CRRAO Advanced Institute of Mathematics, Statistics and Computer Science (AIMSCS)

Research Report

Author (s): **Thiyyagura Kranthi, Siddani Bhaskara Rao and Palanisamy Manimaran**

Title of the Report: **Identification of synthetic lethal pairs in biological systems through network information centrality**

Research Report No.: **RR2013-07**

Date: **June 3, 2013**

Prof. C R Rao Road, University of Hyderabad Campus,
Gachibowli, Hyderabad-500046, INDIA.
www.crraoaimscs.org

Identification of synthetic lethal pairs in biological systems through network information centrality

Thiyyagura Kranthi[1, 2], **Siddani Bhaskara Rao**[1] **and Palanisamy Manimaran**[1,*]

[1] C R Rao Advanced Institute of Mathematics, Statistics, and Computer Science, University of Hyderabad Campus, GachiBowli, Hyderabad - 500046, INDIA.

[2]Department of Biotechnology, School of Life Sciences, University of Hyderabad, Hyderabad - 500046, INDIA.

[*]**Corresponding author:** Palanisamy Manimaran, Assistant Professor, C R Rao Advanced Institute of Mathematics, Statistics, and Computer Science, University of Hyderabad Campus, GachiBowli, Hyderabad - 500046, INDIA.

Phone: +91-9989952362, **Fax:** +91-40-23013118,

Email: maran@crraoaimscs.res.in, pa.manimaran@gmail.com

Keywords: Information centrality, Protein-protein interactions, Synthetic lethality.

Total no. of words: 3,396

Abstract

The immense availability of the protein interaction data, provided with an abstract network approach is valuable for improved interpretation of biological process and protein functions globally. The connectivity of a protein and its structure is related to its functional properties. The highly connected proteins are often functionally important and knockout of such proteins leads to lethality. In this paper, we make use of a graph information centrality concept for the identification of synthetic lethal pairs in biological systems. With the advent of molecular networks comprising of normal, cancerous cells, we apply graph centrality based method on Human cancer protein interaction network. Our approach effectively predicts the potential lethal pairs, which were analogous to the experimental and computational inferences.

1 Introduction

Majority of the systems either available in nature or manmade are complex. Understanding these complex systems requires a bottom up approach i.e. breaking the system into small elementary constituents. Mapping out the interactions between these components can be characterized as network. In terms of network, a system is a modeled graph where the nodes are the elementary constituents and the edges represent the interactions between them. Networks enable a simple and uniform representation of complex structures, processes and finds wide range of applications in various fields such as social, physical and biological sciences [1-5]. Structural analysis of networks can lead to new insights into complex systems and can be studied through the standard models like random, small world, scale free network [6-8]. The topological properties such as degree distribution, clustering coefficients, centrality measures, community structures etc. of the network models help us to understand the functional properties through their structure [9, 11].

The graph centrality measure concept plays a vital role in identifying the potential nodes that are functionally important in a network. In the recent past, various centrality measures such as degree, closeness, betweenness, Eigen vector, information centrality etc. have been developed for predicting the potentiality of a node [12-14].

The above mentioned centrality measures aids us in analyzing the various underlying process especially of biological networks and also identifies the key players in biological processes. A correlation between a node's structural importance in the protein-protein interaction network and its functional importance commonly referred as centrality-lethality rule is well understood using centrality concepts [15]. In a protein interaction network, the connectivity of a protein is related to its functional properties. As evident from the earlier studies, the proteins with the high degrees and high centralities are found to be highly essential in the network. Knock out of such essential proteins from the genome is more likely lethal to the organism [16-18].

Synthetic lethality (SL) which may be viewed as an extension of essentiality thus can be clearly established using the network centrality concepts. Two genes are said to be synthetically lethal, if mutation of either gene alone is compatible with viability but simultaneous mutation of both leads to death [19-22]. A large number of studies have been carried out recently in the context of synthetic lethality for characterizing the functions of genes in various organisms such as Saccharomyces cerevisiae, Zebra fish , Drosophila [23-26] etc. and also for the development of drugs for different diseases like Cancer HIV , Mycobacterium tuberculosis [27-29] etc. Many experimental and computational methods have been emerged out for the identification of SL pairs. The experimental techniques include Kinzler method [30], Cannani method [31], Synthetic lethal screens, unbiased chemical and genetic screens [19] etc. The limitations of the

experimental methods urge for development of computational techniques such as decision trees machine learning methods , Simulation of double gene knock-down and assigning each pair a synergy score [32-34].Synthetic lethal interactions though identified are difficult to rationalize; hence their study is restricted to model organisms such as yeast and bacteria [35]. Lack of conservedness in the model organisms provoked the researchers to study the SL pairs in humans.

