



OPEN

## Identification of anticancer drug target genes using an outside competitive dynamics model on cancer signaling networks

Tien-Dzung Tran<sup>1,2</sup>✉ & Duc-Tinh Pham<sup>1,3</sup>

Each cancer type has its own molecular signaling network. Analyzing the dynamics of molecular signaling networks can provide useful information for identifying drug target genes. In the present study, we consider an on-network dynamics model—the outside competitive dynamics model—wherein an inside leader and an opponent competitor outside the system have fixed and different states, and each normal agent adjusts its state according to a distributed consensus protocol. If any normal agent links to the external competitor, the state of each normal agent will converge to a stable value, indicating support to the leader against the impact of the competitor. We determined the total support of normal agents to each leader in various networks and observed that the total support correlates with hierarchical closeness, which identifies biomarker genes in a cancer signaling network. Of note, by experimenting on 17 cancer signaling networks from the KEGG database, we observed that 82% of the genes among the top 3 agents with the highest total support are anticancer drug target genes. This result outperforms those of four previous prediction methods of common cancer drug targets. Our study indicates that driver agents with high support from the other agents against the impact of the external opponent agent are most likely to be anticancer drug target genes.

Drugs bind to their target proteins/genes, which regulate downstream effectors and ultimately perturb the transcriptome of a cancer cell. Identification of novel drug target genes is a significant challenge in anticancer drug development<sup>1–3</sup>. In recent studies, the phenotypic effects and chemical structures of drugs have been used to infer drug–gene pairs. The phenotypic effect-based approaches exploit the various phenotypic responses, such as expression profiles and side effects, to external anticancer compounds<sup>4–7</sup>. On the assumption that structurally similar drugs tend to bind to similar genes, chemical structure-based approaches have been implemented and have shown promising results<sup>8–10</sup>. Although substantial progress has been made in this field, numerous challenges remain to be addressed. In phenotypic effect-based approaches, the drugs affecting different targets in the same pathway or in the same biological process may cause similar drug responses; in addition, gene expression patterns cannot distinguish target genes from downstream-regulated genes. Moreover, reportedly, the gene expression of drug targets is typically insignificantly affected by drug perturbation. Therefore, autonomous gene expression changes following drug treatment are insufficient to identify drug targets<sup>11</sup>. Chemical structure-based approaches often rely on a few proteins<sup>12,13</sup>, such as those with known interacting drugs<sup>14,15</sup> or with known three dimensional (3D) structures<sup>8,10</sup>. These approaches are insufficient for most proteins without such prior information.

In the past decade, anticancer drug target prediction has gained more interest with the availability of molecular biological network data, such as metabolic networks<sup>16,17</sup>, protein–protein interaction networks, gene regulatory networks<sup>18</sup>, heterogeneous network<sup>19–22</sup>, and cancer signaling networks<sup>23</sup>. Description and analysis of the network data provide a systematic understanding of drug action and disease complexity as well as help improve the efficiency of anticancer drug design<sup>24</sup>. Therefore, network-based methods have been developed to analyze

<sup>1</sup>Complex Systems and Bioinformatics Lab, Faculty of Information and Communication Technology, Hanoi University of Industry, Bac Tu Liem District, 298 Cau Dien street, Hanoi, Vietnam. <sup>2</sup>Department of Software Engineering, Faculty of Information and Communication Technology, Hanoi University of Industry, Bac Tu Liem District, 298 Cau Dien street, Hanoi, Vietnam. <sup>3</sup>Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam. ✉email: trantd@hau.edu.vn

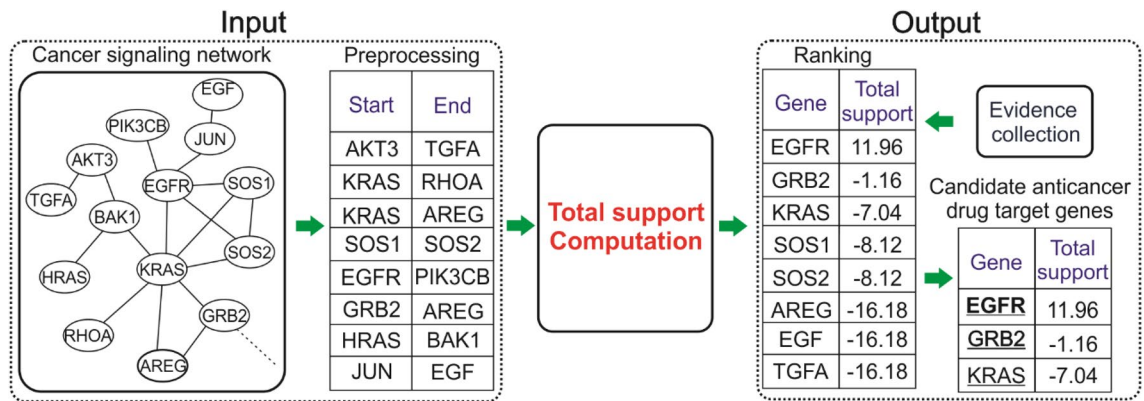
the structure and dynamics of molecular biological networks for improving several stages of drug discovery, particularly to predict drug target genes and designate new therapies<sup>25–29</sup>. Among cancer-related networks, cancer signaling networks are a heterogeneous network type and provide the most informative data for dynamics analysis because they contain both directed and undirected interaction types, rather than containing only one interaction type as observed with the other network types<sup>30,31</sup>. Additionally, if no additional data (such as gene expression data) are integrated into the analysis, computation on these networks often yield more precise prediction results compared with that on the others<sup>23</sup>. Structural analysis revealed that cancer biomarker genes, which lead to cancer via mutations, often reside at high hierarchical closeness positions in the innermost core of the cancer signaling networks<sup>25</sup>. Particularly, in dynamics analysis, the backbone driver nodes found in cancer signaling networks can drive the network into a cancer phenotype as well as steer it into a healthy phenotype. This means that backbone driver genes could be cancer biomarkers as well as cancer therapeutic targets<sup>32</sup>. Unfortunately, determining optimal driver nodes for drug targets in actual biological networks remains a challenge<sup>33</sup>, and a dynamic model for anticancer drug target identification requires further studies.

Recently, Zhao et al.<sup>34</sup> introduced a dynamic model that involved competition among two competitors for obtaining a maximum number of votes from other agents in a social network, where the two competitors within the same network have fixed and different states and each normal agent adjusts its state according to a distributed consensus protocol. The model can predict the bias of each normal agent and thus predict the competitor that will win. They found that the competition result completely depended on the network structure and positions of competitors in the network. Furthermore, it was observed the competitor with higher PageRank in a directed network or higher Katz Centrality in an undirected network has the highest likelihood to be the winner. Although these findings are extremely interesting, the research did not consider the case that one competitor is inside the network whereas the other is outside. This case is an extremely common phenomenon that often occurs in the field of social network and in the field of molecular biological network. In a social network, leaders inside the network often must counter the influence of competitors outside the system<sup>35,36</sup>. Similarly, in a molecular biological network, an environmental agent, such as UV radiation, drugs, chemicals, and viruses, can be considered the external competitor that causes perturbation against the signals of driver agents within the network<sup>37</sup>. Therefore, the external competition between two competitors can be considered a competition between an internal leader and an external opponent competitor in the social network or between a driver agent and an environmental agent in the cancer signaling network for obtaining maximum support from other agents in the network.

