Journal of Theoretical Biology 362 (2014) 17–34

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Core and specific network markers of carcinogenesis from multiple cancer samples

Yung-Hao Wong, Ru-Hong Chen, Bor-Sen Chen *

Lab of Control and Systems Biology, Department of Electrical Engineering National Tsing Hua University, Hsinchu 30013, Taiwan

HIGHLIGHTS

- Using a systems biology approach to construct the protein–protein interaction (PPI) networks of four cancers and the non-cancers.
- Find significant proteins with large PPI changes during carcinogenesis process.
- A total of 28 significant proteins were identified as core network markers in the carcinogenesis of four types of cancer and a specific network marker is also found for each other.
- Pathway analysis of these significant proteins reveals the hidden cancer mechanism.

ARTICLE INFO

Article history:
Received 12 November 2013
Received in revised form
19 May 2014
Accepted 28 May 2014
Available online 9 July 2014

Keywords:
Network markers
Carcinogenesis relevance value
Protein–protein interaction

ABSTRACT

Cancer is the leading cause of death worldwide and is generally caused by mutations in multiple proteins or the dysregulation of pathways. Understanding the causes and the underlying carcinogenic mechanisms can help fight this disease. In this study, a systems biology approach was used to construct the protein–protein interaction (PPI) networks of four cancers and the non-cancers by their corresponding microarray data, PPI modeling and database-mining. By comparing PPI networks between cancer and non-cancer samples to find significant proteins with large PPI changes during carcinogenesis process, core and specific network markers were identified by the intersection and difference of significant proteins, respectively, with carcinogenesis relevance values (CRVs) for each cancer. A total of 28 significant proteins were identified as core network markers in the carcinogenesis of four types of cancer, two of which are novel cancer-related proteins (e.g., UBC and PSMA3). Moreover, seven crucial common pathways were found among these cancers based on their core network markers, and some specific pathways were particularly prominent based on the specific network markers of different cancers (e.g., the RIG-I-like receptor pathway in bladder cancer, the proteasome pathway and TCR pathway in liver cancer, and the HR pathway in lung cancer). Additional validation of these network markers using the literature and new tested datasets could strengthen our findings and confirm the proposed method. From these core and specific network markers, we could not only gain an insight into crucial common and specific pathways in the carcinogenesis, but also obtain a high promising PPI target for cancer therapy.

1. Introduction

Cancer is the leading cause of one in eight deaths worldwide (Stratton et al., 2009). It is a large group of over 100 distinct diseases with various risk factors and its development is a multi-step process, including the dysfunction of genes that regulate the proliferation, division, and death of cells, as well as the tissue microenvironment. Defects and glitches in these gene-encoded proteins lead to a series of acquired functional capabilities that allow undisciplined growth and invasion of abnormal cells into other tissues (Neal and Yu, 2010). Thus, evidence suggests that most cancers are not caused by one single factor or event; instead, a normal cell develops through a sequence of pre-malignant states into an invasive cancer. To determine the molecular mechanisms of cancer, the detection of potential cancer in early stages before tumors are visible is practicable.

* Corresponding author.
E-mail addresses: yvonwong.nthu@gmail.com (Y.-H. Wong), leslie760917@hotmail.com (R.-H. Chen), bschen@ee.nthu.edu.tw (B.-S. Chen).

http://dx.doi.org/10.1016/j.jtbi.2014.05.045
0022-5193/

The identification of biomarkers that can distinguish normal tissue from cancerous tissue is one of the main goals of cancer research. Much effort has been made to identify the causes of cancer, which can help our understanding of cancer and assist in the design of drugs for effective therapies. For example, comparisons within a single cancer type or a cancer family have revealed hidden oncogenic mechanisms that have been successfully applied to establish therapeutic strategies for individual types of cancer (Cowling and Cole, 2006; Pacal and Bremner, 2006; Contessa et al., 2002; Codegoni et al., 1998). Additionally, several bioinformatics methods have been developed and applied to detect differentially expressed genes between normal and cancer samples (Golub et al., 1999; Thomas et al., 2001; Han et al., 2002), and these genes have been determined to be related to cancers and can be used as biomarkers (Liu et al., 2012). Unfortunately, comparing data from single types of cancer has little success at recognizing compatible changes among the distinct types of malignant tumors (cancers) (Uramoto et al., 2006), and, many of the differentially expressed genes detected in one experiment do not always correspond to another experiment for the same cancer, especially for complex cancers (Liu et al., 2012). While complex cancers are usually caused by the dysregulation of functional networks that consist of a set of multi-regulated genes (Liu et al., 2012; Peltonen and McKusick, 2001; Glazier et al., 2002; Merikangas et al., 2006), how to identify a carcinogenic mechanism that is related to cancer development becomes an important research topic, and the first step is to discover common and specific features among various cancer tissues. Once we identify the different features, we can determine appropriate therapeutic targets and then target at risk patients for further medication and treatment (Rybaczyk et al., 2008).

