Graph centrality of BRCA1 in Protein-Protein Interaction network: An Overview

[1]Sminu Izudheen, [2]Sheena Mathew

[1] Department of Computer Science & Engineering, Rajagiri School of Engineering & Technology, Kakkanad, India, [2] Division of Computer Science, School of Engineering, Cochin University of Science & Engineering, Kerala, India
[1]sminu_i@rajagiritech.edu.in,[2]sheenamathew@cusat.ac.in

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 4 June 2021

Abstract— Cancer is a leading cause of death and around 10 million people died in 2020. In terms of new cases reported in 2020, leading was breast cancer with 2.26 billion cases. In most cases, reason for the disease is stated as a germline mutation on tumor suppressor gene, BRCA1. Even though dysfunction of BRCA1 is the reason for cancer development, the pathway determining this malignant cellular transformation is poorly defined. Identifying the mechanism behind functional loss of this proteins is one of the major challenges faced by cancer biologists. This paper makes an attempt to characterize the behavior of BRCA1 by identifying its topological properties in a Protein-Protein Interaction Network (PIN). .

Index Terms—*Biological networks, Protein-protein interaction, Cancer protein identification, BRCA1*

I. INTRODUCTION

Cancer is a disease caused by uncontrolled cell growth. According to World Health Organization, in 2020 it has taken 10 million lives. Around 2.26 million breast cancer cases were reported in 2020 which is the largest when compared with other cancer types. However, breast cancer mortality rate can be reduced if diagnosed and treated early. Mutation in BRCA1 is sited as the reason for fifty percent cases in breast and ovarian cancers [1]. These mutations can also cause fallopian tube cancer, prostate cancers, leukemias and lymphomas.

BRCA1 is a complex protein made with 1863 amino acids. It when interacts with its RING domain (BARD1) *in vivo*[2] forms heterodimers. BRCA1 gets its tumor suppression ability from these hetrodimers[3]. BRCA1's interaction with RAD51 activates p21 which contribute to cell-cycle arrest and growth suppression. DNA double-strand breaks (DSBs) when compared with other DNA damage is considered to be the most dangerous as it causes cell death. To repair DSBs, cells make use of homologous recombination (HR) and nonhomologous recombination (NHR) [4,5]. BRCA1 maintains genomic stability by regulating HR and NHR [6].

Even though vital the dysfunction of BRCA1 is the reason for cancer development. ER-negative and EGFR-positive breast cancers are caused by alteration in BRCA1-mutant Primary Mammary Epithelial Cells (PMEC)[7]. In this paper the topological behavior of BRCA1 in a PIN is discussed. It was observed that, compared with non-cancer proteins, BRCA1 exhibit entirely different topological behavior.

II. LITERATURE REVIEW

Identifying genes associated with a specific phenotype is a key step towards understanding complex disease mechanisms and thereby developing targeted diagnostic and therapeutic interventions[8]. Role of protein-protein interaction at the N-terminal region of BRCA1 in cancer risk classification is discussed[9]. Protein inhibitors of BRCA1 were identified from small molecule microarrays [10]. A study on BRCA1's direct interaction with DNA is presented [11]. Vadiraj Kurdekar et.al.[12] reports effect of small-molecule inhibitor Bractoppin on BRCA1 tBRCT domain.

Researchers were mainly concentrated on identifying mutations in protein-protein interactions of BRCA1 based on its physical properties. Very few works have addressed the topological behavior of BRCA1 in a protein-protein interaction network. This paper provides a detailed discussion on behavior of BRCA1, especially the graph centralities of BRCA1 in PIN.

III. METHODS AND DATA

A. Data

Protein-protein interaction for the proposed work was downloaded from Human Protein Reference Database (HPRD) [13]. After removing redundancy and self interactions there were 39,240 protein interactions pairs containing 7932 proteins. From this the subset of interactions of non-cancer proteins and BRCA1 protein are generated. The subset contains 7102 interactions of 1137 non-cancer proteins, BRCA1 and 1097 cancer proteins.

B. Grpah Centrality of Proteins in PIN

A protein's significance in a PIN can be estimated through its various centrality values in the network. Proteins are ranked based on these four centrality measures. Various centrality measures used in the proposed method are listed below.