Here, we propose a dynamical network model called an outside competitive dynamics model in which an inside leader (driver agent, e.g., a drug target gene) and an opponent competitor outside the system (environmental agent, e.g., a drug) have fixed and different states. If any normal agent links to the external environmental competitor, the state of each normal agent will converge to a stable value, indicating support to (or impact from) the leader against the impact of the competitor. We showed an illustrative example of the working of the model in a disease network, which is formed by the integration of pathways related to a human disease such as a cancer, and the influence of adjacency weights on the outside competition results. We calculated the total support of normal agents to each leader in various networks (i.e., 17 actual and 100 random networks) and observed that the total support positively correlates with hierarchical closeness, the highest rankings of which were used to identify biomarker genes, which have been reported as therapeutic cancer targets in a cancer signaling network<sup>32</sup>. To reinforce the result observed, we gave an illustrative example to show that hierarchical closeness outperforms other popular centralities in the prediction for total support of a node. Interestingly, by experimenting on 17 cancer signaling networks downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG), we found that 82% of the genes among the top 3 agents with the highest total support are anticancer drug target genes. This result implies that genes with high support from the other genes against the impact of the external opponent agent are most likely to be anticancer drug target genes in a cancer signaling network. In other words, the top three agents with the highest total support may play driver nodes in a complex network. Finally, we used prediction results on common cancer drug targets of four previous network-based methods to validate our results. As a result, our top 1 prediction shared the highest consistency with the previous predictions. Overall, the outside competitive dynamics model contributes to the identification of both anticancer drug targets and driver agents in the cancer signaling network.

## Material and methods

**Overview of the process for identifying anticancer drug target genes.** The process to identify anticancer drug target genes using the outside competitive dynamics model is presented in Fig. 1. In the viewpoint of our study, a drug target gene should be a driver node of a disease network<sup>32</sup>. Therefore, this process factually detects the driver nodes of a disease network. First, a disease network is used as the input data of the process. In this study, each of 17 cancer signaling networks downloaded from the KEGG database was used as the input data in turn. Second, the network is pre-processed by replacing each group node (if any) with single nodes interacting with each other to form a heterogeneous network that includes two types of undirected and directed links. Third, the outside competitive model is applied to the network to identify driver nodes. To deploy the model on a network, each normal node is randomly assigned a state between -1 and 1, but the inside leader node, e.g., a driver gene, and the outside opponent competitor, e.g., a drug, that links to the normal node are fixed by opposition states. A random walk process is then simulated to determine whole network steady states, where can thereby compute the total steady-state, called total support, of every normal node to the leader. The algorithm of total support computation considers each node as a trial leader in competition with the outside competitor, and the trial leader node with the highest total support is eventually selected as the driver node. Because identification of driver nodes may be approximate due to noise from network construction, we selected driver nodes from the top three highest total support, and they are also candidate anticancer drug target genes.



**Figure 1.** Process to identify anticancer drug target genes using an outside competitive dynamics model. From an input network, the process uses an algorithm based on the proposed model to compute the total support score of nodes in the network to each node for selecting the top three highest-ranking nodes, which are considered drug target genes as the output.

Finally, these candidate anticancer drug target genes are matched with available evidence to identify real drug target genes and promising ones. Besides cancer, this process can be applied widely for cardiovascular, neurodegenerative, or other diseases with pathways available in open databases such as KEGG and BioCarta.

**Cancer signaling networks.** We downloaded 17 cancer pathways from the KEGG database ([www.genome.jp/kegg](http://www.genome.jp/kegg))<sup>38–40</sup> for conducting network analysis. Other pathway databases, such as BioCyc ([biocyc.org](http://biocyc.org))<sup>41</sup>, Reactome ([www.reactome.org](http://www.reactome.org))<sup>42</sup>, and BioGRID ([thebiogrid.org](http://thebiogrid.org))<sup>43</sup>, were not considered for analysis because they do not include pathways corresponding to a specific cancer site. Each cancer pathway was represented by a heterogeneous network, wherein a node and a link correspond to a protein and a protein–protein interaction, respectively. In the network, undirected links represent protein–protein interactions including binding/association and dissociation whereas directed links represent activation, inhibition, expression, indirect effect, interaction via compound, missing interaction, and phosphorylation. Cytoscape plugin KEGGParser<sup>44</sup> was used to correct the pathways after they were downloaded from KEGG pathway database because the original KGML (KEGG Markup Language) files were inconsistent to the static pathway map image. In addition, an interaction from a protein  $A$  to a group of proteins  $\{B_1, B_2, \dots, B_k\}$  in the original KEGG pathways was transformed into  $k$  different interactions of  $A \rightarrow B_1, A \rightarrow B_2, \dots, A \rightarrow B_k$  in the signaling network<sup>45</sup>.

**Computation of centrality measures.** Considering a heterogeneous network  $G(V, E)$ , we briefly introduce two well-known structural centrality measures as follows.

Closeness centrality: The closeness centrality<sup>46</sup> of a node  $u$  is defined as follows:

$$C_{clo}(u) = \frac{1}{\sum_{v \in V \setminus \{u\}} d(u, v)}, \quad (1)$$

where  $d(u, v)$  is the shortest distance from node  $u$  to node  $v$ . This measure has previously been used to prioritize disease genes in a protein–protein interaction network<sup>47,48</sup>. However, the definition of  $C_{clo}(u)$  is not proper in cases where there is a node  $v$  that is not reachable from  $u$  because  $C_{clo}(u)$  eventually becomes zero. Therefore, we used another version of closeness<sup>49</sup> as follows:

$$C_{clo-v}(u) = \frac{1}{|V| - 1} \sum_{v \in V \setminus \{u\}} \frac{1}{d(u, v)} \quad (2)$$

Hierarchical closeness: Hierarchical closeness<sup>29</sup> of a node  $u$  is  $C_{hc}(u)$  and is determined by combining reachability and closeness measures as follows:

$$C_{hc}(u) = N_R(u) + C_{clo-u}(u) \quad (3)$$

where  $N_R(u) \in [0, |V| - 1]$  is the reachability of a node  $u$  defined by  $N_R(u) = |\{v \in V \mid \exists \text{ a path from } u \text{ to } v\}|$ . Hierarchical closeness measure was successfully used to identify biomarker genes<sup>25</sup> and disease genes<sup>29</sup> on heterogeneous biological networks such as cancer signaling networks. In the present study, we used the finding regarding hierarchical closeness as evidence to support our results.

## Results

**A dynamic model for external competition.** We consider a disease network  $G(V, E)$  with  $N$  agents and  $M$  links. The agent (node/gene) set is denoted as  $V = \{1, 2, \dots, N\}$ , and the topology of the network is described by an adjacency matrix  $S = (s_{uv})_{N \times N}$ . If agent  $u$  is directly influenced by agent  $v$ , then there is a link from agent  $u$  to agent  $v$  and  $s_{uv} \in (0, 1]$  indicating the weight of the link; otherwise,  $s_{uv} = 0$ . For example,  $s_{uv} \in \{0, 1\}$  indicating the existence of links in biological networks<sup>50</sup>, and  $s_{uv} \in (0, 1]$  indicating the strength of ties in weighted biological

network<sup>51</sup>. We assumed that node  $\alpha \in V$  is a leader agent (a driver agent, e.g., a drug target gene) and node  $\beta \notin V$  is an outside opponent competitor (an environmental agent, e.g., a drug), where the state of the leader agent and the opponent agent have fixed and different states as follows:

$$x_\alpha(t) = +1, \forall t \geq 0; x_\beta(t) = -1, \forall t \geq 0 \tag{4}$$

