BioSystems 137 (2015) 64–72

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Effective Boolean dynamics analysis to identify functionally important genes in large-scale signaling networks

Hung-Cuong Trinh, Yung-Keun Kwon*

School of Electrical Engineering, University of Ulsan, 93, Daehak-ro, Nam-gu, Ulsan 680-749, Republic of Korea

ARTICLE INFO

Article history:
Received 23 January 2015
Received in revised form 13 July 2015
Accepted 16 July 2015
Available online 12 August 2015

Keywords:
Boolean dynamics
Boolean sensitivity
Update-rule perturbation
Centrality measure
Essential genes
Drug targets
Signaling network

ABSTRACT

Efficiently identifying functionally important genes in order to understand the minimal requirements of normal cellular development is challenging. To this end, a variety of structural measures have been proposed and their effectiveness has been investigated in recent literature; however, few studies have shown the effectiveness of dynamics-based measures. This led us to investigate a dynamic measure to identify functionally important genes, and the effectiveness of which was verified through application on two large-scale human signaling networks. We specifically consider Boolean sensitivity-based dynamics against an update-rule perturbation (BSU) as a dynamic measure. Through investigations on two large-scale human signaling networks, we found that genes with relatively high BSU values show slower evolutionary rate and higher proportions of essential genes and drug targets than other genes. Gene-ontology analysis showed clear differences between the former and latter groups of genes. Furthermore, we compare the identification accuracies of essential genes and drug targets via BSU and five well-known structural measures. Although BSU did not always show the best performance, it effectively identified the putative set of genes, which is significantly different from the results obtained via the structural measures. Most interestingly, BSU showed the highest synergy effect in identifying the functionally important genes in conjunction with other measures. Our results imply that Boolean-sensitive dynamics can be used as a measure to effectively identify functionally important genes in signaling networks.

1. Introduction

Essential genes are those that are necessary for growth in a rich medium where all the required nutrients are available (Acencio and Lemke, 2009). The deletion of any essential gene can lead to a lethal effect or cause infertility (Kamath et al., 2003; Winzeler et al., 1999). Drug targets are also believed to be functionally important because they can undergo a specific interaction with drug chemicals (Imming et al., 2006). Such the functionally important genes or proteins can be identified through experimental procedures including single gene knockouts (Giaever et al., 2002), conditional knockouts (Roemer et al., 2003) and RNA interference (Cullen and Arndt, 2005); however, these experimental methods require significant time and resources. Previous studies have proposed network structural properties for identifying the important genes, the simplest of which uses the degree to indicate the ability to communicate directly with other proteins (Jeong et al., 2001). It has been observed in several species such as *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus*, and *Homo sapiens*, that proteins of a high degree are more likely to be essential (Hahn and Kern, 2005; Liang and Li, 2007; Lin et al., 2009). In addition to degree, the betweenness (Freeman, 1977) and closeness (Wuchty and Stadler, 2003) based on the shortest paths have been proposed. Their relations to gene essentiality were partially proven in previous studies (Pržulj et al., 2004; Yu et al., 2007). Other measures such as stress (Shimbel, 1953) and eigenvector (Bonacich, 1987) were also used to identify essential genes (Li et al., 2011a, 2012; Park and Kim, 2009). Similarly, a number of experimental approaches based on proteomics technologies have been developed for target identification in drug development (Katayama and Oda, 2007; Sleno and Emili, 2008; Wang et al., 2007). Various network topology-based approaches have also been investigated in an effort to more efficiently detect new drug targets (Florez et al., 2010; Hwang et al., 2008; Li et al., 2011b; Lu et al., 2012).

These previous studies propose novel measures based on network structural properties to identify functionally important genes. However, further investigation into network dynamics-based indicators is needed, as these solutions cannot sufficiently elucidate causality. Some studies developed dynamics-based methods to

* Corresponding author.
E-mail address: kwonyk@ulsan.ac.kr (Y.-K. Kwon).

http://dx.doi.org/10.1016/j.biosystems.2015.07.007
0303-2647/

identify essential components (Azuaje et al., 2010; Li et al., 2006). A network perturbation analysis was carried out to identify potential targets in two protein signaling systems, the caspase-3 network (including 22 components and 26 interactions) and the adenosine network (including 30 components and 44 interactions) (Azuaje et al., 2010). In another study, a dynamics analysis model was suggested to predict essential components of the guard cell signal transduction network for abscisic acid-induced stomatal closure (approximately 40 components) (Li et al., 2006). However, they were limited to analyses on small signaling networks. Thus, it is necessary to employ a dynamics-based measure that can effectively identify functionally important genes in a large-scale biological network. In this paper, we consider Boolean sensitivity dynamics based on update-rule perturbation (BSU) as an effective dynamic measure which was proposed in a previous study (Le and Kwon, 2013). Through investigations on two large-scale human signaling networks, we found that genes with relatively high BSU values are likely to show slower evolutionary rate and higher proportions of essential genes and drug targets than the genes with relatively low BSU values. It was also observed through a gene ontology analysis that the former and latter sets of genes are enriched differently. We compared BSU and five well-known structural measures in order to assess the identification accuracies for essential genes and drug targets. Although BSU was not the best predictor in either case, it effectively identified the putative genes, which are significantly different from those obtained by the structural measures. It was also interesting that BSU showed the highest synergy effect in identifying the functionally important genes in conjunction with other measures.