In this paper, we make use of the theoretical graph centrality measure i.e., information centrality for the identification of lethal pairs [13]. Here the information centrality concept is used to quantify the relevance of pair of the nodes in the network. Knocking out a pair of nodes considerably affects the system as in like the synthetic lethality concept i.e. mutation of both the genes leads to death. We apply this procedure on the Human cancer protein interaction network (HCPIN) to identify the lethal pairs for enhancing the personalized cancer therapy. Cancer, a dreadful disease is the result of gain of functional mutation of oncogene and loss of functional mutation of tumor suppressor genes [36]. Developing drugs that could selectively kill cancerous cells without affecting the normal cells remains a considerable challenge. In this context, synthetic lethality succors for the development of drugs that could theoretically target cancerous cells imposed due to loss and gain of function mutations, while sparing the cells with normal copy of mutated cancer relevant genes.

2 Methods

2.1 Data Collection

For data collection and construction of HCPIN we have followed the procedure adopted by Gozde et al [37]. The Human proteinóprotein interaction (HPPI) data was collected from the Human Protein Reference Database (HPRD) [38]. It consists of 9617 proteins with 39240 interactions. After the removal of self-interactions, palindrome interactions and repeated interactions the HPPI network consists of 9454 proteins with 36867 interactions. A list of 291 cancer genes was collected from the comprehensive census of cancer gene database provided by Futreal and his co-workers [39]. In addition three sets of cancer genes consisting 873 tumor suppressor genes,495 oncogenes and 1023 stability genes were collected CancerGene database [40]. The tumor, onco and stability gene sets have some of the genes in common and the redundant genes of the three sets were omitted, unified to a gene set which then consists of 1927 genes. The list of 291 cancer genes when combined with 1927 gene set resulted in a unique cancer gene list consisting of 2218 cancer genes.

2.2 Construction of Human cancer protein interaction network

By mapping the generated cancer gene list on to the HPPI data, only those interactions between the cancer genes were considered. Thus we have obtained the HCPIN, from which all the orphan nodes were removed and the giant component was considered for further analysis. The resultant core HCPIN consists of 1539 proteins and 6471 interactions.

2.3 Identification of synthetic lethal pairs using information centrality

The graph information centrality measures the potential of an individual node based on the change in efficiency of the network observed after its knockout [13]. Network analysis of biological systems predicts that knocking out a pair of genes have a significant effect on the system compared to knocking out a single gene. This approach correlates with the concept of synthetic lethality as the lethal pairs are those in which mutation of alone is compatible with viability but mutation of both leads to death. Thus, to obtain the synthetic lethal pairs, we have modified the above centrality measure by knocking out a pair of nodes.

If G is the graph representing the network, and G´ is the graph after the removal of the nodes ømø and ÷n', then the information centrality ÷$C_{m,n}$ø for the knocked out nodes is given as

$$C_{\,,} = \frac{\Delta}{\;} = \frac{[\;']}{[\;]} \qquad \text{------ ( )}$$

Where E is the efficiency of the graph, which is a measure of robustness. This quantity is based on the assumption that the information/signaling in a network travels along the shortest paths and it is defined for a graph G as follows:

$$E\ G = \frac{\;}{(\quad)}\sum_{\in} \frac{\;}{\;} \qquad \text{------ ( )}$$

Here ÷d_{ij}ø refers the shortest path between the nodes ÷iø and ÷jø, where ÷Nø represents the total number of nodes in the network.

The detailed procedure for identifying and ranking of the synthetic lethal pairs is as follows:

1. Calculate the overall efficiency E [G] of the network.

2. Knock-out a pair of nodes at random and calculate the efficiency of the network E [G].

3. Calculate the drop in the efficiency of network using equation (i).

4. Repeat from step (ii) for all the possible pair of nodes.

Using the above procedure we have ranked the lethal pairs for HCPIN.

3 Results

3.1 Human cancer protein interaction network

The HCPIN, constructed by mapping the cancer genes on HPPI data consists of 1539 proteins with 6471 interactions. We found that degree distribution of HCPIN possess power-law behavior which is evident from its degree scaling exponent 1.71. The other topological properties such as average degree 23.67, average clustering coefficient 0.1418, and assortivity coefficient 0.4706 also were calculated. The diameter and the efficiency of the network were found to be 10 and 0.2952.