There is an unknown link that may connect from  $\beta$  to any node in the network for causing perturbation against  $\alpha$ . Biologically, this link represents the transmission of the effect of drug  $\beta$  on any node against the impact of the driver agent  $\alpha$ . Therefore, it was assumed that an undirected link was temporarily added between node  $\beta$  and any node  $\gamma \in V$ , whenever  $\gamma$  adjusts its state. Every agent (called a normal node) denoted as  $u \in V/\{\alpha, \beta\}$  has a random initial state and updates its state as follows:

$$x_u(t + 1) = x_u(t) + \varepsilon \sum_v^{N_u} S_{uv}(x_v(t) - x_u(t)) \tag{5}$$

where  $x_u(t)$  is the state of agent  $u$  at time  $t$ ; the parameter  $0 < \varepsilon < W_{max}^{-1}$  captures the level of neighbors' influence, with  $W_{max}$  being the largest total weights of out-links of nodes in the network; and  $N_u = \{v \in V | S_{uv} \in (0, 1]\}$  is the set of neighboring nodes of node  $u$  that can directly influence node  $u$ . Equation (5) represents distributed consensus protocols proposed in the classical model of DeGroot<sup>52</sup>. The existence of competitors in the network disallows global consensus. Equation (5) biologically implies that expression state of a gene  $u$  for next period denoted by  $y_u(t+1) \in [0, 1]$  can be predicted by its current state  $y_u(t)$  plus an error term, where  $y_u(t) = (x_u(t) + 1)/2, x_u(t) \in [-1, 1]$ . With  $t \rightarrow \infty$ , the state of each normal node  $u$  converges to a steady value  $\bar{x}_u$ , which is a convex combination of the opponent states and independent of the initial states of nodes. The sign of the steady state of each normal node:  $\bar{x}_u > 0$  ( $\bar{x}_u < 0$ ) implies that node  $u$  will finally support or impacted by leader node  $\alpha$  (opponent node  $\beta$ ), and  $|\bar{x}_u|$  corresponds to the degree of support or impact.  $\bar{x}_u = 0$  if node  $u$  is a neutral node. In other words, this result means that if the random walk process<sup>53</sup> in Eq. (5) converges, it will determine whether the expression state of each normal gene is eventually more affected by drug  $\beta$  or by driver agent  $\alpha$ .

**Theorem 1** Given is a set  $X_{norm} \in R^{N-2}$ , which represents the state vector of all normal nodes in the network  $G(V,E)$  above.

$$X_{norm}(t) \xrightarrow{t \rightarrow \infty} \bar{X} \triangleq (\bar{D} - \bar{S})^{-1} [c_\alpha c_\beta] \begin{bmatrix} +1 \\ -1 \end{bmatrix}, \tag{6}$$

where  $\bar{D}, \bar{S}, [c_\alpha c_\beta]$  can all be derived from the network adjacency matrix  $S$ . If  $x_u(0) \in [-1, +1], \forall u \in V/\{\alpha, \beta\}$ , then  $x_u(t) \in [-1, +1], \forall t > 0$ .

*Proof of Theorem 1* We can rewrite Eq. (4+5) in the following matrix form:

$$X(t + 1) = (I_N - \varepsilon H L) X(t) = T X(t) \tag{7}$$

where  $I_N$  is an identity matrix;  $L = D - S$  is the Laplacian matrix,  $D$  is the diagonal matrix where  $D_{uu} = \sum_v S_{uv}$ ;  $H$  is an indicative diagonal matrix with  $H(i, i) = 0$  if agent  $i$  is a competitor and  $H(i, i) = 1$  otherwise. Indeed, the sum of each row of matrix  $T$  equals 1 because the sum of each row of matrix  $L$  always equals 0; therefore, the sum of each row of the matrix  $\varepsilon H L$  equals 0. For convenience, we reordered the agents to ensure that the two competitors come last. Accordingly, we have the following:

$$D = \begin{bmatrix} \bar{D} & 0 & 0 \\ 0 & d_\alpha & 0 \\ 0 & 0 & d_\beta \end{bmatrix} \text{ and } S = \begin{bmatrix} \bar{S} & c_\alpha & c_\beta \\ r_\alpha & 0 & * \\ r_\beta & * & 0 \end{bmatrix} \tag{8}$$

where  $d_\alpha$  and  $d_\beta$  denote the total weights of out-links of competitor  $\alpha$  and  $\beta$ , respectively; vectors  $c_\alpha, c_\beta, r_\alpha$ , and  $r_\beta$  contain the corresponding elements in the reordered adjacency matrix. Hence, Eq. (7) can be rewritten as follows:

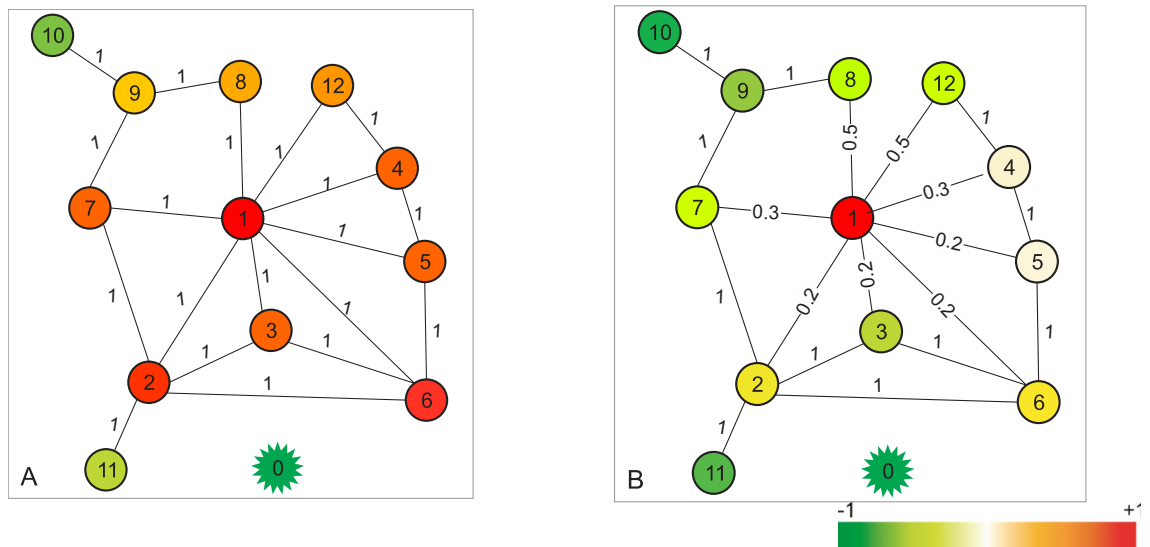
$$\begin{bmatrix} X_{norm}(t + 1) \\ x_\alpha(t + 1) \\ x_\beta(t + 1) \end{bmatrix} = \begin{bmatrix} Q & B \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_{norm}(t) \\ x_\alpha(t) \\ x_\beta(t) \end{bmatrix}, \tag{9}$$

where  $X_{norm} \in R^{N-2}$  represents the state vector of all normal agents;  $Q = I_{N-2} - \varepsilon(\bar{D} - \bar{S})$  and  $B = \varepsilon [c_\alpha c_\beta]$ . Thus,

$$X_{norm}(t) = Q X_{norm}(t - 1) + B \begin{bmatrix} x_\alpha(t - 1) \\ x_\beta(t - 1) \end{bmatrix} = Q^t X_{norm}(0) + \sum_{u=0}^{t-1} Q^u B \begin{bmatrix} x_\alpha(0) \\ x_\beta(0) \end{bmatrix} \tag{10}$$

If each normal agent has a path connecting to at least one competitor, then  $\bar{D} - \bar{S} \in R^{N-2}$  is invertible. Because  $0 < \varepsilon < W_{max}^{-1}$ , from the Gersgorin disk theorem, it can be demonstrated that the spectral radius of  $Q$  is less than 1. Thus, as  $t \rightarrow \infty$ , we have the following:

$$X_{norm}(t) \rightarrow (I_{N-2} - Q)^{-1} B \begin{bmatrix} x_\alpha(0) \\ x_\beta(0) \end{bmatrix} = (I_{N-2} - I_{N-2} + \bar{D} - \bar{S})^{-1} [c_\alpha c_\beta] \begin{bmatrix} x_\alpha(0) \\ x_\beta(0) \end{bmatrix} = (\bar{D} - \bar{S})^{-1} [c_\alpha c_\beta] \begin{bmatrix} +1 \\ -1 \end{bmatrix} \tag{11}$$



**Figure 2.** An illustrative example of how network structure influences the competitive impact results between driver gene and drug. A disease network with 12 genes and 19 interactions is given. Node 1 (red) is a driver gene whose state is fixed by 1. Node 0 (green) is a drug whose state is fixed by  $-1$ . An undirected interaction is temporarily added between the drug and each normal gene for computing support of the normal gene to the driver gene against impact from the drug. The state of each gene converges to a steady value which is a convex combination of the competitors' states, and does not depend on the initial states of genes. The color gradient represents support bias to two competitors. **(A)** The weights are kept at the default value of 1 for all links. The result shows that most genes in the network impacted by the driver gene, except 10 and 11 **(B)** A handful of links are changed in weight. Interestingly, there are large fluctuations in the network. Most genes in the network turn to impacted by the drug.

