

# NetCAD: a network analysis tool for coronary artery disease-associated PPI network

Guomin Ren and Zhihua Liu\*

Department of Computational Biology and Bioinformatics, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences &amp; Peking Union Medical College, Beijing 100193, China

Associate Editor: Martin Bishop

## ABSTRACT

**Summary:** The systematic and unbiased charting of protein–protein interaction (PPI) networks relevant to health or diseases has become an important and burgeoning challenge in systems biology. Further, current reports have supported that good correlation exists between the topological properties and biological function of protein nodes in networks. Coronary artery disease (CAD, also called coronary heart disease) is the most common type of heart disease worldwide. Traditional approaches of studying individual gene or protein have shown their weakness in such complex disease. Here, we provide NetCAD, a web-based tool for systematic investigation of CAD-specific proteins in human PPI network. The features of NetCAD includes the following: proposing a novel method combining biological principles and graph theory, quantified topological analysis tools, build-in PPI information database consolidated from major public databases, creating CAD-associated subnetwork and visualizing network graph with good visual effects. NetCAD may provide important biological information for uncovering the molecular mechanisms and potential targets for therapies of CAD, which could not be found merely through molecular biology methods.

**Availability and implementation:** NetCAD is freely available at: <http://www.herbbol.org/netcad/>.

**Contact:** zhliu@implad.ac.cn or zhliu.liulab@foxmail.com

Received on October 30, 2012; revised on November 7, 2012; accepted on November 9, 2012

## 1 INTRODUCTION

Protein–protein interaction (PPI) is an important layer of connectivity between cell processes. The disturbance of PPIs may result in the disturbance of the cell process to which they contribute, compromising the cell viability or even leading to cell death. With the development of high-throughput screens, such as yeast two-hybrid assays, affinity purification coupled to mass spectrometry and synthetic-lethal and synexpression, great volume of PPI data has been yielded, forming complex PPI networks (PPINs) of varied species. These large networks provide a global view of cellular function and biological processes, which might bring biological and computational researchers new opportunities and challenges.

Recently, many studies indicated that the topology of PPINs is intimately related to biological functions and has potential applications, such as predicting disease-causing genes (Ideker and Sharan, 2008), annotating genes and proteins (Jiang *et al.*,

2011; Liu *et al.*, 2012), elucidating the mechanisms of diseases (Yang *et al.*, 2012) and identifying diseases biomarkers (Liu *et al.*, 2012; Xiong *et al.*, 2010).

As a leading cause of death and disability in western countries, coronary artery disease (CAD) has been the focus of the research for decades, and massive data have been accumulated such as candidate gene and single nucleotide polymorphism information. However, the progress is still limited even advanced genetic approaches are resorted to (Cohen, 2006). Noticing that graphic theories in PPINs may contribute to the research of complex diseases such as CAD, we have developed a web-based tool, NetCAD, for topological analysis and visualization of CAD-associated PPINs.

## 2 DATA SOURCES AND FUNCTIONALITY

For a comprehensive and unbiased study of CAD, NetCAD integrates 129 546 experimental PPI data of 19 480 human proteins from seven major public PPI databases including BioGRID, DIP, HPRD, IntAct, MINT, BIND and iRefWeb. By manually curating from literature and extracting the most reliable ones from the CADgene database (Liu *et al.*, 2011), 265 CAD-related genes were obtained. Then we mapped these genes into proteins and searched them in the PPI database we formerly built. Despite four proteins, which could not be found in the PPI database, 261 CAD-associated proteins were successfully mapped. Next, we searched all paths to find all connections between any two of these proteins directly and indirectly. Finally, we got a CAD-associated subgraph (subnetwork) of 261 nodes (CAD-associated proteins) with 2310 edges (326 direct interactions edges and 1984 indirect interactions edges).

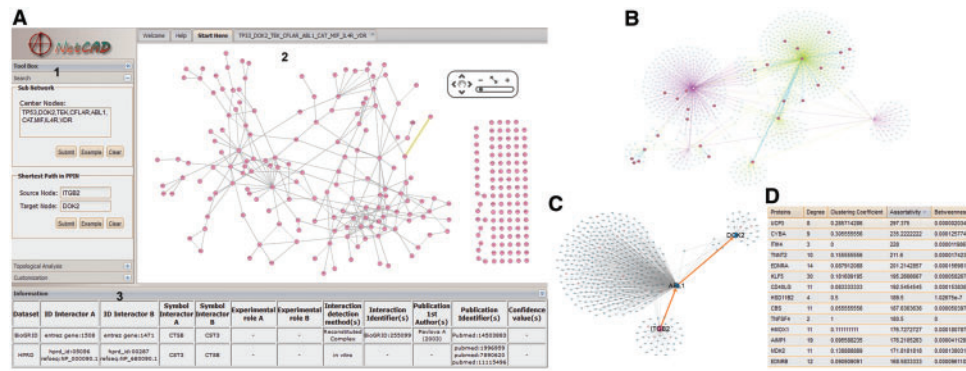
NetCAD implements topological analysis tools, which focus on this CAD-associated subgraph as a partial subgraph of the largest PPIN with 19 480 nodes (proteins) and 129 546 edges (PPIs). Currently, we offer the computation of six network measures to evaluate the significance of specified nodes.