3.2 Synthetic lethality in HCPIN

Using our approach we have calculated the information centrality scores for all the possible pairs. We have ranked these pairs based on their scores i.e. in the order of the effect produced by them on the network by their knock out. The obtained pairs showed a significant variance in drop in their efficiency and the same is clearly represented in the figure 1.The ranked gene pairs that showed a transition in the centrality scores were enumerated and found to be nearly 100 and these pairs were considered for our further analysis. We have observed that all the pairs were found to impose a profound effect on the network when knocked out as a pair satisfying the

paradigm of synthetic lethality. Interestingly, the l00 pairs were composed of tumor and oncogene as the interacting partners. Hence, we consider the above 100 pairs to be synthetically lethal/sick. The list of these top 100 SL pairs and the type of cancers associated with them was provided in the **supplementary information 1 and 2** respectively.

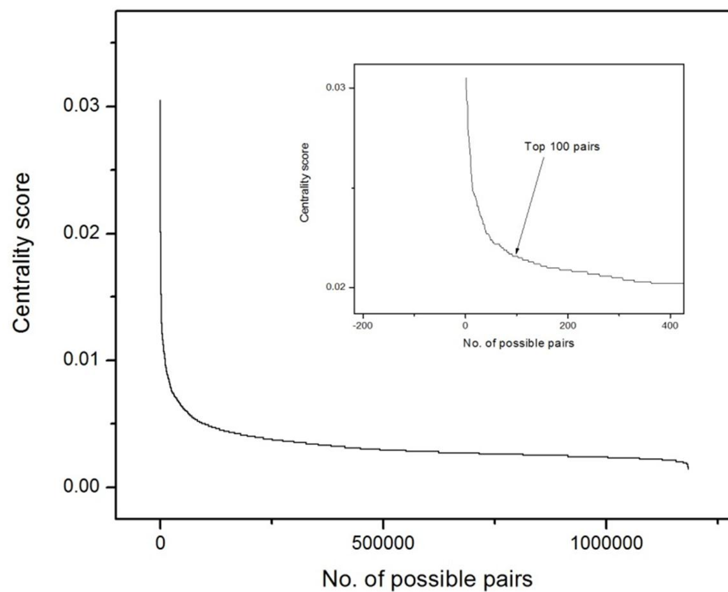

Figure 1: The main graph depicts the ranked centrality scores for all the possible pairs of HCPIN. The inset graph shows the transition in the centrality score near top 100^{th}pair and these pairs were considered for our analysis.

3.3 Analysis of top 100 lethal pairs

The analysis of top 100 lethal pairs of HCPIN yielded some alluring information; we have found out that 97 pairs possess the tumor suppressor protein, ÷TP53ø as an interacting partner with the cancer relevant gene. We have observed that there are many possible shortest paths for a given

lethal pair with a maximum path length of 3. Out of 100 lethal pairs 33 pairs were found to have a direct interaction with its interacting partner. These findings help us to have detailed information about the interacting partners between the pairs during drug development. The list and the number of possible shortest paths for the 100 SL pairs, with their path lengths were provided in **supplementary information 3**.

4. Discussion

With the recent improvements in network biology, the structure of protein interaction networks themselves provide valuable information and hence can be used for the identification of SL pairs. From our HCPIN analysis we have observed that both the genes of the top 100 lethal pairs, were found to participate in the same/compensatory pathways as is evident from their ability to cause same type of cancer, thus satisfying the conditions of synthetic lethality. The pharmacological approach to treat cancer till date is unclear, because of the induced functional mutation of tumor suppressor gene by drugs. Targeting the protein products that are in synthetic lethal interaction with non-druggable cancer causing loss of function mutations provides an elegant solution to the problem.

From the literature survey itøs clear that TP53 is mutated in 50% of human tumors, therefore synthetic lethality in the context of TP53 loss of function mutations is an important aspect for anticancer therapy. This strengthens our results, as 97 0f 100 top lethal pairs consist of TP53 as one of the interacting partner with the cancer related genes. Our approach was able to identify some experimentally and computationally identified SL interactions of TP53 with BRCA1 [41], CDKNA1 (PAK3) [42], and CDKN2A (P19ARF) [43] and MET [44] which are present among top 100 lethal pairs. Also we have tried to corroborate through literature survey for validation of

our approach. The lethal pairs obtained through our method were found to quench the concepts of synthetic lethality and some of them were annotated for their lethality as follows

CTNNB1-TP53: TP53 mutation is frequently associated with CTNNB1 mutation. Accumulation of catenin as a result of CTNNB1 mutation leads to high levels of wild type P53 which becomes favorable for tumor cells deficient for functional P53.Thus the gain and loss of function mutations of both TP53 and CTNNB1 together eventually leads to tumor progression [45]. This implies that CTNNB1 and TP53 can be essentially considered as lethal pair which can be effectively targeted in case of Medullo Blastoma anti-cancer therapy, since mutation of either gene alone doesnøt contribute for carcinogenesis.

