree of the GMI network considerably follows a power-law distribution whereas that of the GDI network does not (Fig. 3a). In particular, the hub nodes with a relatively high degree were more abundant in the GDI network than in the GMI network. Through additional comparisons with respect to the in-degree and the out-degree distributions (Fig. 3b and c, respectively), we found that the difference of the out-degree distribution was larger than that of in-degree distribution. All these results indicate that the overall topology of the GDI network is considerably different from that of the GMI network. We further wondered if a degree of a node in the GDI network is related or not to that in the GMI network. To answer this question, we compared the correlation coefficients
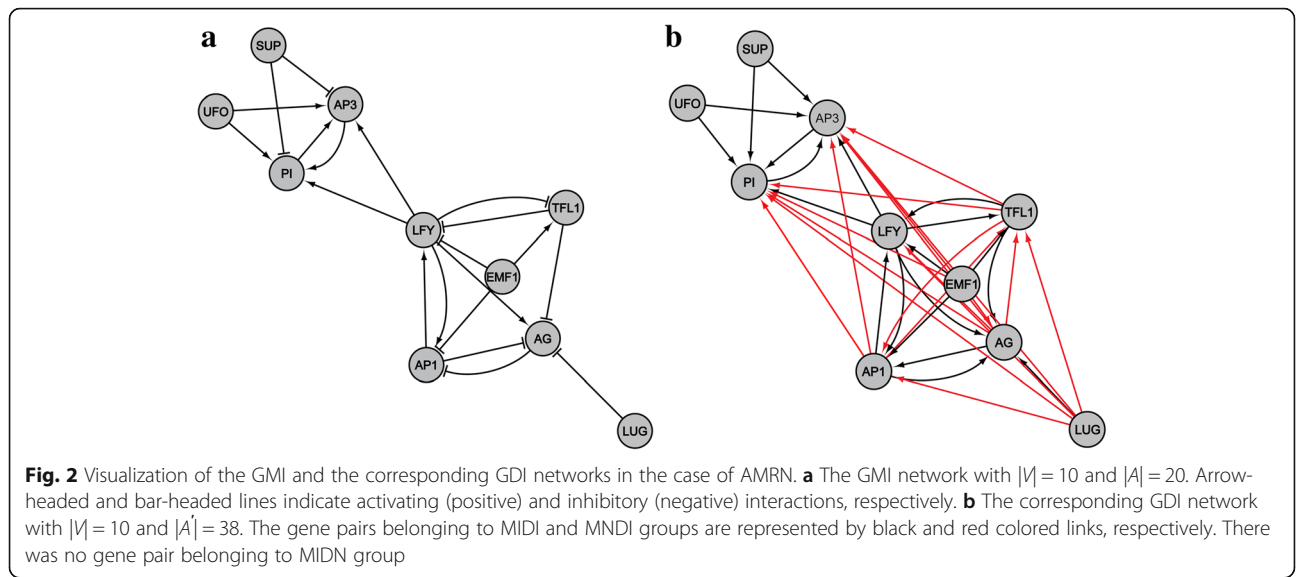

Fig. 2 Visualization of the GMI and the corresponding GDI networks in the case of AMRN. **a** The GMI network with $|V| = 10$ and $|A| = 20$. Arrow-headed and bar-headed lines indicate activating (positive) and inhibitory (negative) interactions, respectively. **b** The corresponding GDI network with $|V| = 10$ and $|A'| = 38$. The gene pairs belonging to MIDI and MNDI groups are represented by black and red colored links, respectively. There was no gene pair belonging to MIDN group

Mazaya *et al. BMC Systems Biology* 2017, **11**(Suppl 7):133

Page 106 of 174

Table 1 The number of gene pairs in groups classified by comparing the GMI and the GDI networks

