Network Analysis Reveals the Connection Between Chronic Pancreatitis and Pancreatic Cancer

Li-sha LIa,*, He-ping WANG[b] and Yu-lin Li[c]

[a]The Key Laboratory of Pathobiology, Ministry of Education, College of Basic Medical Sciences, Jilin University, China

[b]Department of Neurosurgery, Tongji Hospital, Tongji Medical School

[c]The Key Laboratory of Pathobiology, Ministry of Education, College of Basic Medical Sciences, Jilin University, China

*lilisha@jlu.edu.cn

*Corresponding author

Keywords: Transcriptome, Chronic Pancreatitis, Pancreatic Cancer, Network Analysis.

Abstract. The relationship between chronic pancreatitis and pancreatic cancer needs further exploration in a system view. We constructed chronic pancreatitis and pancreatic cancer network based on gene expression data and protein-protein interaction, and then analyzed the global topological features of chronic pancreatitis and pancreatic cancer genes. Transcriptome profiling-based network analysis reveals the importance of chronic pancreatitis genes in pancreatic cancer by computational method.

Background

Exploring the mechanism of disease progression is a meaningful challenge in biomedicine. Many approaches have been developed to identify disease-associated genes and pathways [1, 2]. The biological network is the basis of complex biological systems and a leading method deciphering in complex diseases [3, 4]. Conventional approaches are not enough in clarifying complex diseases, because most molecules function through complex networks in an integrated fashion [5].

Network-based approaches to inflammation and cancer research are developed with the high-throughput data [6, 7]. This paper focuses on inflammation which is one of the important drivers of cancer [8, 9]. Integrating the information from inflammation and cancer is helpful to understand the nature of cancer. Inflammation and cancer are complex processes regulated by both environment and genetics [10]. Here, we emphasize the relationships between chronic pancreatitis and pancreatic cancer at a network level since chronic inflammation is a process of gradual decay of homeostasis increasing the susceptibility to cancer. Moreover, the risk of pancreatic cancer (PC) is significantly elevated in subjects with chronic pancreatitis (CP) [10]. In this paper, we used the transcriptional profiles, protein-protein interaction (PPI), pathway into network, and constructed a chronic pancreatitis-pancreatic cancer network (CPPCN) in gene level.

Methods

Gene Expression Data and Differential Expressed Genes

Gene expression profiling data E-EMBL-6 were obtained from the EMBL European Bioinformatics Institute database [11], which contain the stages of normal state, chronic pancreatitis and pancreatic cancer, each with 9 samples. The 27 samples of normal state, chronic pancreatitis, pancreatic cancer were used in this research. The differential expressed genes of CP and PC were obtained by t-test with the local false discovery rates (FDR)<0.05 using the fdrtool package [12].

Network Construction and Analysis

The differential expressed genes are mapped to the PPI network combining KEGG and Human Protein Reference Database (HPRD) and the connected components were extracted, CP, PC and CP-PC network were constructed [13]. The network parameters were analyzed by cytoscape [14].

Results

The Chronic Pancreatitis - Pancreatic Cancer Gene Network

The networks of different diseases can reveal the particular features of disease-related genes. Thus, we analyzed the pancreatic cancer progression by constructing a chronic pancreatitis (CP) network, a pancreatic cancer (PC) network and a chronic pancreatitis - pancreatic cancer network (CPPCN) (Fig. 1A). In these networks, the nodes are differentially expressed genes compared with the normal pancreas extracted from expression profiling data. There are 532 differentially expressed genes in CP, 1843 differentially expressed genes in PC and totally 1942 genes related to CP-PC (Fig. 1B). There are 190 nodes in CP network, 1011 nodes in PC network, and 1077 nodes in CPPCN.