Degree Centrality (DC): Number of edges incident upon the node is given by its degree centrality. It represents the count of immediate neighbours of the node[14] as given in Equation 1.

$$DC(i) = \sum_{j=1}^{n} A_{ij} \qquad (1)$$

Average Shortest path (SP): How well a node is connected in the network can be estimated though its average shortest path distance to other nodes as given in Equation 2.

$$SP(i) = \frac{1}{n} \sum_{j=1}^{n} p_{ij} \qquad (2)$$

where p_{ij} denotes the shortest path from the node to any other node in the network.

Betweenness Centrality (BC): It is the aggregate of shortest paths passing through that node [15]. It shows the efficiency of a node to act as a bridge between two other nodes. BC of a node can be calculated as given in Equation 3.

$$BC(i) = \sum_{s \neq i \neq t} \frac{\sigma_{s,t}(i)}{\sigma_{s,t}} \qquad (3)$$

$\sigma_{s,t}(i)$ is the shortest path from node *s* to node *t* passing through node *i*, where *s* and *t* are two other nodes in the network.

Clustering Coefficient: Clustering coefficient of a node gives an insight on how connected a nodes' neighbours are among themselves. It is the count of triangles in the network as in Equation 4.

$$CC(i) = \sum_{j,k \in N} \frac{2.A_{ij}.A_{jk}.A_{ki}}{l_i.(l_i - 1)} \qquad (4)$$

where l_i is the count of edges possible to *i*.

IV. RESULTS AND DISCUSSION

From the protein-protein interaction network various centrality values were calculated. Proteins are ranked according to the centrality values, expecting BRCA1 protein to hold top rank in the list. Some of the observations noted are discussed below.

A. BRCA1 has higher degree

One of the most elementary characteristics of a network is the degree/connectivity. First parameter considered was the connectivity or degree of a protein in a PIN. The degree of BRCA1 was 106 which is far higher when compared with non-cancer protein with an average degree 3.88. Detailed plot showing the degree distribution of 1137 non-cancer proteins is given in Figure 1. From the figure, there are 1082 proteins whose degree is less than 26, of which 749 had degree 1. It may be noted that only 4 (0.35 percent) non-cancer proteins have there degree greater than BRCA1.From the result it is clear that when compared with non-cancer proteins, proteins encoded with BRCA1 proteins interacts strongly with other proteins and shows higher connectivity.

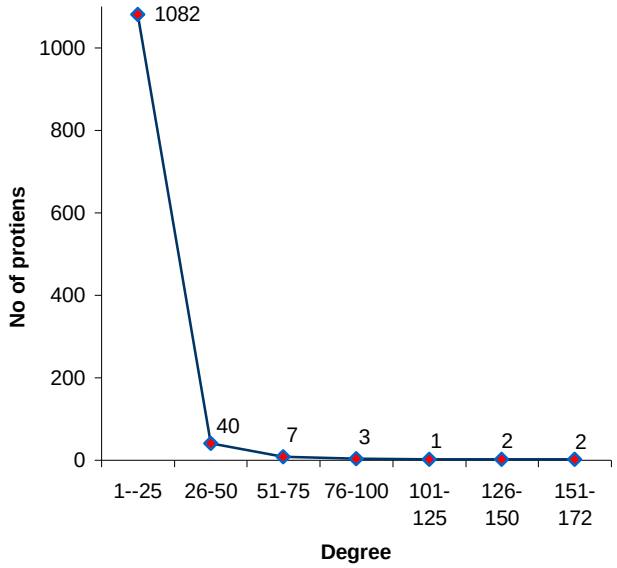

Fig. 1. Degree distribution of non cancer proteins

B. BRCA1 has lower shortest path distance

Next parameter considered was shortest path distance to other proteins in the network. Average shortest path distance of BRCA1 was only 3, while on an average a non-cancer protein has shortest path distance as large as 4.07. To have a more detailed view of shortest path distance; proteins are separated by their average shortest path to other nodes and examined their shortest path distribution, which is given in Figure 2. It may be noted from the graph that all non cancer proteins have their shortest path distance either greater than or equal as that of BRCA1, of which 1042 (91.65 percent) have greater shortest path than BRCA1. The graph again asserts that BRCA1 interacts stronger in protein interaction network than non-cancer proteins.