According to Lemma 4<sup>54</sup>, each entry of  $(D-S)^{-1}[c_\alpha c_\beta]$  is nonnegative and each row sum of  $(D-S)^{-1}[c_\alpha c_\beta]$  is equal to 1. Therefore, the steady state of each normal agent is a convex combination of  $+1$  and  $-1$ .

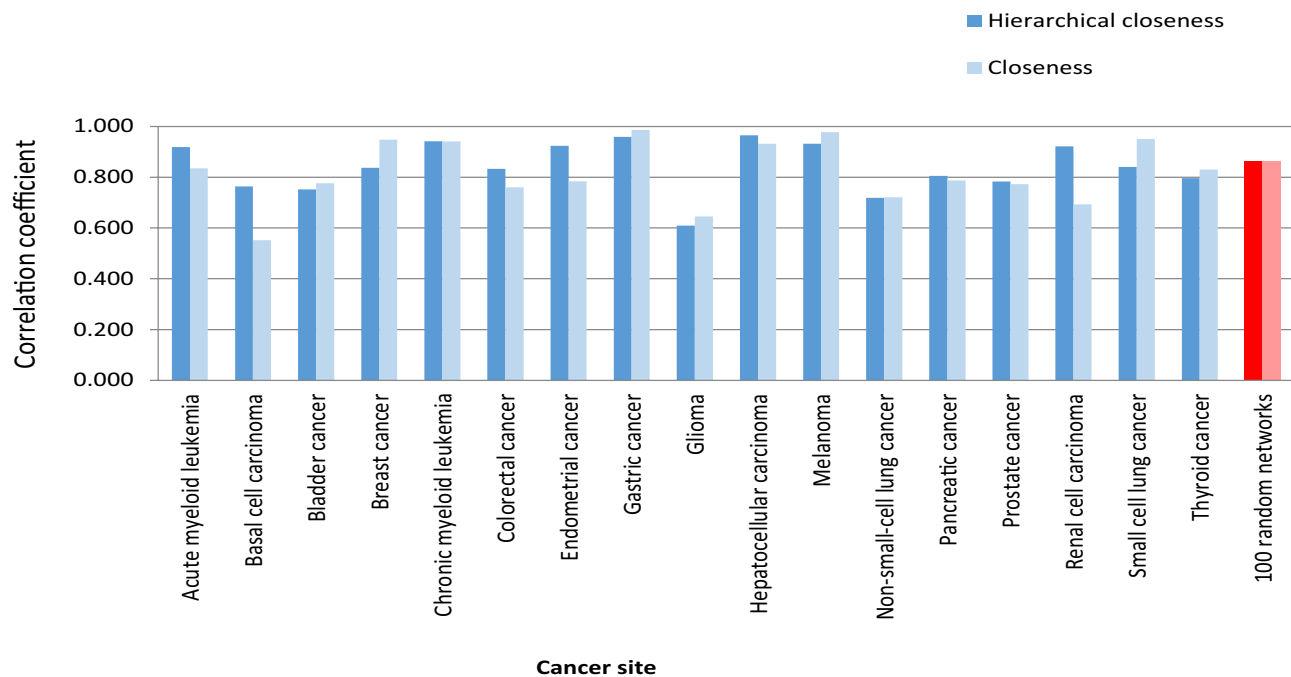
Further, we propose the measure of total support of normal nodes to node  $\alpha$  (hereafter referred to as total support/impact of  $\alpha$ ) against perturbation from  $\beta$  as follows:

$$ToS(\alpha) = \sum_{\gamma \in V \setminus \{\alpha\}} \bar{x}_\gamma \quad (12)$$

where  $\bar{x}_\gamma$  is the steady value of node  $\gamma$  if an undirected link is added between node  $\beta$  and  $\gamma$ . Total support of  $\alpha$ — $ToS(\alpha)$ —is computed by Algorithm S1 (see in Supplementary file). Driver agent of the network is identified by  $C = \max_{\alpha \in V} ToS(\alpha)$ . Driver agents may include a few nodes with the same total support, whose value may be only approximately computed because of network noise from the issues of measurement techniques and inherent natural variation<sup>51</sup>. Therefore, we selected driver agents from the top three nodes with the highest total support, which are also considered drug target genes in a disease network<sup>32</sup>.

**An illustration example.** Figure 2 shows outside competitive dynamics on two disease networks that have the same number of genes but different adjacency structures. We have considered a driver gene (node 1) and a drug (node 0) as two competitors in each network with fixed states  $x_1 = +1$  and  $x_0 = -1$ . To model the interaction between the drug and each normal gene, an undirected link was temporarily added between the drug and the normal gene, whenever this gene adjusts its state. We then computed the support of each gene to the driver gene against opposition impact from the drug. Stable states of normal genes were computed according to Eq. (6). A red (green) node represented a gene with a positive (negative) state. The darker the color, the larger was the absolute value of the state. White color nodes represented neutral agents. For network (A), the weights were maintained at the default value of 1 for all links. The result showed that most genes in the network, except 10 and 11, were impacted by the driver gene. For network (B), a handful of links were changed in weight. Most genes in the network converted to support the opponent drug outside the network. We observed that adjacency weights influenced the outside competition results, demonstrating the presence of large fluctuations in the network. In the following section, we will consider unweighted networks by considering that the weight of every link in a network is 1.

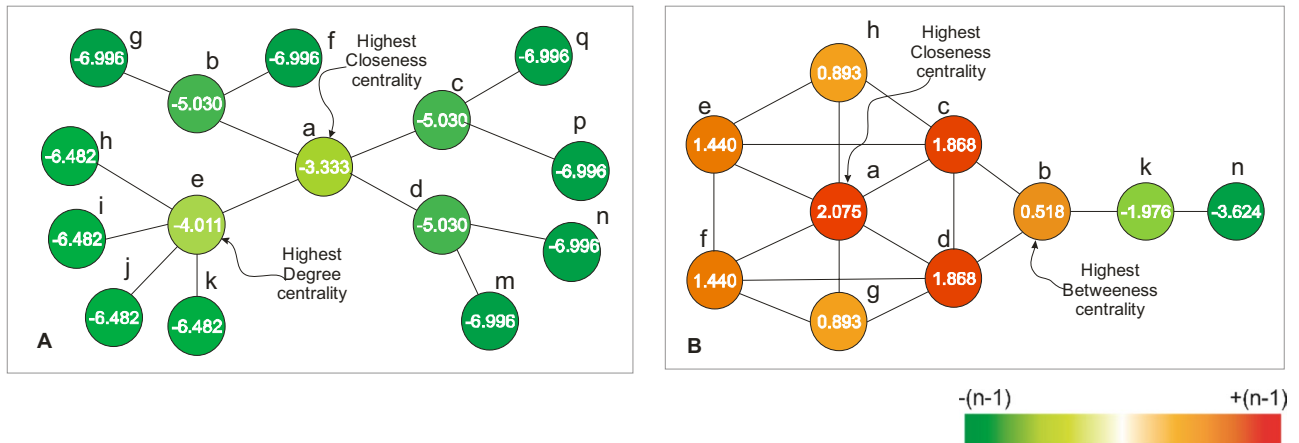
**Relationship between total support and hierarchical closeness.** Closeness is one of the most well-known structural centrality measures<sup>46</sup> wherein a node is defined as the inverse of the total sum of the shortest distance to the remaining nodes in an undirected network, and its effectiveness for disease gene prediction has frequently been reported for undirected biological networks<sup>47,48,55,56</sup>. Moreover, the closeness definition can be