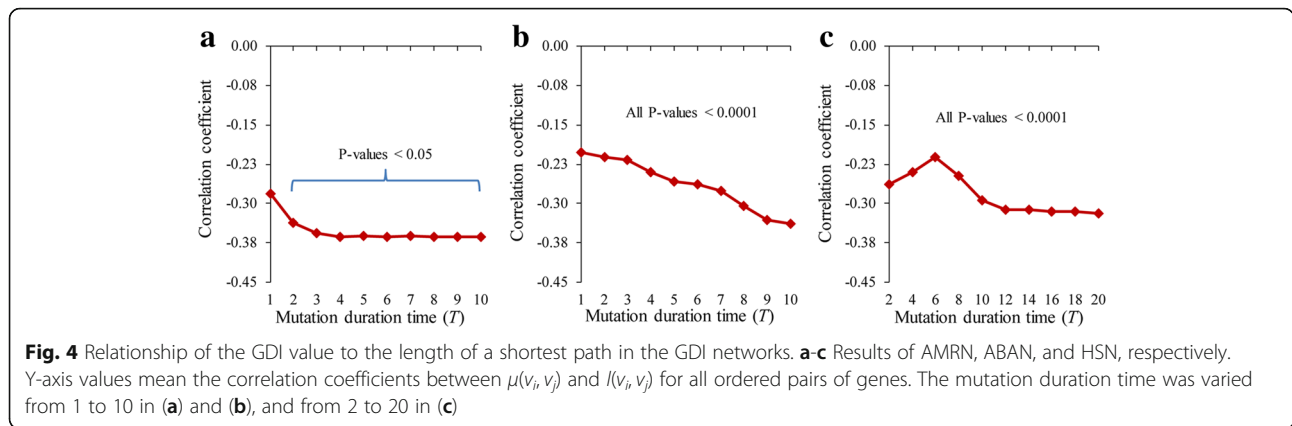|	AMRN	ABAN	HSN
The number of MIDI gene pairs (A)	20	63	1311
The number of MIDN gene pairs (B)	0	15	3752
The number of MNDI gene pairs (C)	18	603	19,910
The number of links in the molecular interaction network (A + B)	20	78	5063
The number of links in the GDI network (A + C)	38	666	21,221

between degree/in-degree/out-degree values of a node in the GMI and the GDI networks (Fig. 3d). As shown in the figure, each of them showed a significant positive correlation, irrespective of the mutation duration time. This means that the degree/in-degree/out-degree of a node in the GDI network is likely to be larger as that in the GMI network gets larger. Taken together, the topology of a GMI network can be partially helpful in predicting the topology of a GDI network although the latter is denser than the former.

Relation of dynamics influence values with structural characteristics in GDI networks

To discover a network-based principle about the GDI value, we investigated some structural properties in the

GDI networks. Here, we considered the relationships of the GDI value of a directed edge ($\mu(v_i, v_j)$) to three edge-based structural properties, the length of a shortest path from v_i to v_j ($l(v_i, v_j)$), the number of paths from v_i to v_j ($n(v_i, v_j)$), and the number of feedback-loops involving v_i and v_j ($f(v_i, v_j)$) in the GDI networks (see Methods for the definitions). We considered these three structural properties because they have been frequently used to show structural characteristics of functionally important genes or interactions in signaling networks [52, 53]. Figure 4 shows the correlation coefficients between $\mu(v_i, v_j)$ and $l(v_i, v_j)$ in the GDI networks, and they showed significant negative relations, irrespective of the mutation duration time. In other words, the dynamics influence of v_i on v_j is likely to be higher as the length of a shortest path from v_i to v_j is shorter. We infer that the information flow from the source gene to the target gene in the pair is less interfered by other genes when they are connected by a path of a short length. We note that this result is relevant to a previous study having shown that diseases whose associated genes are connected by a relatively short path tend to be comorbid [40]. In addition, the negative relation was more obvious as the duration time increases. To clarify that this finding is a general principle, we examined the correlation coefficients between $\mu(v_i, v_j)$ and $l(v_i, v_j)$ in two types of random networks generated by the shuffling and BA models (see Methods), and could observe the
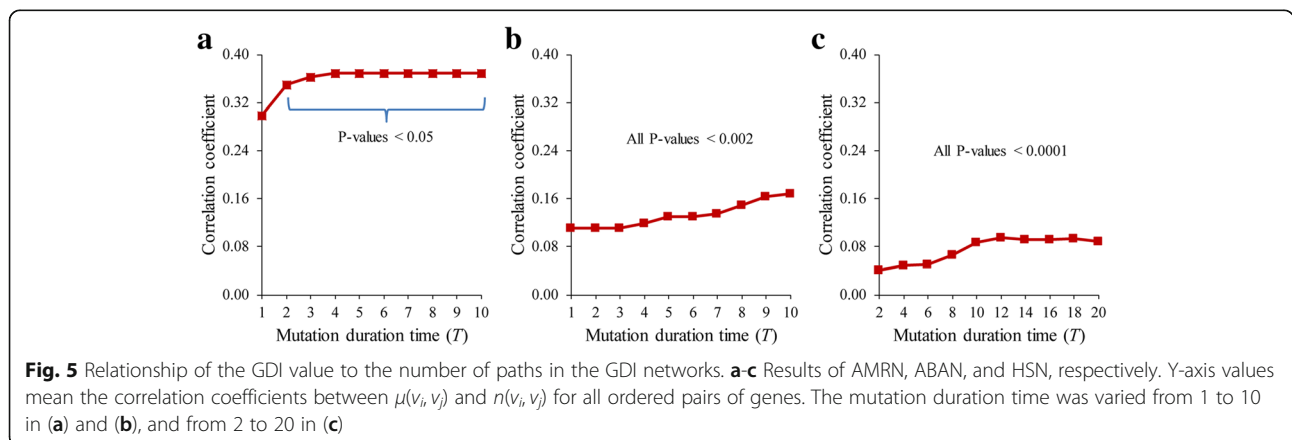