**Degree:** The degree  $D_v$  of a node  $v$  is the number of edges adjacent to node  $v$ . Results might be biased when only node degrees were used to measure topology of PPINs, as some proteins are ‘hotspot’ and studied better than others, which lead to more PPI data correlated to these proteins.

**Clustering coefficient:** The clustering coefficient of a node quantifies how close its neighbors tend to cluster together. It is defined as:

$$C_v = \frac{2T_v}{D_v \times (D_v - 1)}$$

\*To whom correspondence should be addressed.



**Fig. 1.** NetCAD interface and three sample outputs. (A) NetCAD user interface: (i) tool box, (ii) network graph panel (iii) information box. (B) Subnetwork graph constructed with random selected nodes in NetCAD database, pink round nodes are CAD related, and blue diamond nodes are not CAD related. (C) A sample shortest path found by NetCAD from ITGB2 to DOK2. (D) Results generated by topological analysis tool in NetCAD are sorted in descending order according to the value of assortativity of each node

where  $T_v$  is the number of triangles through node  $v$  and  $D_v$  is the degree of  $v$ .

**Assortativity:** Assortativity  $A_v$ , or average neighbor degree of node  $v$ , measures the average degree of the neighborhood of node  $v$ . It is given by:

$$A_v = \frac{1}{|N_i|} \sum_{k \in N_i} D_k$$

where  $N_i$  are the neighbors of node  $i$  and  $D_k$  is the degree of node  $j$ , which belongs to  $N_i$ .

**Shortest path:** The shortest path  $SP_{v-t}$  is the minimum sum of weights path from source node  $s$  to target node  $t$  (PPIN is unweighted graph, which means all the edges are the same weight,  $SP_{v-t}$  here represents the one contains the minimum number of edges among all the paths from  $v$  to  $t$ ).

**Betweenness:** Betweenness  $B_v$  of a node  $v$  is the sum of the fraction of all-pairs shortest paths that pass through:

$$B_v = \sum_{s \neq t \neq v} \frac{\sigma_v(s, t)}{\sigma(s, t)}$$

where  $\sigma(s, t)$  is total number of shortest paths from node  $s$  to node  $t$  and  $\sigma_v(s, t)$  is the number of those paths that pass through  $v$ .

**Closeness:** The farness  $F_v$  of a node  $v$  is defined as the sum of its distances to all other nodes, and closeness is the inverse of the farness. It can be regarded as a measure of how long it takes to spread information from  $v$  to all other nodes sequentially. The classic closeness of PPIN should be 0, as the whole PPIN of human is not a connected graph. Thus, in NetCAD, we computed the closeness for each connected part separately then normalize to:

$$\frac{n-1}{size(G)-1}$$

where  $n$  is the number of nodes in the connected part of graph containing the node.

### 3 INTERFACE AND EXAMPLE

NetCAD has a friendly interface presented by a JavaScript framework Ext JS (Fig. 1A). The interactive network graph is

generated by Cytoscape Web (Lopes *et al.*, 2010). Edges and nodes are in the different shapes and colors according to their disease-associated attributes. Users can click on nodes or edges for detailed information, which will be shown in the box at the bottom of the browser window. Subnetworks can be constructed from the specific proteins lists which the users query. Topological tools will be available once any subnetwork is created. A detail demonstration of topological analysis can be accessed from the help tab. Three example outputs are shown in Figure 1B–D as the visualization of subnetwork, shortest path finding and the computation of topological attributes, respectively.

**Funding:** Basic Scientific Research Operation Cost of State-Leveled Public Welfare Scientific Research Courtyard (YZ-12-15), National Natural Science Foundation of China (81102746, 81100077), the Research Fund for the Doctoral Program of Higher Education (20111106120028), ‘Major Drug Discovery’ Major Science and Technology Research ‘12th Five-Year Plan’ (2012ZX09301-002-001-025), Program for New Century Excellent Talents in University (NCET, to Z.L.).

**Conflict of Interest:** none declared

### REFERENCES

- Cohen, J.C. (2006) Genetic approaches to coronary heart disease. *J. Am. Coll. Cardiol.*, **48**, A10–A14.
- Ideker, T. and Sharan, R. (2008) Protein networks in disease. *Genome Res.*, **18**, 644–652.
- Jiang, X. *et al.* (2011) Network-based auto-probit modeling for protein function prediction. *Biometrics*, **67**, 958–966.
- Liu, H. *et al.* (2011) CADgene: a comprehensive database for coronary artery disease genes. *Nucleic Acids Res.*, **39**, D991–D996.
- Liu, Z. *et al.* (2012) MiR-106b and MiR-15b modulate apoptosis and angiogenesis in myocardial infarction. *Cell. Physiol. Biochem.*, **29**, 851–862.
- Lopes, C.T. *et al.* (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.
- Xiong, J. *et al.* (2010) Protein-protein interaction reveals synergistic discrimination of cancer phenotype. *Cancer Inform.*, **9**, 61–66.
- Yang, D. *et al.* (2012) Ischemia/reperfusion-induced MKP-3 impairs endothelial NO formation via inactivation of ERK1/2 pathway. *PLoS ONE*, **7**, e42076.