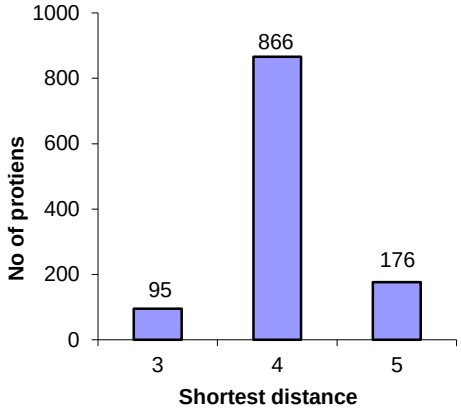

Fig. 2. Shortest path distance distribution of non cancer proteins

C. BRCA1 has higher betweenness centrality

In a biological network the significance of Betweenness centrality is that it reflects how many signals pass through the node. When calculated the betweenness value for all proteins, the average betweenness value of non-cancer proteins reported by the method was 2586.94 which is lesser than BRCA1 with betweenness 7533. Detailed plot showing the betweenness distribution of 1137 non-cancer proteins is given in Figure 3. From the graph only 2 (0.18 percent) non-cancer protein is having betweenness value greater than BRCA1. Hence the result suggest that more signals pass through BRCA1 than non-cancer proteins.

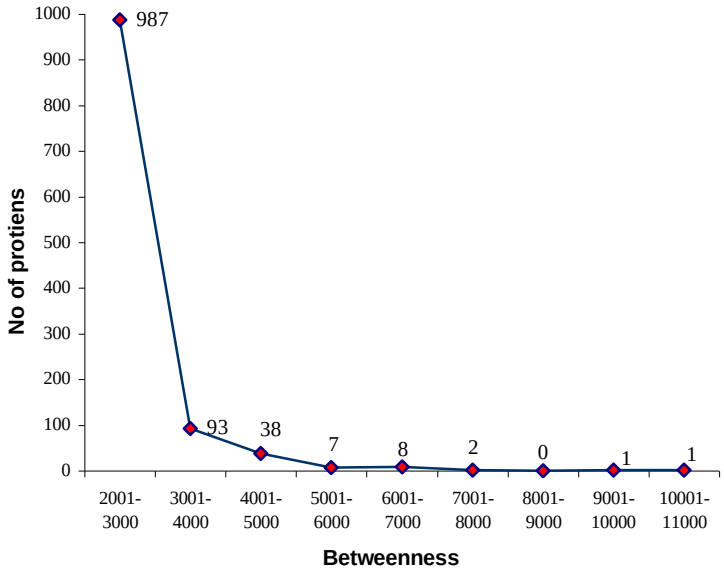

Fig. 3. Betweenness distribution of non cancer proteins

D. BRCA1 has lower clustering coefficient

Clustering coefficient of a node shows how well connected a node is with its neighbours. Results from the method shows that clustering coefficient of BRCA1 is much lesser than average clustering coefficient of non cancer proteins. Exempting the nodes with clustering coefficient zero, average clustering coefficient of the non-cancer proteins was $3.78e^{-1}$, which is greater than that of BRCA1 ($1.44e^{-3}$). Detailed plot showing the clustering coefficient of non-cancer proteins is given in Figure 4. Again note from the graph that only 5 (3.6 percent) of non-cancer protein have their clustering coefficient lesser than BRCA1.