Fig. 3 Degree distributions of the GMI and the corresponding GDI networks in HSN. **a-c** Results of degree, in-degree, and out-degree distributions, respectively. The mutation duration time (*T*) was set to 20 in generating the GDI network. Y-axis represents a log-scaled frequency and blank triangle points (Δ) mean zero value of a frequency. We observed that the proportions of a same bin in GMI and GDI networks are significantly different from each other for most ranges (All *P*-values were less than 0.003 in (**a**) except for degree range '0–4', less than 0.0001 in (**b**) except for in-degree range '0–4', and were less than 0.002 in (**c**) except for out-degree range '0–4'. **d** Correlation coefficients between degree/in-degree/out-degree values of a node in the GMI and the GDI networks. The mutation duration time was varied from 2 to 20. Each of degree, in-degree, and out-degree of a node in the GDI network showed a significant positive relationship with that in the GMI network (All *p*-values <0.0001)

Mazaya *et al. BMC Systems Biology* 2017, **11**(Suppl 7):133

Page 107 of 174

Fig. 4 Relationship of the GDI value to the length of a shortest path in the GDI networks. **a-c** Results of AMRN, ABAN, and HSN, respectively. Y-axis values mean the correlation coefficients between $\mu(v_i, v_j)$ and $l(v_i, v_j)$ for all ordered pairs of genes. The mutation duration time was varied from 1 to 10 in (**a**) and (**b**), and from 2 to 20 in (**c**)

consistent results (see Additional file 1: Figure S5). To find another structural property, we examined the correlation coefficients between $\mu(v_i, v_j)$ and $n(v_i, v_j)$ (Fig. 5), and found significant positive relations, irrespective of the mutation duration time. This means that the dynamics influence of v_i on v_j tends to be higher as a larger number of paths connect from v_i to v_j. We infer that the information flow from the source gene to the target gene in the pair is more reinforced when they are connected by a larger number of paths. In addition, we examined the correlation coefficients between $\mu(v_i, v_j)$ and $n(v_i, v_j)$ in both the shuffled and the BA random networks, and could observe the consistent results (see Additional file 1: Figure S6). This implies that the positive relation between the number of paths and the dynamics influence can be a general property in various structural networks. Finally, we examined the relationship between $\mu(v_i, v_j)$ and the number of feedback loops involving the gene pair $f(v_i, v_j)$ (see Additional file 1: Figure S7) and found no consistently significant relationships in both real GMI networks and the random networks. Considering that the feedback loop structure was successfully used to predict functionally important genes or interactions [22, 26], our finding implies that the structural characteristics in the GDI networks can be different from those in the GMI networks.

Comparison of GDI network with knockout experiments

To validate our approach, we investigated how much a GDI network is consistent to real knockout experiments. To this end, we used an *E. coli* microarray dataset (E_coli_v4_Build_6 version) from the Many Microbe Microarrays database (M3D) [56] which contains the expression levels of 4297 genes from 446 samples. In addition, we also employed the RegulonDB database [57] which contains the information about the transcriptional regulations of *E. coli*. We integrated these two databases to identify the set of common genes, and then constructed a GMI network of *E. coli* with 1424 genes and 3114 edges. From the GMI network, we also generated the corresponding GDI network of *E. coli*, and denoted a set of out-going genes from a gene g by $GDI_O(g)$. To compare the GDI network result with the real knockout experiments, we identified 7 genes of which real knockout experimental results were included in both the GDI network and M3D database (There were 21 knockouts and 9 relevant wild-type experiments in M3D database). We converted real-valued expression to Boolean-valued

Fig. 5 Relationship of the GDI value to the number of paths in the GDI networks. **a-c** Results of AMRN, ABAN, and HSN, respectively. Y-axis values mean the correlation coefficients between $\mu(v_i, v_j)$ and $n(v_i, v_j)$ for all ordered pairs of genes. The mutation duration time was varied from 1 to 10 in (**a**) and (**b**), and from 2 to 20 in (**c**)