**Figure 3.** Correlation coefficient between total support and closeness, hierarchical closeness. Blue columns indicate the results on 17 cancer signaling networks and the red represents those on 100 random directed networks generated with  $|V|=50$  and  $49 \leq |E| \leq 100$ . Dark red represents the value of hierarchical closeness ( $R=0.866$ ;  $P=0.0001$ ) whereas light red is the value of closeness ( $R=0.866$ ;  $P=0.0001$ ) (see Table S1 for details).

slightly modified to be properly used in a directed network<sup>49</sup>. In another study, Tran (2014) proposed an extensive closeness centrality called hierarchical closeness, which is a generalized measure of closeness centrality because it provides ranking results similar to that of closeness on an undirected network as well as functions efficiently on a directed or unconnected network<sup>29</sup>. The study found that hierarchical closeness outperforms other structural centrality measures in disease gene prediction. Moreover, the study showed that genes with a high level of hierarchical closeness were able to encode proteins in the extracellular matrix and receptor proteins in a human signal network. Particularly, hierarchical closeness was used to identify biomarker genes<sup>25</sup>, which have also been reported as cancer therapeutic targets in cancer signaling networks<sup>32</sup>. Both findings suggest that hierarchical closeness denotes both biomarker genes and drug target genes in cancer signaling networks. Considering the above results, we examined the relationship between hierarchical closeness of a node and total support of the node. To this end, the Barabasi–Albert network growth model<sup>57</sup> was used to generate random directed networks with the scale-free property inducing a few hub nodes and several non-hub nodes, as observed in real signaling networks. Investigation on the random networks generated by the model could statistically prove a network property in a generalized case. Experiment on 17 molecular signaling networks of cancers and 100 random directed networks generated with  $|V|=50$  and  $49 \leq |E| \leq 100$  showed that total support of each node positively correlates with both closeness and hierarchical closeness of the node (Fig. 3) because the correlation coefficients between total support and closeness (hierarchical closeness) on random networks was 0.866 ( $P < 0.05$ ). To illustrate the performance of hierarchical closeness on the prediction of total support, we demonstrated that the closeness centrality of an existing node is a significantly better predictor of total support of the node than the degree/betweenness centrality (Fig. 4). This result further provided an important basis to suggest that total support can be predicted by hierarchical closeness which indicates the closeness of a node to all other nodes. To clarify the role of total support in the identification of cancer therapeutic targets, we conducted an extensive experiment, the findings from which are detailed in the following section.

**Identification of anticancer drug target genes by total support.** Recent research on cancer signaling networks has shown that genes with high hierarchical closeness values were considered cancer biomarkers<sup>25</sup>, which also are often cancer therapeutic targets<sup>32</sup>. In the previous section, it was demonstrated that the hierarchical closeness of a node correlates with the total support of the node. These findings suggest that total support can accurately predict biomarker genes as well as drug target genes in a cancer signaling network. Considering these findings, we examined three genes with the highest total support of 17 cancer cell signaling networks (Table 1). Interestingly, 42 out of 51 genes (82%) were previously identified as anticancer drug target genes. For example, the three genes GRB2, FLT3, and PML, which were found in the acute myeloid leukemia (AML) signaling network, were considered key drug target genes. FLT3 is a common therapeutic target because it is frequently over-expressed or mutated, and its mutations indicate poor prognosis in AML. The development of FLT3 inhibitors leads to the recent approval of two drugs: Midostaurin (PKC412) and Gilteritinib (ASP2215) for the treatment of FLT3 mutant AML<sup>58</sup>. AML has a subtype called Acute promyelocytic leukemia (APL). The treatment of APL has

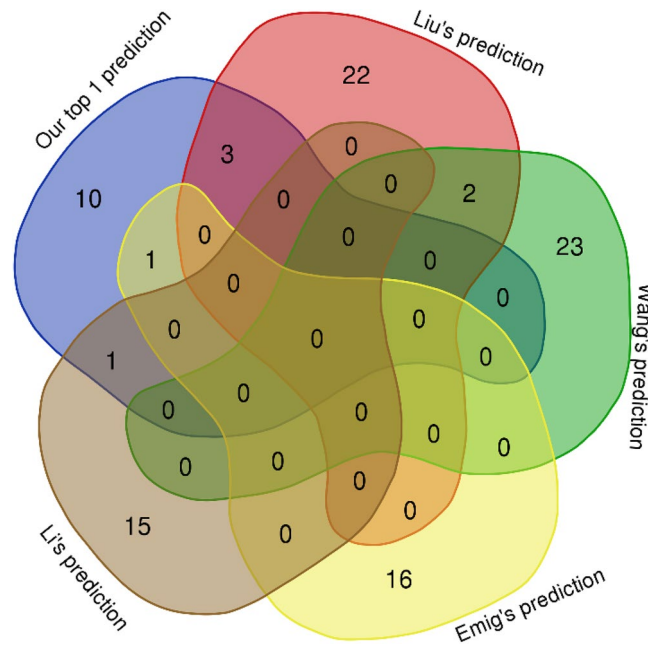


**Figure 4.** An illustrative example of the comparison between hierarchical closeness and other popular centralities for the prediction of total support. (A) Comparison with degree centrality. The highest closeness centrality outperforms the highest degree centrality in total support from the other nodes. (B) Comparison with betweenness centrality. The highest closeness centrality outperforms the highest betweenness centrality in total support from the remaining nodes. The color gradient represents the total support of a node ranging from  $-(n - 1)$  to  $+(n - 1)$ , where  $n$  is the number of nodes. Note that hierarchical closeness and closeness exhibit the same ranking result in these two networks.

Cancer site	Network properties		Candidate anticancer drug target genes by Totalsupport		
	The number of nodes	The number of links	C1	C2	C3
Acute myeloid leukemia	66	183	<u>GRB2</u>	<b>FLT3</b>	<b>PML</b>
Basal cell carcinoma	59	550	<u>SUFU</u>	<b>SMO</b>	<b>GLI3</b>
Bladder cancer	29	58	<u>RASSF1</u>	<b>FGFR3</b>	<b>HRAS</b>
Breast cancer	144	773	<u>LRP6</u>	<b>LRP5</b>	<b>WNT1</b>
Chronic myeloid leukemia	76	182	<u>CRK</u>	<b>CRKL</b>	<b>GAB2</b>
Colorectal cancer	74	195	<u>EGFR</u>	<b>GRB2</b>	<b>KRAS</b>
Endometrial cancer	51	117	<u>EGF</u>	<b>EGFR</b>	<b>AXIN1</b>
Gastric cancer	148	682	<u>LRP6</u>	<b>LRP5</b>	<b>WNT7A</b>
Glioma	74	310	<u>CALM1</u>	<b>CALML5</b>	<b>CALM2</b>
Hepatocellular carcinoma	167	769	<u>LRP6</u>	<b>WNT3A</b>	<b>WNT7A</b>
Melanoma	69	580	<u>FGF2</u>	<b>FGF1</b>	<b>HGF</b>
Non-small-cell lung cancer	65	157	<u>ALK</u>	<b>EML4</b>	<b>KRAS</b>
Pancreatic cancer	75	163	<u>KRAS</u>	<b>AKT2</b>	<b>AKT1</b>
Prostate cancer	84	375	<u>IGF-1</u>	<b>INS</b>	<b>PDGFB</b>
Renal cell carcinoma	56	154	<u>HGF</u>	<b>MET</b>	<b>EGLN2</b>
Small cell lung cancer	91	356	<u>ITGB1</u>	<b>COL4A1</b>	<b>LAMB3</b>
Thyroid cancer	37	84	<u>NTRK1</u>	<b>TPR</b>	<b>TPM3</b>