2. Materials and methods

2.1. Datasets

Cellular signaling networks play pivotal roles in fundamental processes of a cell. We employed two large-scale human signaling networks, KEGG and WANG networks. The former was used in a previous study (Kim et al., 2011) and it consists of 1659 genes and 7964 interactions constructed by integrating all the human signaling pathways in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa and Goto, 2000). The latter network obtained from Cui and Wang papers (Cui et al., 2007, 2009) consists of 1609 genes and 5063 interactions by manually curating signaling pathways from BioCarta (http://www.biocarta.com/genes/Cellsignaling.asp) and Cancer Cell Map (Croft et al., 2011). To examine the evolutionary rates (human–mouse dN/dS) of genes, we used the Evola database (Matsuya et al., 2008). In addition, we considered essential or drug-targeted genes as functionally important. By using the DEG (Database of Essential Genes, version 5.4) database (Zhang and Lin, 2009), we found 473 and 615 essential genes included in the KEGG and WANG networks, respectively. We also found 353 and 392 drug target genes in the KEGG and WANG networks, respectively, using the DrugBank database (Knox et al., 2011).

2.2. Boolean dynamics in a Boolean network model

In order to show the effectiveness of Boolean dynamics in identifying functionally important genes, we employed a Boolean network model that has been frequently used to investigate the complex dynamics of biological networks (Chaouiya et al., 2013; Kauffman et al., 2003; Kwon et al., 2007; Mendes et al., 2013; Saez-Rodriguez et al., 2007; Shmulevich et al., 2003). A Boolean network is represented by a directed graph $G(V,A)$, where V is a set of Boolean variables and A is a set of ordered pairs of the Boolean variables called directed links. Each $v_i \in V$ has a value of 1 ("on") or 0 ("off"), which represents the possible states of the corresponding gene. A state of a Boolean network is defined as a vector of the states of all nodes. A directed link (v_i, v_j) has a positive ("activating") or negative ("inhibiting") relationship from v_i to v_j. The value of each variable v_i at time $t+1$ is determined by the values of k_i other variables $v_{i_1}, v_{i_2}, \ldots, v_{i_{k_i}}$ with a link to v_i at time t by the Boolean function $f_i : \{0, 1\}^{k_i} \rightarrow \{0, 1\}$; all variables are synchronously updated. Hence, the update rule can be written as $v_i(t+1) = f_i(v_{i_1}(t), v_{i_2}(t), \ldots, v_{i_{k_i}}(t))$. Here, we employ a nested canalyzing function (NCF) model (Kauffman et al., 2003; see Supporting Text in Supplementary Information for details) which can represent f_i as follows:

$$v_i(t+1) = \begin{cases} O_1 & \text{if } v_{i_1}(t) = I_1 \\ O_2 & \text{if } v_{i_1}(t) \neq I_1 \text{ and } v_{i_2}(t) = I_2 \\ O_3 & \text{if } v_{i_1}(t) \neq I_1 \text{ and } v_{i_2}(t) \neq I_2 \text{ and } v_{i_s}(t) = I_3 \\ & \vdots \\ O_{k_i} & \text{if } v_{i_1}(t) \neq I_1 \text{ and} \cdots \text{and } v_{i_{k_i-1}}(t) \neq I_{k_i-1} \text{ and } v_{i_{k_i}}(t) = I_{k_i} \\ 1 - O_{k_i} & \text{if } v_{i_1}(t) \neq I_1 \text{ and} \cdots \text{and } v_{i_{k_i}}(t) \neq I_{k_i} \end{cases}$$

where I_m and O_m ($m = 1, 2, \ldots, k_i$) are canalyzing and canalyzed Boolean values which are randomly specified. The NCF model can generate various canalyzing rules which are ubiquitously found in molecular interactions (Harris et al., 2002; Samal and Jain, 2008). It was also shown that NCFs properly fit the experimental data obtained from a literature (Kauffman et al., 2003), and many logical interaction rules inferred from gene expression data can be represented by NCFs (Faure et al., 2006; González et al., 2006; Harris et al., 2002; Murrugarra and Laubenbacher, 2011; Naldi et al., 2010). For example, 133 out of 139 rules compiled from a dataset about a transcriptional regulatory network (Harris et al., 2002) and 39 out of 42 rules inferred from a dataset about signaling pathways (Naldi et al., 2010) were NCFs.

In this paper, we extend a network sensitivity against perturbations in terms of Boolean dynamics used in a previous study (Le and Kwon, 2013). Given a Boolean network, a state trajectory starts from an initial state and eventually converges to either a fixed-point or limit-cycle attractor. These attractors can represent diverse biological network behaviors such as multi-stability, homeostasis, and oscillation (Bhalla et al., 2002; Ferrell and Machleder, 1998; Pomerening et al., 2003). Based on the definition of the attractor, we introduce the notion of sensitivity in terms of converging dynamics. Here, we consider an update-rule perturbation, which can describe a dysfunction in interactions between genes/proteins in pathways. Previous studies having shown the significant impact of updating-function perturbations on network dynamics (Xiao and Dougherty, 2007; Xiaoning and Dougherty, 2008). If an update-rule perturbation occurs, the landscape of the network state transitions is typically changed. When the trajectory starting from the same initial state along with the update-rule perturbation converges to a different attractor, the network is determined to be sensitive to the perturbation.

Given a Boolean network $G(V,A)$ with a set of update-rules $F = \{f_1, f_2, \ldots, f_N\}$ and an initial state $s = [v_1(0), v_2(0), \ldots, v_N(0)]$, an update-rule perturbation at a node $v_i \in V$ involves a scenario where F is changed to $F' = \{f_1, \ldots, f_i', \ldots, f_N\}$, where f_i' denotes that every canalyzing and canalyzed values are flipped (i.e., all I_m and O_m are changed into $1 - I_m$ and $1 - O_m$, respectively). The update-rule perturbation may represent a deleterious change in the function of a protein or gene (Ng and Henikoff, 2003). For a set of