Mazaya *et al. BMC Systems Biology* 2017, **11**(Suppl 7):133

Page 108 of 174

one by using a discretization method based on K-means clustering algorithm [58]. It assigns 1 (the 'on' state) and 0 (the 'off' state) if the expression value is larger and lower, respectively, than the average expression of a gene. For each mutant gene g, we denote by $EXP_O(g)$ a set of genes of which Boolean expression values are differently observed between the knockout and the wild-type experiments. In other words, $GDI_O(g)$ and $EXP_O(g)$ represent a set of genes of which dynamics are affected by a knockout mutation at gene g through the GDI network analysis and the real experiments, respectively. To assess how much proportion of $EXP_O(g)$ is predicted by $GDI_O(g)$, we examined a precision ratio defined as follows:

$$ratio(g) = \frac{|EXP_O(g) \cap GDI_O(g)|}{|EXP_O(g)|}$$

As shown in Table 2, we found that the ratio ranges from 0.35 to 0.63 except for two genes, cspA and appY. This implies that the GDI network analysis can predict a relatively high portion of genes of which the expression was changed by the knockout experiment, although it did not explain all the experiments. This partially supports the validation of the GDI network-based analysis.

Analysis of drug-target genes based on GDI networks

Some previous approaches investigated drug-target genes through network structure analysis [59, 60]. For example, it was found that drug-target genes are more centrally located as well as more evolutionary than non-drug target genes [61]. It was also shown that the connectivity of drug-targets was significantly different from that of non-drug targets [62, 63]. Inspired by these results, we examined the structural characteristics of drug-targets in the GDI network. More specifically, we classified all genes in HSN into three groups of non-drug targets, drug-targets without side-effects, and drug-targets with side-effects (see Methods). We examined the average in- and out-degrees of three groups in the GDI network derived from the GMI network of HSN (Fig. 6). As shown in the figure, the average out-degree (in-degree) of the drug-targets with

side-effects was larger (resp., smaller) than those of non-drug targets and drug-targets without side-effects, almost irrespective of the mutation duration time. In other words, the drug-targets with side-effects are more likely to influence other genes, whereas they are less likely to be influenced by other genes. This result supports some experimental studies having shown that drug-targets with side-effects have a relatively larger impact on other genes than non-drug targets [60, 62, 63]. This case study supports the usefulness of GDI network analysis.

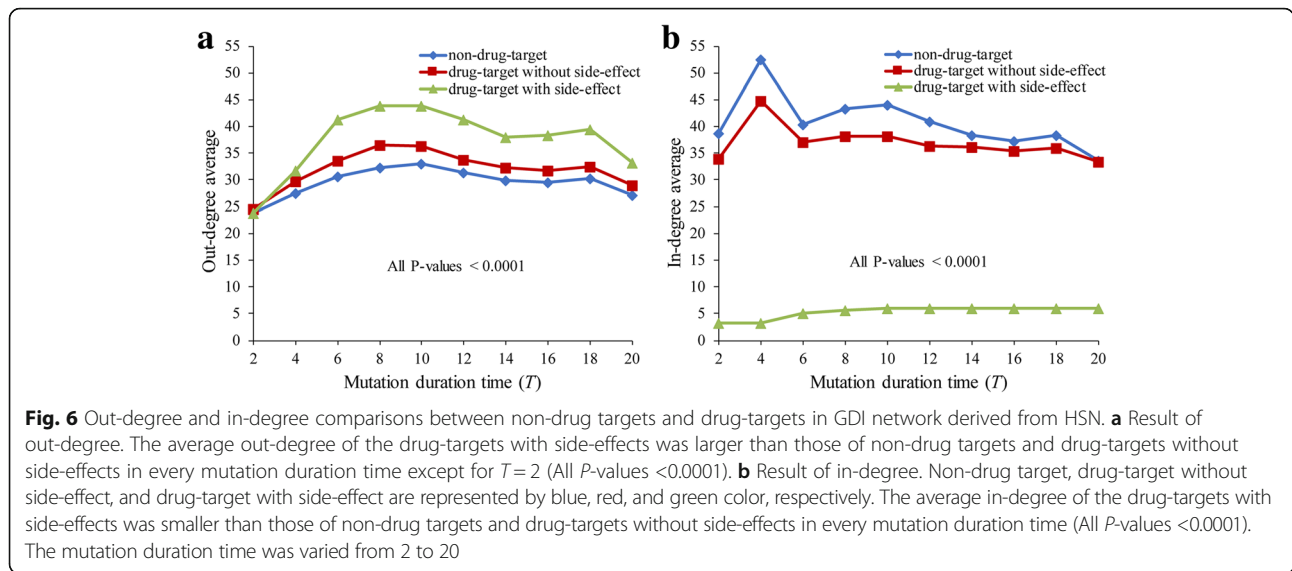### Biological evidence of novel gene-gene relations