YWHAG-TP53: YWHAG (14-3-3 gamma) was recently found to have oncogenic activity and is observed to be over expressed in Human lung cancers. The 14-3-3 gamma negatively regulates the P53 protein that results in loss of P53 function eventually over-expressing YWHAG [46]. Thus the gain of function mutation of YWHAG coupled with the loss of function mutation of TP53 promotes lung carcinoma indicating that afore mentioned pair can be considered as synthetically sick/lethal.

EP300-TP53: EP300 modulates P53 pathway at multiple levels. Mutations of P300 could disrupt P53 activation, stability and cell cycle arrest thereby promoting carcinogenesis. Mutations in P300 and P53 are not mutually exclusive, suggesting that mutation of P300 doesnøt abate selective pressure of P53 mutation [47].Hence P300 and P53 with their loss and gain of functional mutations can be regarded as Potential synthetic targets for anti-cancer therapy.

AR-TP53: The loss of function mutation of TP53 leads to over expression of AR through its gain of function mutation. Thus the positive feedback regulation of AR produces low levels of

Androgen and thereby braces the Prostate cancer progression [48, 49].The AR-P53 lethal pair can be used as potential markers for prostate cancer. Similarly we have observed that the down regulation of BCL2 gene expression because of the loss of function mutation of tumor suppressor protein P53and also provokes tumor genesis making the pair essentially lethal [50]. The above annotations prove that the pairs obtained through our method are lethal. We were also able to identify other potential SL interactions like CTNNB1, YWHAG, EP300, CSNK2A1, and SMAD3 with TP53 which were present in top 5 of the 100 synthetic lethal pairs. The list of other potential lethal pairs was provided in the **supplementary information 1** which can be further validated by any of the validation approaches described below.

Analysis of the expression levels of the intermediary proteins of the lethal pairs helps in further validation of lethal pairs. The up regulation of all the intermediary nodes which ensue in the shortest paths of lethal pairs (provided in supplementary information 3) indicates the un interrupted flow of information and contributes to carcinogenesis which provided with mutation information of the pair proves the lethality. Since carcinogenesis is mainly linked with the cell cycle progression one can also validate the lethal pairs by considering the list of all the genes involved in cell cycle progression, tumor and onco gene pairs and mapping out the interactions for the three sets of genes. The analysis of the pathways in which the three gene sets are involved helps in commuting further the validation process.

Our study suggests that, one can specifically kill cancerous cells by targeting the cells consisting of a mutant TP53 associated with an oncogene. These circumstances may lead to the development of personalized cancer therapy. The approach we employed using the network centrality measure was quite efficient for predicting the SL pairs in HCPIN. This sheds light on the importance of the network analysis methods in understanding the physiological and

functional process underlying various biological systems and also for the drug target identification for various diseases. Hence, one can use the network based approach in identifying the lethal pairs for any other complex systems.

5. **References**

1. S. Wasserman and K. Faust, Cambridge University Press, 1994.
2. J. Scott, Sage 2nd edition, 2000.
3. S.H. Strogatz, *Nature*, 2001, 410, 268-276.
4. S.N. Dorogovtesev and J.F.F. Mendes, Oxford University Press, 2003.
5. R. Albert and A.L. Barabasi, *Rev. Mod. Phys*, 2002, 74, 47.
6. P. Erdos and A. Renyi, Publ. Math. Debrecent, 1959, 6,290-297.
7. D.J. Watts and S.H. Strogatz, *Nature*, 1998, 393, 440ó442.
8. A.L. Barabasi and R. Albert, *Science*, 1999, 286, 509ó511.
9. P. Holme, M. Huss and H. Jeong, *Bioinformatics* 2003, 19, 532ó538.
10. S. Wuchty and P.F. Stadler, *J. Theor Biol*, 2003, **223**, 45ó53.
11. S.R. Hegde, P. Manimaran and S.C. Mande, *PLoS Comput Biol*, 2008, **4**, e1000237.
12. P. Manimaran, S.R. Hegde and S.C. Mande, *Mol BioSyst,* 2009 **5**, 1936.
13. D. Koschützki and F. Schreiber, *Proc. German Conf Bioinformatics*, 2004, **53**,199ó206.
14. V.Latora and M. Marchiori*, New J Phys*, 2007, **9**, 188.
15. H. Jeong, S.P. Mason, A.L. Barabasi and Z.N. Oltvai, *Nature*, 2001,**411**, 41ó42
16. H. Yu, D. Greenbaum, L.H. Xin, and X. Zhu, *Trends Genet*, 2004, **20**, 227ó231.
17. M.W. Hahn and A.D. Kern, *Mol Biol Evol*, 2005, **22**, 803ó806.
18. S.Wuchty, *Proteomics*, 2002,**2**, 1715ó1723.
19. W.G.Jr. Kaelin, *Nature*, 2005, **5,** 689-698.
20. Th. Dobzhansky, *Genetics*, 1946, **31**, 269-290.
21. J.C. Lucchesi, *Genetics*, 1968**, 59**, 37-44.
22. L.H. Hartwell, P. Szankasi and C.J. Roberts, *Science*, 1997, **278**, 1064-1068.
23. S.L.Ooi, X. Pan and B.D. Peyser, *Trends Genet*, 2005, **22**, 56-63.
24. P.Ye, B.D. Peyser and X.Pan, *Mol Syst Biol*, 2005, **1**, p0026.
25. V.A. Hajeri, and J.F. Amatruda, *Dis Model Mech*, 2012, **5** 3337.