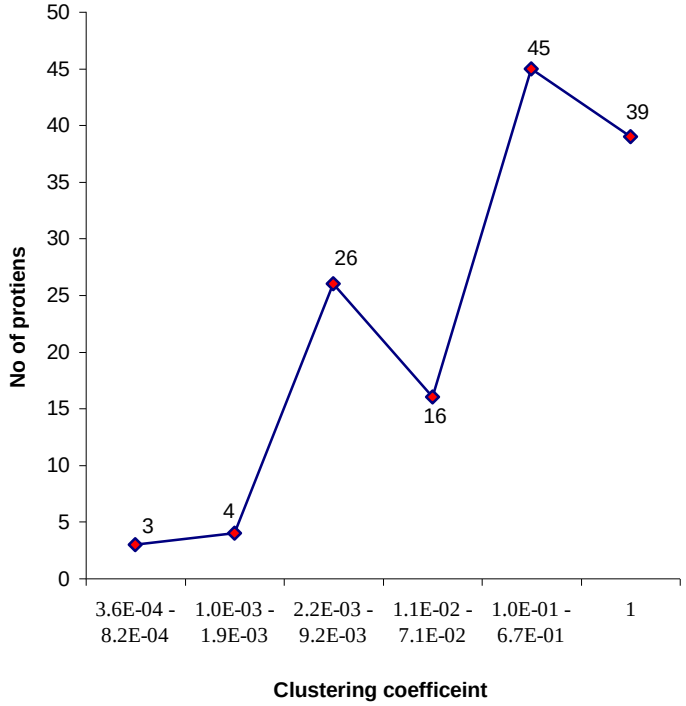

Figure 4. Clustering Coefficient of non cancer proteins

The significance of the results obtained was statistically verified through t-test with a significance level of 10% (i.e. 90% confidence). The null hypothesis (H_0) selected for the analysis is given below.

$H_0 : \mu_{BRCA1} - \mu_{non\ cancer} = 0$

i.e., we assume that for any centrality measure there is no difference between centrality value of BRCA1 and average non cancer proteins. Table 1 summarizes the t-test for degree centrality, shortest path, betweenness centrality and clustering coefficient.

Degree and betweenness centrality were tested using right-tailed test and shortest path and clustering coefficient were tested using left-tailed test. The result of t-test for various centrality parameters is given in Table 1.
We performed right-tailed test for degree and betweenness centrality with a critical point 1.28 at 10% level of significance. It may be noted that the value for degree centrality and betweenness centrality are 1.831 and 6.309 respectively. Since the value is larger than the critical point, the null hypothesis is rejected with 97 and 99 percent confidence respectively. Left-tailed test with critical point -1.28 was performed at 10% level of significance for shortest path and clustering coefficient. As the values obtained were much less than the table values, null hypothesis for shortest path and clustering coefficient were rejected the 91, 96 and 96 percent respectively. Hence we conclude that degree and betweenness centrality of BRCA1 is greater than non-cancer proteins. But shortest path and clustering coefficient of BRCA1 is lower than non-cancer proteins.

Table 1. *t*-test on centrality

Parameter	**t-statistic**	**Confidence %**
$degree_{BRCA1}$ Vs. $degree_{non\ cancer}$	1.831	97
BC_{BRCA1} Vs. $BC_{non\ cancer}$	6.309	99
SP_{BRCA1} Vs. $SP_{non\ cancer}$	-1.3	91
CC_{cancer} Vs. $CC_{non\ cancer}$	-1.738	96

CONCLUSION

Mutation in BRCA1 is the reason for about fifty percent cases breast and ovarian cancers. But the pathway determining this malignant cellular transformation is poorly defined. As an aid to this, the paper presents a comparison between different graph centralities of BRCA1 protein over non cancer proteins in a PIN. Result demonstrate that when compared with non-cancer proteins BRCA1 exhibits an entirely different network signature, interacting strongly and showing high tendency to form hubs in a protein-protein interaction network.

REFERENCES

[1] Rahman, N., M. R. Stratton**,** 1998. "The genetics of breast cancer susceptibility," Annu. Rev. Genet, Vol. 32, pp**:**95–121, 1998.

[2] Wu, L. C., Wang, Z. W., Tsan, J. T., Spillman, M. A., Phung, A., Xu, X. L., Yang, M.-C. W., Hwang, L.-Y., Bowcock, A. M., and Baer, R., "Identification of a RING protein that can interact *in vivo* with the *BRCA1* gene product," Nature Genetics, *Vol.*14, pp**:**430–440, 1996.

[3] Jose E. Meza, Peter S.Brzovic, Mary-Claire King, Rachel E. Klevit, "Mapping the functional domains of BRCA1 Interactions of the ring domains of BRCA1 and BARD1," The Journal of Biological Chemistry, Vol. 274, pp. 5659–5665, Feb. 1999.

[4] Khanna, K. K., and S. P. Jackson, "DNA double-strand breaks: signaling, repair and the cancer connection," Nature Genetics, Vol. 27, pp:247–254, 2001.

[5] Van Gent, D. C., J. H. Hoeijmakers, and R. Kanaar, "Chromosomal stability and the DNA double-stranded break connection," Nat. Rev. Genet., Vol. 2, pp:196–206, 2001.