**Table 1.** Anticancer drug target genes are identified by total support ranking. In the table, C1, C2, and C3 denote NCBI gene symbols of the top three genes with the highest total support. The underlined and bold genes (42 out of 51) were previously reported as anticancer drug target genes (see Table S2), in which genes in bold are approved for drug manufacture whereas the remaining ones are clinical trial/potential. The remaining neither underlined nor bold genes in the top three highest-ranking genes that have not yet been fully investigated may be promising anticancer drug target genes.

dramatically been improved with the introduction of two synergism drugs of all-trans retinoic acid (ATRA) and arsenic trioxide (ATO) related to the targeting of genes PML<sup>59</sup>. The effect of GRB2 in AML treatment is under clinical trial with a phase II study of BP1001 that is a liposome-incorporated GRB2 antisense oligonucleotide for inhibition of GRB2 expression<sup>60</sup>. For another example, among the three genes HGF, MET, and EGLN2 identified in the renal cell carcinoma signaling network, HGF<sup>61</sup> and MET<sup>62,63</sup> have been known as potential and approved drug target genes, respectively; the third gene EGLN2 may be a promising anticancer drug target gene. A recent study demonstrated that the efficiency of cancer treatment may be significantly enhanced by combining drugs against multiple tumor specific drivers genes<sup>64</sup>. Taken together, 42 of the 51 genes (82%) that were reported as therapeutic targets confirmed that the top three genes with the highest total support score have the highest likeli-



**Figure 5.** Comparison with other prediction results. The Venn diagram was drawn based on the intersection of the predicted anticancer drug target genes in four previous reports and our top 1 prediction. Our top 1 prediction shares consistency with the most predictions of different methods, i.e., 3 out of 4 predictions. The figure was drawn with the online tool <http://bioinformatics.psb.ugent.be/webtools/Venn/>.

hood to be anticancer drug target genes; the 8 remaining genes that have not yet been completely investigated may be promising anticancer drug target genes. Although the predicted top three target genes of 8 networks don't have approved target genes, these predicted rank lists are also meaningful for further reference.

**Comparison with other methods.** We compared our method with four previous ones, namely: Liu's prediction<sup>65</sup>, Wang's prediction<sup>66</sup>, Emig's prediction<sup>67</sup>, and Li's prediction<sup>68</sup>. These methods are network-based approaches to predict drug targets of common cancers. Liu's prediction suggested 27 potential anticancer drug targets by a network-based screening of gene pairs on human cancer signaling network. Wang's prediction identified 25 candidate cancer drug targets by network score from genes sensitive with p53 mutation, which occurs in more than half of all human cancer cases. Emig's prediction found 17 cancer drug target genes by a combination of four network methods, namely: Neighborhood Scoring, Interconnectivity, Network Propagation, and Random Walks on a molecular interaction network associated with microarray experiment data. Li's prediction proposed 16 candidate anticancer drug targets using Random Walks on heterogeneous networks integrated from multi-source data. To compare with the above methods, we used our top 1 prediction (C1 in Table 1) including a list of 15 unique elements for common cancers. Note that the number of elements in our list was the smallest among the five predictions. We used top 1 prediction rather than top 3 prediction to guarantee that the comparison is not biased towards us because the size of our list is greater than those of the other lists. The Venn diagram in Fig. 5 showed that our's prediction and Liu's prediction had the same biggest number of intersection elements, i.e., 5 genes. Our intersection genes were composed of HGF, FGF2, ITGB1, EGFR, and GRB2, which most relates to growth factor. Especially, with the smallest number of elements, our prediction shared consistency with three different methods whereas Liu's prediction agreed with only two methods (see Table S3). This implies that our prediction outperforms the others, for it shared overlap with most of the remaining predictions whereas the number of elements was the smallest. All these methods have been used separately for predicting anticancer drug targets, and we believed using them together will provide better results.

## Discussion

Knowledge discovery of anticancer drug target genes is key to the research and development of successful drugs for cancer treatment. In the present study, an outside competitive dynamics model was used on the complex network to identify the drug target genes by calculating the total support of each agent in the cancer signaling network. First, we proposed a novel dynamic model for outside competition in a complex network and defined the total support of a node in the model. Total support of a leader node indicates the support degree of the other nodes to the leader as well as, conversely, represents the impact of the leader on the remaining nodes against the outside impact. Therefore, nodes with the highest total support score act as driver nodes that easily navigate the state of the other nodes in the network via signaling to the neighbors from the drivers. Second, we showed an illustrative example of the working of the model in a disease network and the influence of adjacency weights on the outside competition results. This example demonstrates that the number of links as well as the quality of links decides the impact of a leader on other nodes. Third, by experimenting on 17 molecular signaling



networks of cancers and 100 random directed networks, we showed that the total support of each node positively correlates with closeness and hierarchical closeness of the node. This finding indicates that the closer a node is to the remaining nodes, the more it receives support from them and the better it influences them. Note that a node with the highest closeness ranking is unnecessarily a network hub. Therefore, it suggests that a driver node often is at the central position of a disease network, where it may have only a few interactions to control the remaining nodes. Recent studies have reported that driver nodes often have the highest betweenness<sup>69</sup> or the highest degree<sup>70</sup> in some network types. We compared the prediction of total support between hierarchical closeness and degree/betweenness to demonstrate that a node with the highest closeness may be a better driver node than the others (Fig. 4). This finding was relatively different from that of the previous reports<sup>69,70</sup>. Note that we do not refute the previous findings because the closeness centrality used in the present study is a new version that was different from that used previously. Fourth, considering a finding that the top highest hierarchical closeness-ranking nodes often play biomarker genes<sup>25</sup> and a report that backbone driver nodes often act as both biomarkers and drug target genes in the cancer signaling network<sup>32</sup>, we conducted an extensive investigation on the performance of total support in the identification of drug target genes. Of note, when we examined the top 3 genes with the highest total support, 82% of the genes were previously found to be anticancer drug target genes and the remaining genes that have not yet been completely investigated may be promising anticancer drug target genes. This result is consistent with that of a previous report<sup>32</sup> and is evidence that genes with high support from the other genes in a cancer signaling network are most likely to be anticancer drug target genes and act as driver nodes in the network. Although the predicted top three target genes of 8 networks don't have approved target genes, these predicted rank lists are also meaningful for further reference. Finally, we validated our cancer drug target prediction by comparison with four previous network-based methods. As a result, our top 1 prediction shared the highest consistency among predictions by different methods. Overall, the outside competitive dynamics model contributes to the identification of both anticancer drug targets and driver agents in the cancer signaling network.

Our study showed that total support is a novel dynamic centrality measure for effective identification of anticancer drug target genes on cancer signaling networks. Although we only applied the outside competitive dynamics model on cancer signaling networks to identify anticancer drug target genes, the model can also be generalized to identify driver agents in a complex network. The positive relationship between total support and hierarchical closeness suggests that a driver agent in a network might use techniques such as hierarchical closeness optimization to adjust the network structure for winning the competition to the outside opponent competitor in various network types. However, obtaining valid evidence to demonstrate that the model functions satisfactorily in such a network type remains a key challenge. In addition, the algorithm to compute the total support presented in this study is limited by the long-running time required, particularly for large networks. All these issues will be considered in future studies.