26. C.L. Tucker and F. Stanley, *Nat Gent,* 2003, **35**, 204-205.
27. C. Scholl, S. Frohling, I.F Dunn and A.C. Schinzel, *Cell,* 2009, **137**, 821-834.
28. S. Brouillet, T. Valere, E. Ollivier and L. Marsan, *Biology Direct*, 2010, **5**, 40.
29. R. Kalscheuer, K. Syson, U. Veeraraghavan and B.Weinrick, *Nat Chem Biol*, 2010, **6**, 376384.
30. C.J. Torrance, V. Agrawal, B. Vogelstein and DK.W. Kinzler ,*Nature Biotechnol,* 2001, **19**, 940-945.
31. A. Simons, N. Dafni, I. Dotan and Y. Oron, *Genome Res,* 2001, **11**, 266-273
32. O. Folger, L. Jerby, C. Frezza and E. Gottlieb, *Mol Syst Biol* ,2011, **7**, 501-511.
33. S.R. Paladugu, S. Zhao, A. Ray and A.Raval, *BMC Bioinformatics,* 2008, **9**, 426.
34. D. Hanahan and R.A. Weinberg, *Cell*, 2000, **100**, 57-70.
35. W.G.Jr. Kaelin, *Genome Med*, 2009, **1**, 99
36. J.L. Hartman, B. GarvikandL. Hartwell, *Science*, 2001, **291**, 1001-1004.
37. G. Kar, A. Gursoy and O. Keskin, *PLoS Comput Biol*, 2009, **5**, e1000601.
38. T.S. Keshava Prasad, R. Goel, K. Kandasamy and S.Keerthikumar, *Nucleic Acids Res*, 2009, **37**, 767-72.
39. P.A. Futreal, L. Coin, M. Marshall and T. Down, *Nat Rev Cancer*, 2004, **4**, 177-83.
40. E.H. Maureen, C. Martine, E.M. John and S. Chris, *Nucleic Acids Res,* 2006, **35**, D721-D726.
41. S.R. Bartz, Z. Zhang, J. Burchard and M.Imakura, *Mol Cell Biol*, 2006, **26**, 9377-9386.
42. A. Baldwin, D.A. Grueneberg, K. Hellner and J. Sawyer, *PNAS*, 2010, **107**, 12463-12468.
43. K. Christopher, M. Russell, K. Michael and X. Chang,. EMBO Reports 2009, **10**, 87.
44. K.D. Sullivan, N. Padilla, R.E. Henry and C.C. Porter, *Nat ChemBiol*, 2012, **8**, 646-54.
45. P. Elke, M. Remke and D. Sturm, *J Clin Oncol*, 2010, **28**, 5188-5196.
46. V.M. Radhakrishnan, C.W. PutnamandW. Qi, *BMC Cancer,* 2011, **11**,378.
47. N.G. Iyer, H.Ozdag and C.Caldas, *Oncogene*, 2004, **23**, 422564231
48. M. Castagnaro, D.W. Yandell and C. Poremba, *Verh Dtsch Ges Pathol*, 1993, **77**,119-23.
49. A.Vellaichamy, A. Sreekumar, T. Rajendiran and J. Yu, *PLoS One*, 2009, **4,** e7075.
50. T. Miyashita, M. Harigai, M. Hanada and J.C. Reed, *Cancer Res*,1994, **54(12)**, 3131-5.

5. Acknowledgement

The authors would like to thank the Department of Science and Technology for their financial support (DST-CMS GoI Project No. SR/S4/MS: 516/07 Dated 21.04.2008).

6. Conflicts of Interest

The authors declare that they have no conflict of interest.