To reveal novel gene-gene relations by means of the GDI network analysis, we profiled the gene pairs which are included in the GDI network but not included in the GMI network (i.e., gene pairs of MNDI group found in Table 1, Fig. 2, and see Additional file 1: Figures S3-S4), and some of them were listed in Table 3. Interestingly, we could find some biological evidences relevant to the gene pairs in the table. For example, the relation from EMF1 to AG found in AMRN can explain that EMF1 played an important role in maintaining AG development in *A. thaliana* [16, 64–66], and the relation from TFL1 to AP1 can explain that the addition of the TFL1 mutation induces the AP1 mutant which changes the *phyllotaxy* of lateral flowers [67]. In addition, it was reported in ABAN that ABA gene cooperates with S1P on slow anion channels [24, 68] or induces NO productions abolished in either NOS or NIA12 [31]. We also note previous studies having shown the CSK mutant can affect a regulatory polymorphism in B-cell signaling [69, 70] or the complement relationship of GHR and IGF1R [71]. These evidences imply that the GDI network-based analysis can reveal novel gene-gene relations which are not well-known yet.

Results and discussion

Gene-gene relationships have been investigated in many studies, most of which focused on epistasis and statistical correlation analysis. However, they have a limitation in identifying more complicated relationships and hence some network-based approaches have been proposed to

Table 2 Comparison between two sets of knockout-affected genes identified through a GDI network and *E-coli* knockout experiment, respectively

Knocked-out gene (g)	The number of genes in $EXP_O(g)$	The number of genes in $GDI_O(g)$	The number of genes in $EXP_O(g) \cap GDI_O(g)$	Ratio(g)
cspA	151	2	0	0.00
soxS	199	524	70	0.35
oxyR	235	539	97	0.41
appY	250	9	5	0.02
arcA	311	453	111	0.35
crp	342	891	217	0.63
fnr	502	616	263	0.52

Mazaya *et al. BMC Systems Biology* 2017, **11**(Suppl 7):133

Page 109 of 174

Fig. 6 Out-degree and in-degree comparisons between non-drug targets and drug-targets in GDI network derived from HSN. **a** Result of out-degree. The average out-degree of the drug-targets with side-effects was larger than those of non-drug targets and drug-targets without side-effects in every mutation duration time except for $T = 2$ (All *P*-values <0.0001). **b** Result of in-degree. Non-drug target, drug-target without side-effect, and drug-target with side-effect are represented by blue, red, and green color, respectively. The average in-degree of the drug-targets with side-effects was smaller than those of non-drug targets and drug-targets without side-effects in every mutation duration time (All *P*-values <0.0001). The mutation duration time was varied from 2 to 20

overcome it. It is a still open problem because they did not incorporate analysis about the dynamical relationships. In this regard, we first proposed a measure to quantify the gene-gene dynamics influence using a Boolean network model and eventually constructed a GDI network. To find characteristics of the GDI network, we compared the topologies of the GMI and the GDI networks and observed that the latter is denser than the former. This was because a lot of hub nodes were generated in the GDI network. In addition, the degree distributions were also different between them. Despite these topological differences, we found an interesting similarity such that the degree value of a node was positively correlated between the GMI and the GDI networks. For further investigations about the structure of the GDI networks, we examined the relations of three structural properties to the GDI value, and found that the length of a shortest path and the number of paths have negative and positive correlations, respectively, whereas the number of feedback loop showed no relation. In addition, we observed that a GDI network could

predict a set of genes whose steady-state expression is affected in *E. coli* gene-knockout experiments. It was more intriguing to observe that the drug-targets with side-effects are more likely to influence the dynamics of other genes, but less likely to be influenced by other genes through the GDI network-based analysis. We note that it is possible to reveal novel gene-gene relationships by considering gene pairs which are not molecularly interacting but dynamically influential. Taken together, the GDI network can be a useful method to explain various dynamical behavior caused by complex gene-gene relations in GMI networks. A future study will include a more generalized analysis considering various mutation types, an examination of novel structural characteristics in the GDI network and an investigation on the dynamical influence among three or more genes based on multiple mutations.

Additional file

Table 3 Example of gene pairs which are molecularly non-interacting but dynamically influential (MNDI)

Network	Gene pairs	Sub-graph	$\mu(v_i, v_j)$	$l(v_i, v_j)$
AMRN	EMF1, AG	EMF1→ TFL1⊣ AG	0.146	2
	TFL1, AP1	TFL1⊣ LFY→ AP1	0.0125	2
ABAN	ABA, S1P	ABA→ SphK→ S1P	0.498	2
	ABA, NIA12	ABA→ RCN1→ NIA12	0.498	2
HSN	CSK, BCR	CSK⊣ LYN→ BCR	0.503	2
	GHR, IGF1R	GHR→ IGF1→ IGF1R	0.382	2

Source and target genes are represented by orange- and green-colored nodes, respectively. Sub-graph field means the shortest path with respect to the gene pair in the GMI network. Arrow-headed (→) and bar-headed (⊣) lines indicate activating (positive) and inhibitory (negative) interactions, respectively