Many systems biology approaches have been applied in various areas, such as reconstruction and design of genetic network (Lin et al., 2009; Lin and Chuang, 2010; Chuang and Lin, 2010; Wu et al., 2004; Lin et al., 2012a, 2012b, 2013; Wu, 2011; Yang and Wu, 2012; Wu and Li, 2008). Despite extensive research, there are still large gaps and controversies in the understanding of the fundamental molecular mechanisms of cancer in various tissues. Complex diseases are well-recognized by the deregulation of biological systems or molecular networks. Generally, molecule regulations and interactions vary at different times and in different tissues, while their changes are causally related to cancer progression. Therefore, molecules that are found to interact with different members of the molecular interaction network can be connected with the relevant cancer (Liu et al., 2012). Moreover, protein–protein interaction (PPI) networks also provide valuable information for understanding cellular functions and biological processes (Lin et al., 2009; Lee et al., 2006). Due to the substantial increase in human protein interaction data, PPI network analysis is used to unravel the molecular mechanisms of disease, and, in particular, to analyze cancer mechanisms (Kar et al., 2009; Liu et al., 2013; Liu et al., 2013; Wang and Chen, 2011). For example, the cancer-perturbed PPI networks of apoptosis shed light on the mechanism of cancer at the systems level and allow the identification of potential novel drug targets (Chu and Chen, 2008). Therefore, it is critical to identify relevant PPI networks to investigate cancer development processes. Recently, Wang et al. proposed a network approach through the construction of protein association networks to successfully shed light upon the pathways and mechanisms in the lung carcinogenic process and to provide potential therapeutic targets for novel drug design (Wang and Chen, 2011).

In this study, we analyzed various cancers, specifically, bladder, colorectal, liver, and lung cancers (Chen et al., 2012; Wang et al., 2012; Lin et al., 2010), through regression modeling, microarray data, maximum likelihood parameter estimation, and big database mining. Based on their PPI information and the gene expression data from normal and patient samples, two PPI networks with quantitative protein association abilities for cancer and the surrounding non-cancerous tissue of each cancer were constructed, respectively. The cancer and non-cancer PPI networks for each cancer were then compared with respect to their network structure and their protein association abilities for obtaining sets of significant proteins with large PPI variations from noncancer to cancer, which play important roles in carcinogenesis for each cancer type. The intersections of the four sets of significant proteins in each cancer were considered to be core network markers of these four cancers, while the distinctive significant proteins unique to one cancer type were considered to be a specific network marker for each cancer. Therefore, a network marker of each cancer can be divided into a core network marker and a specific network marker. We found that there were 28 significant proteins that were core network markers of bladder, colorectal, liver, and lung cancers; the correlated pathways and molecular mechanism are further discussed. These significant proteins are involved in cancer-related pathways and play important roles in carcinogenesis. Finding the core network marker is a novel approach to investigate the common network mechanism of carcinogenesis from the perspective of multiple cancers during the process of oncogenesis. After establishing the core network marker from each cancer network marker, the specific network marker of each cancer was determined. From the specific network markers of each cancer, we can investigate specific pathways or molecular mechanisms for gaining an insight into the specific carcinogenesis of each cancer. Furthermore, these analyses demonstrate the efficiency of our method and the validity of the identified network markers, which are consistent with results reported in the literature and further validate our results.

2. Materials and methods

2.1. Overview of the construction process of core and specific network markers

A flowchart representing the construction of core and specific network markers for carcinogenesis is shown in Fig. 1. In summary, three kinds of data, specifically, (1) microarray data of cancer and non-cancer samples of bladder, colorectal, liver, and lung cancers, and data from (2) the Gene Ontology database, and (3) the PPI databases, were required to construct the PPI network for each type of cancer and non-cancer. These data were used for PPI pool selection and then the selected PPIs and the microarray data were used for PPI network (PPIN) construction, resulting in a cancer PPIN (CPPIN) and a non-cancer PPIN (NPPIN) by regression modeling and the maximum likelihood parameter estimation method. The two constructed cancer and non-cancer PPINs were compared to obtain the sets of significant proteins with large PPI changes between two PPIs for each cancer based on the carcinogenesis relevance value (CRV) for each protein via the statistical assessment. From the intersection of these significant proteins, we were able to determine the core network markers of these four cancers. Distinctive from core network marker for each cancer, the specific network marker was obtained by extracting each unique network marker for each of the four cancers.

2.2. Data selection and preprocessing

The microarray gene expression datasets were obtained from the NCBI Gene Expression Omnibus (GEO). In this study, we chose four different types of cancer datasets as research objects,