In Fig. 1A, CPPCN has 1077 nodes which are more than randomly (Fig. 2A, 1000 degree-conserved networks are randomly chosen as control in human PPI network, and the number of nodes in the connected components is 1028.7±16.0). Therefore, PC related genes are inclined to be linked with CP related genes. Also, CPPCN has 2375 edges, and it is more linked than randomly (See Fig. 2B, 1000 degree-conserved networks are randomly chosen as control, and the number of edges within the connected component is 2163.8±37.9).

The degree distribution follows $P(k) \propto k^{-1.791}$, a power-law (Fig. 2B), so CPPCN is scale-free, error-tolerant and attack-vulnerable [3]. Also, we calculated the topological properties of CPPCN. Clustering coefficient is a measure of the inclination of components to form clusters [3]. Topological coefficient is a measure for the inclination of components to share interaction partners with other components [6]. As shown in Fig. 1C and Fig. 1D, clustering coefficient and topological coefficient decrease with the increase of nodes' degree in CPPCN. These results indicate that CPPCN has a potential hierarchical organization and CPPCN has local clustering, potentially representing functional modules, and more global connectivity mediated via hubs. It can be inferred that CPPCN shares the similar characteristic with Human PPI Network from HPRD [15]. Moreover, the topological coefficient confirms the modular network organization indicated by the clustering coefficient.
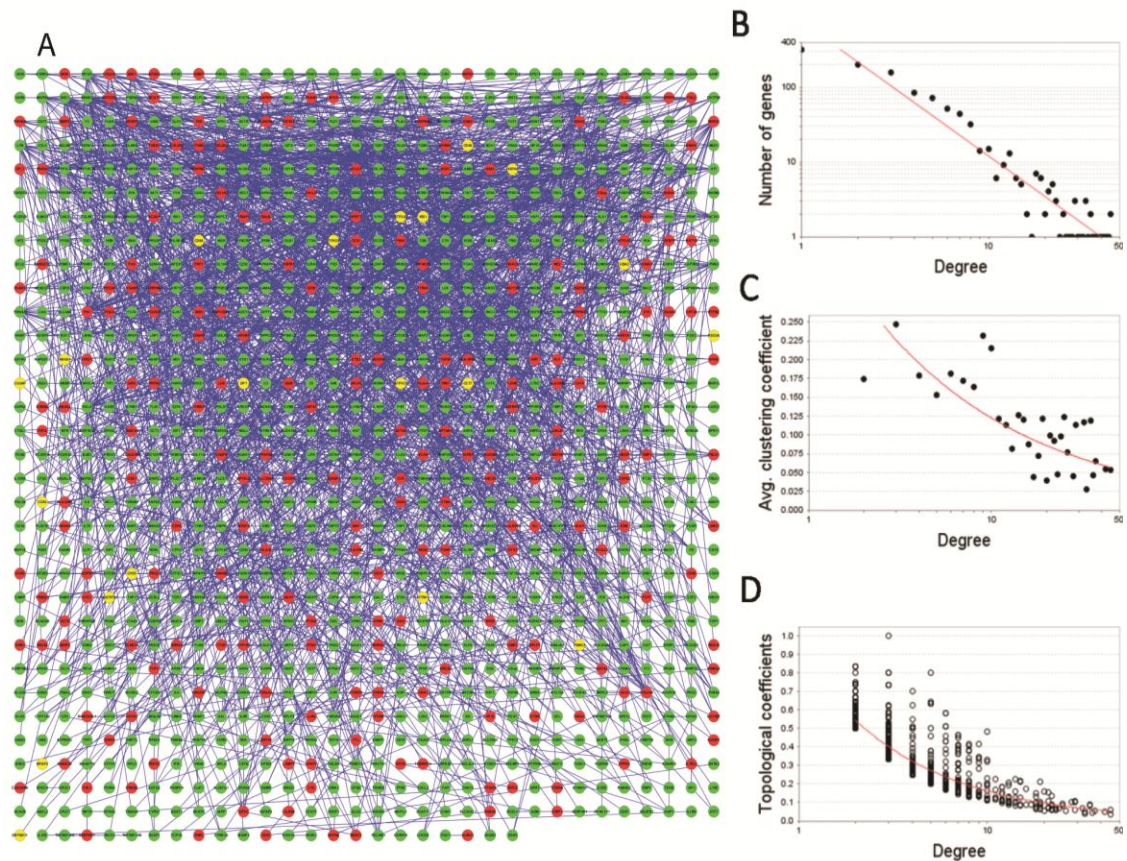
Figure.1. The CPPCN and its topological properties (A) A protein-protein interaction network connecting chronic pancreatitis and pancreatic cancer. Chronic pancreatitis -specific genes are colored in yellow; pancreatic cancer genes are colored in green; chronic pancreatitis and pancreatic cancer shared genes are colored in red. (B–D) Basic network features of CP-PCN. (B) The degree distribution, P (k) of the CP-PCN illustrates its scale-free topology. (C) The scaling of the clustering coefficient C (k) with the degree k illustrates the hierarchical architecture of CP-PCN. (D) The scaling of the topological coefficient with the degree k illustrates that hub genes in CP-PCN do not have more common neighbors than genes with fewer links.