Mazaya *et al. BMC Systems Biology* 2017, **11**(Suppl 7):133

Page 110 of 174

networks with $|V| = 50$, $|A| = 100$. **Figure S7.** Relationship of the GDI value to the number of feedback loops involving the gene pair. Relationship of the GDI value to the number of feedback loops involving the gene pair. (**a–c**) Results of AMRN, ABAN, and HSN, respectively. (**d-f**) Results of the random networks shuffled from AMRN, ABAN, and HSN, respectively. **Table S1.** AMRN dataset consisting of 10 nodes and 20 interactions after removing self-loops from the original dataset. It includes the information about gene name as source and target, and interaction type. **Table S2.** ABAN dataset consisting of 44 nodes and 78 interactions after removing self-loops from the original dataset. It consists the information about gene name as source and target, and interaction type. **Table S3.** HSN dataset consisting of 1609 nodes and 5063 interactions after removing self-loops from the original dataset. It consists the information about gene name as source and target, and interaction type. **Table S4.** Information of drug-targets and side-effects for genes in HSN. It includes information about drug-targets and side-effects genes in HSN. (PDF 4490 kb)

Availability of data and materials
All data generated or analyzed during this study are included in this published article and its supplementary information files.

About this supplement
This article has been published as part of *BMC Systems Biology* Volume 11 Supplement 7, 2017: 16th International Conference on Bioinformatics (InCoB 2017): Systems Biology. The full contents of the supplement are available online at https://bmcsystbiol.biomedcentral.com/articles/supplements/volume-11-supplement-6.

Authors' contributions
YKK designed the study, MM and THC performed simulations, MM, THC, and YKK designed the analysis, MM, THC, and YKK wrote and revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published: 21 December 2017

References
1. Gilbert-Diamond D, Moore JH. Analysis of Gene-Gene Interactions. In: Current Protocols in Human Genetics. Volume 70. 3 edition. Wiley; 2011. pp. 1.14.11-11.14.12.
2. Baryshnikova A, Costanzo M, Myers CL, Andrews B, Boone C. Genetic Interaction Networks: Toward an Understanding of Heritability. Annu Rev Genomics Hum Genet. 2013;14:111–33.
3. Phillips PC. Epistasis - the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet. 2008;9:855–67.
4. Mallet J. Epistasis and the evolutionary process. Science. 2001;291:602–2.
5. Lin HY, Cheng CH, Chen DT, Chen YA, Park JY. Coexpression and expression quantitative trait loci analyses of the angiogenesis gene-gene interaction network in prostate cancer. Transl Cancer Res. 2016;5:S951.
6. QY W, Ye YM, Ho SS, Zhou SG. Semi-supervised multi-label collective classification ensemble for functional genomics. BMC Genomics. 2014;15:S17.
7. QY W, Wang ZY, Li CS, Ye YM, Li YP, Sun N. Protein functional properties prediction in sparsely-label PPI networks through regularized non-negative matrix factorization. BMC Syst Biol. 2015;9:S9.
8. QY W, Ye YM, Ng MK, Ho SS, Shi RC. Collective prediction of protein functions from protein-protein interaction networks. Bmc Bioinformatics. 2014;15:S9.
9. Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet. 2002;11:2463–8.
10. Azpeitia E, Benitez M, Padilla-Longoria P, Espinosa-Soto C, Alvarez-Buylla ER. Dynamic network-based epistasis analysis: Boolean examples. Front Plant Sci. 2011;2
11. Segre D, DeLuna A, Church GM, Kishony R. Modular epistasis in yeast metabolism. Nat Genet. 2005;37:77–83.
12. Olsen C, Fleming K, Prendergast N, Rubio R, Emmert-Streib F, Bontempi G, Haibe-Kains B, Quackenbush J. Inference and validation of predictive gene networks from biomedical literature and gene expression data. Genomics. 2014;103:329–36.
13. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet. 2009;10:392–404.
14. Snijder B, Liberali P, Frechin M, Stoeger T, Pelkmans L. Predicting functional gene interactions with the hierarchical interaction score. Nat Methods. 2013;10:1089–92.
15. Higa CH, Louzada VH, Andrade TP, Hashimoto RF. Constraint-based analysis of gene interactions using restricted boolean networks and time-series data. BMC Proc. 2011;5:S5.
16. Kim SY, Zhu T, Sung ZR. Epigenetic regulation of gene programs by EMF1 and EMF2 in Arabidopsis. Plant Physiol. 2010;152:516–28.
17. H-h J, Sohn K-A. Relvance epistasis network of gastritis for intra-chromosomes in the Korea associated resource (KARE) cohort study. Genomeics & Informatics. 2014;4:216–24.
18. Shervais S, Kramer PL, Westaway SK, Cox NJ, Zwick M. Reconstructability analysis as a tool for identifying gene-gene interactions in studies of human diseases. Stat Appl Genet Mol Biol. 2010;9(1):Article18.
19. Faure A, Naldi A, Chaouiya C, Thieffry D. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. Bioinformatics. 2006;22:E124–31.
20. Kauffman S, Peterson C, Samuelsson B, Troein C. Genetic networks with canalyzing Boolean rules are always stable. Proc Natl Acad Sci U S A. 2004;101:17102–7.
21. Kwon YK, Cho KH. Boolean dynamics of biological networks with multiple coupled feedback loops. Biophys J. 2007;92:2975–81.
22. Kwon YK, Cho KH. Analysis of feedback loops and robustness in network evolution based on Boolean models. Bmc Bioinformatics. 2007;8:430.
23. 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.
24. Chalfant CE, Spiegel S. Sphingosine 1-phosphate and ceramide 1-phosphate: expanding roles in cell signaling. J Cell Sci. 2005;118:4605–12.
25. Son SW, Kim DH, Ahn YY, Jeong H. Response network emerging from simple perturbation. J Korean Phys Soc. 2004;44:628–32.
26. Kwon YK, Cho KH. Quantitative analysis of robustness and fragility in biological networks based on feedback dynamics. Bioinformatics. 2008;24:987–94.
27. del Sol A, Balling R, Hood L, Galas D. Diseases as network perturbations. Curr Opin Biotechnol. 2010;21:566–71.
28. Xiao Y, Gong YH, Lv YL, Lan YJ, Hu J, Li F, JY X, Bai J, Deng YL, Liu L, et al. Gene perturbation atlas (GPA): a single-gene perturbation repository for characterizing functional mechanisms of coding and non-coding genes. Sci Rep. 2015;5:10889.
29. Cornelius SP, Kath WL, Motter AE. Realistic control of network dynamics. Nat Commun. 2013;4:1942.
30. Mendoza L, Thieffry D, Alvarez-Buylla ER. Genetic control of flower morphogenesis in Arabidopsis Thaliana: a logical analysis. Bioinformatics. 1999;15:593–606.
31. Li S, Assmann SM, Albert R. Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. PLoS Biol. 2006;4:1732–48.

