Chapter 12
Incorporating Network Topology Improves Prediction of Protein Interaction Networks from Transcriptomic Data

Peter E. Larsen
Argonne National Laboratory, USA

Frank Collart
Argonne National Laboratory, USA

Yang Dai
University of Illinois at Chicago, USA

ABSTRACT

The reconstruction of protein-protein interaction (PPI) networks from high-throughput experimental data is one of the most challenging problems in bioinformatics. These biological networks have specific topologies defined by the functional and evolutionary relationships between the proteins and the physical limitations imposed on proteins interacting in the three-dimensional space. In this paper, the authors propose a novel approach for the identification of potential protein-protein interactions based on the integration of known PPI network topology and transcriptomic data. The proposed method, Function Restricted Value Neighborhood (FRV-N), was used to reconstruct PPI networks using an experimental data set consisting of 170 yeast microarray profiles. The results of this analysis demonstrate that incorporating knowledge of interactome topology improves the ability of transcriptome analysis to reconstruct interaction networks with a high degree of biological relevance.

DOI: 10.4018/978-1-4666-1785-8.ch012
INTRODUCTION

As the number of published descriptions of protein-protein interactions (PPIs) continues to grow, it becomes necessary to distill this increasingly vast number of facts into useful and biologically relevant associations. Although high-throughput methods for synthesizing and identifying large numbers of PPIs are beyond the reach of many research laboratories, profiling mRNA expression under various conditions is an effective and attainable approach to uncover underlying cellular machinery. In the context of this study, we define a PPI as one of two kinds of interactions: a direct physical interaction between proteins or two proteins that may or may not come into direct contact with one another but are parts of the same macromolecular complex. These types of interactions are distinct from 'functionally associated' proteins which may be co-expressed and working in the same or related biological process but do not require any form of physical interaction with one another for their function. It is well established that genes whose mRNA expression patterns are correlated across many diverse conditions can often be inferred to be functionally associated and these transcriptional co-expression patterns have been used in inferring physical protein interactions (Deane, Salwinski, Xenarios, & Eisenberg, 2002; Jansen et al., 2003). Although transcriptional data cannot consistently distinguish between direct protein binding and membership in a protein complex, strongly co-expressed mRNAs are more likely to indicate long-lived interactions (Ge, Liu, Church, & Vidal, 2001; Jansen, Greenbaum, & Gerstein, 2002; Simonis, Gonze, Orsi, van Helden, & Wodak, 2006). Recently, a supervised machine learning method (support vector machine) was proposed to predict physical PPIs from microarray gene expression data (Soong, Wrzeszczynski, & Rost, 2008). This method used gene expression profiles across 349 yeast microarray experiments to derive hidden factors that underlie the predicted PPI observations. The method was shown to recover more experimentally observed physical interactions than a conventional correlation-based approach. A recent augmentation of this approach used network topology to refine co-expression networks and uncover potential PPIs (Xulvi-Brunet & Li, 2009). The authors noted, however, that gene co-expression by itself does not necessarily correlate with known PPIs. This is not unexpected as for a given biological condition; the number of known interactions comprises only a small subset of all possible interactions likely to actually occur. In addition, only those interactions observed in the experimental condition and also differentially expressed in the context of a transcriptomic experiment have the potential to be detected. A PPI network then should be specifically associated with a biological condition and matched with differentially expressed relevant genes observed via the transcriptomic experiment. Several methods have been proposed to identify condition-specific biological networks or pathways from microarray data (Dittrich, Klau, Rosenwald, Dandekar, & Muller, 2008; Guo et al., 2007; Ideker, Ozier, Schwikowski, & Siegel, 2002) which map condition-specific interactions differentially expressed genes. However, the network structure of these methods is constrained to a pre-assembled PPI network and therefore they cannot discover novel interactions.

An ideal method for identifying PPI’s from transcriptomic data should have the ability to use the large and growing databases of known interactions, and also have the ability to identify novel interactions not previously reported in those databases. This is particularly important for those organisms with a minimal number of experimentally observed PPIs.

The method presented here is an extension of the Asymmetric-N method (Chen, Larsen, Almasri, & Dai, 2008), which constructs a co-expression network of genes through the analysis of microarray gene expression data. This method differs from other correlation-based methods such as, relevance networks (Butte, Tamayo, Slonim,
17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage: www.igi-global.com/chapter/incorporating-network-topology-improves-prediction/66712?camid=4v1

This title is available in InfoSci-Books, InfoSci-Medical, Healthcare, and Life Science and Technology, Communications, Social Science, and Healthcare. Recommend this product to your librarian: www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

Evidence-Based Combination of Weighted Classifiers Approach for Epileptic Seizure Detection using EEG Signals

www.igi-global.com/article/evidence-based-combination-weighted-classifiers/77929?camid=4v1a

Exerting Cost-Sensitive and Feature Creation Algorithms for Coronary Artery Disease Diagnosis

www.igi-global.com/article/exerting-cost-sensitive-feature-creation/74695?camid=4v1a

A Review on Brain Imaging Techniques for BCI Applications
Saugat Bhattacharyya, Anwesha Khasnobish, Poulami Ghosh, Ankita Mazumder and D. N. Tibarewala (2016). *Biomedical Image Analysis and Mining Techniques for Improved Health Outcomes* (pp. 39-70).

www.igi-global.com/chapter/a-review-on-brain-imaging-techniques-for-bci-applications/140484?camid=4v1a

Property and Personality Rights with Regard to Biobanks: A Layered System with Germany as an Example
Jürgen Robienski and Jürgen Simon (2014). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 16-23).

www.igi-global.com/article/property-and-personality-rights-with-regard-to-biobanks/105098?camid=4v1a