Received: 18 February 2021; Accepted: 23 June 2021

Published online: 08 July 2021

## References

- Li, K., Du, Y., Li, L. & Wei, D. Q. Bioinformatics approaches for anti-cancer drug discovery. *Curr. Drug Targets* **21**, 3–17. <https://doi.org/10.2174/1389450120666190923162203> (2020).
- Lindsay, M. A. Target discovery. *Nat Rev Drug Discov* **2**, 831–838. <https://doi.org/10.1038/nrd1202> (2003).
- Nieto Gutierrez, A. & McDonald, P. H. GPCRs: emerging anti-cancer drug targets. *Cell Signal* **41**, 65–74. <https://doi.org/10.1016/j.cellsig.2017.09.005> (2018).
- Parsons, A. B. *et al.* Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat. Biotechnol.* **22**, 62–69. <https://doi.org/10.1038/nbt919> (2004).
- Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L. J. & Bork, P. Drug target identification using side-effect similarity. *Science* **321**, 263–266. <https://doi.org/10.1126/science.1158140> (2008).
- Lamb, J. *et al.* The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935. <https://doi.org/10.1126/science.1132939> (2006).
- Moffat, J. G., Rudolph, J. & Bailey, D. Phenotypic screening in cancer drug discovery—past, present and future. *Nat. Rev. Drug Discovery* **13**, 588–602. <https://doi.org/10.1038/nrd4366> (2014).
- Cheng, A. C. *et al.* Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **25**, 71–75. <https://doi.org/10.1038/nbt1273> (2007).
- Bleakley, K. & Yamanishi, Y. Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics* **25**, 2397–2403. <https://doi.org/10.1093/bioinformatics/btp433> (2009).
- Xie, L., Li, J., Xie, L. & Bourne, P. E. Drug discovery using chemical systems biology: identification of the protein–ligand binding network to explain the side effects of CETP inhibitors. *PLoS Comput. Biol.* **5**, e1000387. <https://doi.org/10.1371/journal.pcbi.1000387> (2009).
- Isik, Z., Baldow, C., Cannistraci, C. V. & Schroeder, M. Drug target prioritization by perturbed gene expression and network information. *Sci. Rep.* **5**, 17417. <https://doi.org/10.1038/srep17417> (2015).
- Cleves, A. E. & Jain, A. N. Robust ligand-based modeling of the biological targets of known drugs. *J. Med. Chem.* **49**, 2921–2938. <https://doi.org/10.1021/jm051139t> (2006).
- Nigsch, F., Bender, A., Jenkins, J. L. & Mitchell, J. B. Ligand-target prediction using Winnow and naive Bayesian algorithms and the implications of overall performance statistics. *J. Chem. Inf. Model.* **48**, 2313–2325. <https://doi.org/10.1021/ci800079x> (2008).
- Nidhi, G. M., Davies, J. W. & Jenkins, J. L. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J. Chem. Inf. Model.* **46**, 1124–1133. <https://doi.org/10.1021/ci060003g> (2006).
- Bleakley, K. & Yamanishi, Y. Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics (Oxford, England)* **25**, 2397–2403. <https://doi.org/10.1093/bioinformatics/btp433> (2009).
- Folger, O. *et al.* Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* **7**, 501. <https://doi.org/10.1038/msb.2011.35> (2011).

17. Chen, J., Ma, M., Shen, N., Xi, J. J. & Tian, W. Integration of cancer gene co-expression network and metabolic network to uncover potential cancer drug targets. *J. Proteome Res.* **12**, 2354–2364. <https://doi.org/10.1021/pr400162t> (2013).
18. Xie, Y., Wang, R. & Zhu, J. Construction of breast cancer gene regulatory networks and drug target optimization. *Arch. Gynecol. Obstet.* **290**, 749–755. <https://doi.org/10.1007/s00404-014-3264-y> (2014).
19. Chen, L., Lu, J., Huang, T. & Cai, Y.-D. A computational method for the identification of candidate drugs for non-small cell lung cancer. *PLoS ONE* **12**, e0183411. <https://doi.org/10.1371/journal.pone.0183411> (2017).
20. Lu, J. *et al.* Identification of new candidate drugs for lung cancer using chemical-chemical interactions, chemical-protein interactions and a K-means clustering algorithm. *J. Biomol. Struct. Dyn.* **34**, 906–917. <https://doi.org/10.1080/07391102.2015.1060161> (2016).
21. Chen, L. *et al.* Finding candidate drugs for hepatitis c based on chemical-chemical and chemical-protein interactions. *PLoS ONE* **9**, e107767. <https://doi.org/10.1371/journal.pone.0107767> (2014).
22. Li, B. Q. *et al.* Identifying chemicals with potential therapy of HIV based on protein-protein and protein-chemical interaction network. *PLoS ONE* **8**, e65207. <https://doi.org/10.1371/journal.pone.0065207> (2013).
23. Prasasya, R. D., Tian, D. & Kreeger, P. K. Analysis of cancer signaling networks by systems biology to develop therapies. *Semin. Cancer Biol.* **21**, 200–206. <https://doi.org/10.1016/j.semcancer.2011.04.001> (2011).
24. Csermely, P., Korcsmáros, T., Kiss, H. J. M., London, G. & Nussinov, R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* **138**, 333–408. <https://doi.org/10.1016/j.pharmthera.2013.01.016> (2013).
25. Tran, T.-D. & Kwon, Y.-K. Hierarchical closeness-based properties reveal cancer survivability and biomarker genes in molecular signaling networks. *PLoS ONE* **13**, e0199109. <https://doi.org/10.1371/journal.pone.0199109> (2018).
26. Hao, T. *et al.* Analyzing of Molecular Networks for Human Diseases and Drug Discovery. *Curr. Top. Med. Chem.* **18**, 1007–1014. <https://doi.org/10.2174/1568026618666180813143408> (2018).
27. Ma, J. *et al.* Network-based method for drug target discovery at the isoform level. *Sci. Rep.* **9**, 13868. <https://doi.org/10.1038/s41598-019-50224-x> (2019).
28. Kotlyar, M., Fortney, K. & Jurisica, I. Network-based characterization of drug-regulated genes, drug targets, and toxicity. *Methods* **57**, 499–507. <https://doi.org/10.1016/j.ymeth.2012.06.003> (2012).
29. Tran, T.-D. & Kwon, Y.-K. Hierarchical closeness efficiently predicts disease genes in a directed signaling network. *Comput. Biol. Chem.* **53**, 191–197. <https://doi.org/10.1016/j.compbiolchem.2014.08.023> (2014).
30. Mousavian, Z., Díaz, J. & Masoudi-Nejad, A. Information theory in systems biology. Part II: protein-protein interaction and signaling networks. *Semin. Cell Dev. Biol.* **51**, 14–23. <https://doi.org/10.1016/j.semcdb.2015.12.006> (2016).
31. Creixell, P., Schoof, E. M., Erler, J. T. & Linding, R. Navigating cancer network attractors for tumor-specific therapy. *Nat. Biotechnol.* **30**, 842–848. <https://doi.org/10.1038/nbt.2345> (2012).
32. Ravindran, V. & Bagler, G. Identification of critical regulatory genes in cancer signaling network using controllability analysis. *Physica A Stat. Mech. Appl.* **474**, 134–143. <https://doi.org/10.1016/j.physa.2017.01.059> (2019).
33. Guo, W.-F. *et al.* A novel algorithm for finding optimal driver nodes to target control complex networks and its applications for drug targets identification. *BMC Genom.* **19**, 924. <https://doi.org/10.1186/s12864-017-4332-z> (2018).
34. Zhao, J., Liu, Q. & Wang, X. Competitive dynamics on complex networks. *Sci. Rep.* **4**, 5858. <https://doi.org/10.1038/srep05858> (2014).
35. Ketchen, D. J., Snow, C. C. & Hoover, V. L. Research on competitive dynamics: recent accomplishments and future challenges. *J. Manag.* **30**, 779–804. <https://doi.org/10.1016/j.jm.2004.06.002> (2004).
36. Hoppe, B. & Reinelt, C. Social network analysis and the evaluation of leadership networks. *Leadersh. Q.* **21**, 600–619. <https://doi.org/10.1016/j.leaqua.2010.06.004> (2010).
37. Wogan, G. N., Hecht, S. S., Felton, J. S., Conney, A. H. & Loeb, L. A. Environmental and chemical carcinogenesis. *Semin. Cancer Biol.* **14**, 473–486. <https://doi.org/10.1016/j.semcancer.2004.06.010> (2004).
38. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
39. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Prot. Sci. Publ. Prot. Soc.* **28**, 1947–1951. <https://doi.org/10.1002/pro.3715> (2019).
40. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucl. Acids Res.* **49**, D545–d551. <https://doi.org/10.1093/nar/gkaa970> (2021).
41. Caspi, R. *et al.* BioCyc: online resource for genome and metabolic pathway analysis. *FASEB J.* **30**, 192. [https://doi.org/10.1096/fasebj.30.1\\_supplement.lb192](https://doi.org/10.1096/fasebj.30.1_supplement.lb192) (2016).
42. Fabregat, A. *et al.* The reactome pathway knowledgebase. *Nucl. Acids Res.* **46**, D649–D655. <https://doi.org/10.1093/nar/gkx1132> (2017).
43. Chatr-aryamontri, A. *et al.* The BioGRID interaction database: 2015 update. *Nucl. Acids Res.* **43**, D470–D478. <https://doi.org/10.1093/nar/gku1204> (2014).
44. Arakelyan, A. & Nersisyan, L. KEGGParser: parsing and editing KEGG pathway maps in Matlab. *Bioinformatics* **29**, 518–519. <https://doi.org/10.1093/bioinformatics/bts730> (2013).
45. Tran, T.-D. & Kwon, Y.-K. The relationship between modularity and robustness in signalling networks. *J. R. Soc. Interface* **10**, 20130771. <https://doi.org/10.1098/rsif.2013.0771> (2013).
46. Sabidussi, G. The centrality index of a graph. *Psychometrika* **31**, 581–603. <https://doi.org/10.1007/BF02289527> (1966).
47. Gottlieb, A., Magger, O., Berman, I., Ruppín, E. & Sharan, R. PRINCIPLE: a tool for associating genes with diseases via network propagation. *Bioinformatics* **27**, 3325–3326. <https://doi.org/10.1093/bioinformatics/btr584> (2011).
48. Hsu, C. L., Huang, Y. H., Hsu, C. T. & Yang, U. C. (2011) Prioritizing disease candidate genes by a gene interconnectedness-based approach. *BMC Genom.* **12**, 25. <https://doi.org/10.1186/1471-2164-12-s3-s25> (2011).
49. Opsahl, T., Agneessens, F. & Skvoretz, J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc. Netw.* **32**, 245–251. <https://doi.org/10.1016/j.socnet.2010.03.006> (2010).
50. Truong, C.-D., Tran, T.-D. & Kwon, Y.-K. MORO: a Cytoscape app for relationship analysis between modularity and robustness in large-scale biological networks. *BMC Syst. Biol.* **10**, 122. <https://doi.org/10.1186/s12918-016-0363-3> (2016).
51. Wang, B. *et al.* Network enhancement as a general method to denoise weighted biological networks. *Nat. Commun.* **9**, 3108. <https://doi.org/10.1038/s41467-018-05469-x> (2018).
52. Degroot, M. H. Reaching a consensus. *J. Am. Stat. Assoc.* **69**, 118–121. <https://doi.org/10.1080/01621459.1974.10480137> (1974).
53. Pearson, K. The problem of the random walk. *Nature* **72**, 294–294. <https://doi.org/10.1038/072294b0> (1905).
54. Meng, Z., Ren, W. & You, Z. Distributed finite-time attitude containment control for multiple rigid bodies. *Automatica* **46**, 2092–2099. <https://doi.org/10.1016/j.automatica.2010.09.005> (2010).
55. Erten, S., Bebek, G., Ewing, R. M. & Koyutürk, M. DADA: degree-aware algorithms for network-based disease gene prioritization. *BioData Min* **4**, 19. <https://doi.org/10.1186/1756-0381-4-19> (2011).
56. Wu, X., Jiang, R., Zhang, M. Q. & Li, S. Network-based global inference of human disease genes. *Mol. Syst. Biol.* **4**, 189–189. <https://doi.org/10.1038/msb.2008.27> (2008).
57. Barabasi, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512. <https://doi.org/10.1126/science.286.5439.509> (1999).