Mazaya *et al. BMC Systems Biology* 2017, **11**(Suppl 7):133

Page 111 of 174

32. Cui QH, Purisima EO, Wang E. Protein evolution on a human signaling network. BMC Syst Biol. 2009;3:21.

33. Cui QH, Ma Y, Jaramillo M, Bari H, Awan A, Yang S, Zhang S, Liu LX, Lu M, O'Connor-McCourt M, et al. A map of human cancer signaling. Mol Syst Biol. 2007;3:152.

34. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, et al. DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs. Nucleic Acids Res. 2011;39:D1035–41.

35. Kuhn M, Al Banchaabouchi M, Campillos M, Jensen LJ, Gross C, Gavin AC, Bork P. Systematic identification of proteins that elicit drug side effects. Mol Syst Biol. 2013;9:663.

36. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. Nucleic Acids Res. 2016;44:D1075–9.

37. Kauffman SA. The origins of order : self-organization and selection in evolution: Oxford University Press; 1993.

38. Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol. 1969;22:437–67.

39. Helikar T, Konvalina J, Heidel J, Rogers JA. Emergent decision-making in biological signal transduction networks. Proc Natl Acad Sci U S A. 2008;105:1913–8.

40. Le DH, Kwon YK. The effects of feedback loops on disease comorbidity in human signaling networks. Bioinformatics. 2011;27:1113–20.

41. 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:e103010.

42. 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;100:14796–9.

43. Trinh HC, Kwon YK. Effective Boolean dynamics analysis to identify functionally important genes in large-scale signaling networks. Biosystems. 2015;137:64–72.

44. Trinh HC, Kwon YK. Edge-based sensitivity analysis of signaling networks by using Boolean dynamics. Bioinformatics. 2016;32:763–71.

45. Samal A, Jain S. The regulatory network of E. Coli metabolism as a Boolean dynamical system exhibits both homeostasis and flexibility of response. BMC Syst Biol. 2008;2:21.

46. Kwon YK, Kim J, Cho KH. Dynamical robustness against multiple mutations in signaling networks. IEEE/ACM Trans Comput Biol Bioinform. 2016;13:996–1002.

47. Campbell C, Albert R. Stabilization of perturbed Boolean network attractors through compensatory interactions. BMC Syst Biol. 2014;8:53.

48. Newman S, Howarth KD, Greenman CD, Bignell GR, Tavare S, Edwards PAW. The relative timing of mutations in a breast cancer genome. PLoS One. 2013;8:e64991.