In CPPCN, CP (nodes with yellow color), PC (nodes with green color) and CP-PC genes tend to share the uniform distribution in CPPCN. As shown in Fig. 2C, closeness centrality of CP, PC and CP-PC genes shows no difference (P>0.05). Betweenness centrality indicates how significant a node is in communicating between 2 nodes. In Fig. 2D, the betweenness centrality of CP genes is bigger than others in CPPCN (P<0.01). Average degree of CP genes is 2.019, which is less than that of CP-PC genes 4.496 and that of PC genes 4.283(P<0.01). This fact implies the significance of CP genes in the connectivity of CPPCN. However, 344 genes with degree, as well as 10 hubs–genes with degree more than 30 are detected. Previous studies have suggested that the hubs in our human interaction network could take more important cellular tasks than other proteins [15]. The above reveals that there are close implications among CP/PC genes, and then we explore how close the relationship is.

Connectivity Between Chronic Pancreatitis and Pancreatic Cancer

In all 532 CP differential expressed genes, 433 (81.39%) are overlapped with PC differential expressed genes (Fig. 2B). In all 303 CP genes connected in human PPI network, 250 (82.51%) are overlapped with pancreatic cancer genes connected in

human PPI network (Fig. 2C). We observed significant overlap between chronic pancreatitis and pancreatic cancer in differential expressed genes and in differential expressed genes that are connected in human PPI network. Overlap is a form of connectivity. Besides the overlapping genes, the relationship between CP and PC was checked by the interactions. In Fig. 2D, there are 57092 interactions in the human PPI network, while 248 interactions are among CP genes and 2213 interactions are among PC genes, and 200 interactions between both chronic pancreatitis and pancreatic cancer genes (p-value<1e-15).The above observation is inferred due to the close relationships between CP and PC. CP and PC are connected in both overlapping genes and interactions in the network level. The protein product of CP and PC are more likely to have physical interactions.
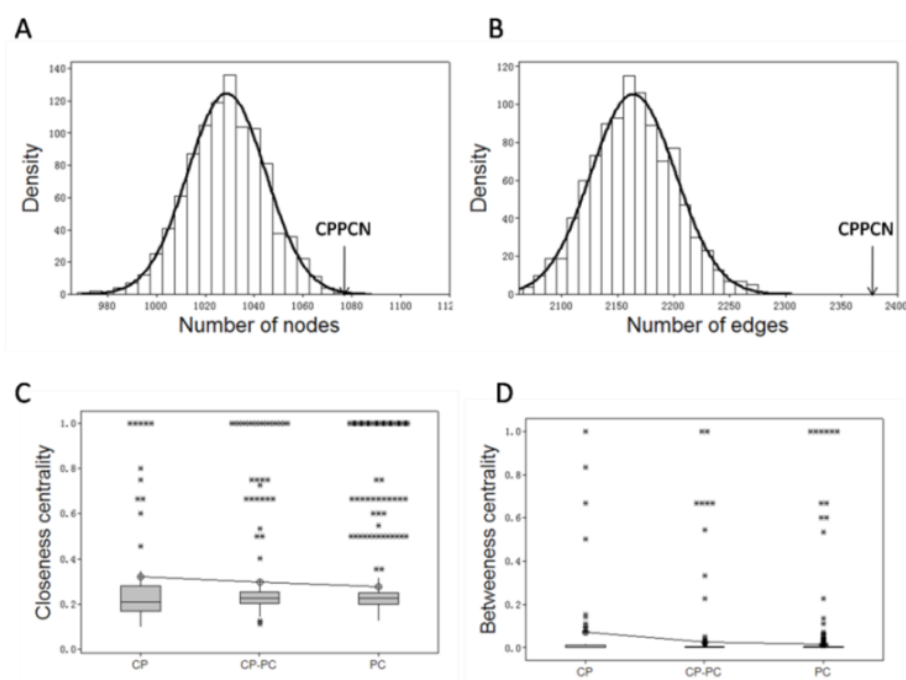