[6] Junran Zhang, Henning Willers, Zhihui Feng, Jagadish C. Ghosh, Sang Kim, David T. Weaver, Jay H. Chung, Simon N. Powell, Fen Xia, "Chk2 Phosphorylation of BRCA1 Regulates DNA Double-Strand Break Repair," Molecular and Cellular Biology, Vol 24, pp . 708–718, Jan 2004.

[7] Laura N. Burga, Nadine M. Tung, Susan L. Troyan, Mihnea Bostina, Panagiotis A. Konstantinopoulos, Helena Fountzilas, Dimitrios Spentzos, Alexander Miron, Yosuf A. Yassin, Bernard T. Lee, Gerburg M. Wulf, Altered, " Proliferation and Differentiation Properties of Primary Mammary Epithelial Cells from BRCA1 Mutation Carriers," Cancer Research Vol.(4), pp: 1273-1278, Feb 2009.

[8] Sahil and Rose Mary Braun (2019) "GeneSurrounder: Network-based identification of disease genes in expression data: the Gene Surrounder method," *BMC Bioinformatics, Vol.* 20, pp:229, 2019.

[9] J.R.Morris and S.V. Tavtigian, "Use of BRCA1 protein:protein interactions to classify cancer risk," Breast Cancer Research 10(2):S1-S49, 2008.

[10] Zhenkun Na, Sijun Pan, Mahesh Uttamchandani and Shao Q Yao, "Protein-Protein Interaction Inhibitors of BRCA1 Discovered Using Small Molecule Microarrays," Methods and Protocols, Methods in Molecular Biology, 1518: 139-156, 2017.

[11] Václav Brázda, Lucia Hároníková, Jack C. C. Liao, Helena Fridrichová and Eva B. Jagelská, "Strong preference of BRCA1 protein to topologically constrained non-B DNA structures," BMC Molecular Biology, Vol. 17(14): https://doi.org/10.1186/s12867-016-0068-6, 2016.

[12] Vadiraj Kurdekar, Saranya Giridharan, Jasti Subbarao, Mamatha B. Nijaguna, Jayaprakash Periasamy, Sanjana Boggaram, Amol V. Shivange, Gayathri Sadasivam, Muralidhara Padigaru, Vijay Potluri, Ashok R. Venkitaraman, Kavitha Bharatham, "Structure-Guided Synthesis and Evaluation of Small-Molecule Inhibitors Targeting Protein–Protein Interactions of BRCA1 tBRCT Domain, " ChemMedChem, Vol. 14(18), pp:1620-1632, 2019.

[13] S. Peri, J. D. Navarro, R. Amanchy, T. Z. Kristiansen, C. K. Jonnalagadda, V. Surendranath, V. Niranjan, B. Muthusamy, T. K. Gandhi, M. Gronborg, N. Ibarrola, N. Deshpande, K. Shanker, H. N. Shivashankar, B. P. Rashmi, M. A. Ramya, Z. Zhao, K. N. Chandrika, N. Padma, H. C. Harsha, A. J. Yatish, M. P. Kavitha, M. Menezes, D. R. Choudhury, S. Suresh, N. Ghosh, R. Saravana, S. Chandran, S. Krishna, M. Joy, S. K. Anand, V. Madavan, A. Joseph, G. W. Wong, W. P. Schiemann, S. N. Constantinescu, L. Huang, R. Khosravi-Far, H. Steen, M. Tewari, S. Ghaffari, G. C. Blobe, C. V. Dang, J. G. Garcia, J. Pevsner, O. N. Jensen, P. Roepstorff, K. S. Deshpande, A. M. Chinnaiyan, A. Hamosh, A. Chakravarti, and A. Pandey, "Development of human protein reference database as an initial platform for approaching systems biology in humans," Genome Res, vol. 13, pp.2363-2371, Oct 2003.

[14] Jeong H, Mason SP, Barabási AL, Oltvai ZN, "Lethality and centrality in protein networks," Nature, Vol. 411(6833):41-42, 2001.

[15] Joy MP, Brock A, Ingber DE, Huang S., "High-betweenness proteins in the yeast protein interaction network," J Biomed Biotechnol, Vol.(2), pp:96-103, 2005.