49. Lecca P, Casiraghi N, Demichelis F. Defining order and timing of mutations during cancer progression: the TO-DAG probabilistic graphical model. Front Genet. 2015;6:309.

50. Bozic I, Nowak MA. Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers. Proc Natl Acad Sci U S A. 2014;111:15964–8.

51. Kaiser M, Hilgetag CC. Edge vulnerability in neural and metabolic networks. Biol Cybern. 2004;90:311–7.

52. Prill RJ, Iglesias PA, Levchenko A. Dynamic properties of network motifs contribute to biological network organization. PLoS Biol. 2005;3:1881–92.

53. Klein C, Marino A, Sagot MF, Milreu PV, Brilli A. Structural and dynamical analysis of biological networks. Briefings in Functional Genomics. 2012;11:420–33.

54. Girvan M, Newman MEJ. Community structure in social and biological networks. Proc Natl Acad Sci U S A. 2002;99:7821–6.

55. Barabasi AL, Albert R. Emergence of scaling in random networks. Science. 1999;286:509–12.

56. Faith JJ, Driscoll ME, Fusaro VA, Cosgrove EJ, Hayete B, Juhn FS, Schneider SJ, Gardner TS. Many microbe microarrays database: uniformly normalized Affymetrix compendia with structured experimental metadata. Nucleic Acids Res. 2008;36:D866–70.

57. Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muñiz-Rascado L, García-Sotelo JS, Weiss V, Solano-Lira H, Martínez-Flores I, Medina-Rivera A, et al. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. Nucleic Acids Res. 2013;41:D203–13.

58. MacQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics; 1967; Berkeley, Calif: University of California Press; 1967. p. 281–97.

59. Puniya BL, Allen L, Hochfelder C, Majumder M, Helikar T. Systems perturbation analysis of a large-scale signal transduction model reveals potentially influential candidates for cancer therapeutics. Front Bioeng Biotechnol. 2016;4:10.

60. Perez-Lopez AR, Szalay KZ, Turei D, Modos D, Lenti K, Korcsmaros T, Csermely P. Targets of drugs are generally, and targets of drugs having side effects are specifically good spreaders of human interactome perturbations. Sci Rep. 2015;5:10182.

61. Lv WH, YD X, Guo YY, ZQ Y, Feng GL, Liu PP, Luan MW, Zhu HJ, Liu GY, Zhang MM, et al. The drug target genes show higher evolutionary conservation than non-target genes. Oncotarget. 2016;7:4961–71.

62. Kotlyar M, Fortney K, Jurisica I. Network-based characterization of drug-regulated genes, drug targets, and toxicity. Methods. 2012;57:499–507.

63. Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M. Drug-target network. Nat Biotechnol. 2007;25:1119–26.

64. Sanchez R, Kim MY, Calonje M, Moon YH, Sung ZR. Temporal and spatial requirement of EMF1 activity for Arabidopsis vegetative and reproductive development. Mol Plant. 2009;2:643–53.

65. Kim SY, Lee J, Eshed-Williams L, Zilberman D, Sung ZR. EMF1 and PRC2 cooperate to repress key regulators of Arabidopsis development. PLoS Genet. 2012;8

66. Calonje M, Sanchez R, Chen LJ, Sung ZR. EMBRYONIC FLOWER1 participates in Polycomb group-mediated AG gene silencing in Arabidopsis. Plant Cell. 2008;20:277–91.

67. Shannon S, Meekswagner DR. Genetic interactions that regulate inflorescence development in Arabidopsis. Plant Cell. 1993;5:639–55.

68. Coursol S, Fan LM, Le Stunff H, Spiegel S, Gilroy S, Assmann SM. Sphingolipid signalling in Arabidopsis guard cells involves heterotrimeric G proteins. Nature. 2003;423:651–4.

69. Manjarrez-Orduno N, Marasco E, Chung SA, Katz MS, Kiridly JF, Simpfendorfer KR, Freudenberg J, Ballard DH, Nashi E, Hopkins TJ, et al. CSK regulatory polymorphism is associated with systemic lupus erythematosus and influences B-cell signaling and activation. Nat Genet. 2012;44:1227–30.

70. Vuica M, Desiderio S, Schneck JP. Differential effects of B cell receptor and B cell receptor-fc gamma RIIB1 engagement on docking of Csk to GTPase-activating protein (GAP)-associated p62. J Exp Med. 1997;186:259–67.

71. Gan YJ, Buckels A, Liu Y, Zhang Y, Paterson AJ, Jiang J, Zinn KR, Frank SJ, Human GH. Receptor-IGF-1 receptor interaction: implications for GH signaling (vol 28, pg 1841, 2014). Mol Endocrinol. 2015;29:332–2.