58. Gebru, M. T. & Wang, H.-G. Therapeutic targeting of FLT3 and associated drug resistance in acute myeloid leukemia. *J. Hematol. Oncol.* **13**, 155. <https://doi.org/10.1186/s13045-020-00992-1> (2020).
59. Testa, U. & Lo-Coco, F. Targeting of leukemia-initiating cells in acute promyelocytic leukemia. *Stem cell investigation* **2**, 8–8. <https://doi.org/10.3978/j.issn.2306-9759.2015.04.03> (2015).
60. Ohanian, M. *et al.* A phase II study of BP1001 (liposomal Grb2 antisense oligonucleotide) in patients with hematologic malignancies. *J. Clin. Oncol.* **38**, 7561. [https://doi.org/10.1200/JCO.2020.38.15\\_suppl.TPS7561](https://doi.org/10.1200/JCO.2020.38.15_suppl.TPS7561) (2020).
61. Giubellino, A., Linehan, W. M. & Bottaro, D. P. Targeting the Met signaling pathway in renal cancer. *Expert Rev. Anticancer Ther.* **9**, 785–793. <https://doi.org/10.1586/era.09.43> (2009).
62. Nandagopal, L., Sonpavde, G. P. & Agarwal, N. Investigational MET inhibitors to treat Renal cell carcinoma. *Expert Opin. Investig. Drugs* **28**, 851–860. <https://doi.org/10.1080/13543784.2019.1673366> (2019).
63. Alonso-Gordoa, T. *et al.* Targeting Tyrosine kinases in Renal Cell Carcinoma: “New Bullets against Old Guys”. *Int. J. Mol. Sci.* **20**, 1901 (2019).
64. Zsákai, L. *et al.* Targeted drug combination therapy design based on driver genes. *Oncotarget* **10**, 5255–5266. <https://doi.org/10.18632/oncotarget.26985> (2019).
65. Liu, L. *et al.* Synthetic lethality-based identification of targets for anticancer drugs in the human signaling network. *Sci. Rep.* **8**, 8440. <https://doi.org/10.1038/s41598-018-26783-w> (2018).
66. Wang, X. & Simon, R. Identification of potential synthetic lethal genes to p53 using a computational biology approach. *BMC Med. Genomics* **6**, 30. <https://doi.org/10.1186/1755-8794-6-30> (2013).
67. Emig, D. *et al.* Drug Target Prediction and Repositioning Using an Integrated Network-Based Approach. *PLoS ONE* **8**, e60618. <https://doi.org/10.1371/journal.pone.0060618> (2013).
68. Li, C. *et al.* Cancer-Drug Interaction Network Construction and Drug Target Prediction Based on Multi-source Data. In *International Conference on Wireless Algorithms, Systems, and Applications*, 223–235. (Springer International Publishing). [https://doi.org/10.1007/978-3-319-94268-1\\_19](https://doi.org/10.1007/978-3-319-94268-1_19) (2018).
69. Abbasi, A., Hossain, L. & Leydesdorff, L. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *J. Informet.* **6**, 403–412. <https://doi.org/10.1016/j.joi.2012.01.002> (2012).
70. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473**, 167–173. <https://doi.org/10.1038/nature10011> (2011).

### Author contributions

D.T.P. drafted the manuscript, conducted experiments, and collected data for the project. T.D.T. conceived the study; designed and coded the experimental software; wrote, revised, and edited the manuscript; and provided guidance for the theoretical and mathematical issues and context. All authors contributed critical feedback and edited the manuscript.

### Funding

This research was funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant No. 102.04-2018.304.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-93336-z>.

**Correspondence** and requests for materials should be addressed to T.-D.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021