Figure.2. The further analysis of CPPCN (A)The number of vertexes of CPPCN is significantly larger than that of degree-conserved random networks (P<0.001). (B) The number of edges of CPPCN is significantly larger than that of degree-conserved random networks (P<0.001). (C) and (D) There is no difference (P>0.05) in closeness centrality and betweeness centrality of CP, PC and CP-PC genes in CPPCN.

Conclusion

There is an urgent need to connect chronic inflammation with cancer, because it can help to explain cancer initiation and progression. We prove that CP and PC are connected based on transcriptome data and protein-protein interaction by computation method.

Acknowledgements

References

[1] Baranzini, S.E., et al., Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. Hum Mol Genet, 2009. 18(11): p. 2078-90.

[2] Tiffin, N., M.A. Andrade-Navarro, and C. Perez-Iratxeta, Linking genes to diseases: it's all in the data. Genome Med, 2009. 1(8): p. 77.

[3] Barabasi, A.L. and Z.N. Oltvai, Network biology: understanding the cell's functional organization. Nat Rev Genet, 2004. 5(2): p. 101-13.

[4] Alon, U., Biological networks: the tinkerer as an engineer. Science, 2003. 301(5641): p. 1866-7.

[5] Hyduke, D.R. and B.O. Palsson, Towards genome-scale signalling network reconstructions. Nat Rev Genet, 2010. 11(4): p. 297-307.

[6] Vodovotz, Y., et al., Translational systems biology of inflammation. PLoS Comput Biol, 2008. 4(4): p. e1000014.

[7] Kreeger, P.K. and D.A. Lauffenburger, Cancer systems biology: a network modeling perspective. Carcinogenesis. 31(1): p. 2-8.

[8] Mantovani, A., et al., Cancer-related inflammation. Nature, 2008. 454(7203): p. 436-44.

[9] Grivennikov, S.I., F.R. Greten, and M. Karin, Immunity, inflammation, and cancer. Cell. 140(6): p. 883-99.

[10] Lowenfels, A.B., et al., Pancreatitis and the risk of pancreatic cancer. International Pancreatitis Study Group. N Engl J Med, 1993. 328(20): p. 1433-7.

[11] Stoehr, P.J. and R.A. Omond, The EMBL Network File Server. Nucleic Acids Res, 1989. 17(16): p. 6763-4.

[12] Strimmer, K., fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics, 2008. 24(12): p. 1461-2.

[13]. Xia, J., E.E. Gill, and R.E. Hancock, NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nat Protoc, 2015. 10(6): p. 823-44.

[14] Su, G., et al., Biological network exploration with cytoscape 3. Curr Protoc Bioinformatics, 2014. 47: p. 8 13 1-8 13 24.

[15] Stelzl, U., et al., A human protein-protein interaction network: a resource for annotating the proteome. Cell, 2005. 122(6): p. 957-68.