Network analysis of common genes related to esophageal, gastric, and colon cancers

Padina Vaseghi Maghvan1, Mostafa Rezaei-Tavirani2, Hakimeh Zali3, Abdolrahim Nikzamir4, Saeed Abdi5, Mahsa Khodadoostan6, Hamid Asadzadeh-Aghdasi7

1Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
3Proteomics Research Center, Department of Tissue engineering and Applied Cell, School of Advanced Technologies in Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
4Faculty of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
5Gastroenterology and Liver Diseases Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran
6Department of Gastroenterology and Hepatolgy, Isfahan University of Medical Sciences, Isfahan, Iran
7Basic and Molecular Epidemiology of Gastrointestinal Disorders Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran

ABSTRACT

Aim: The aim of this study was to provide a biomarker panel for esophageal, gastric and colorectal cancers. It can help introducing some diagnostic biomarkers for these diseases.

Background: Gastrointestinal cancers (GICs) including esophageal, gastric and colorectal cancers are the most common cancers in the world which are usually diagnosed in the final stages and due to heterogeneity of these diseases, the treatments usually are not successful. For this reason, many studies have been conducted to discover predictive biomarkers.

Methods: In the present study, 507 genes related to esophageal, gastric and colon cancers were extracted. The network was constructed by Cytoscape software (version 3.4.0). Then a main component of the network was analyzed considering centrality parameters including degree, betweeness, closeness and stress. Three clusters of the protein network accompanied with their seed nodes were determined by MCODE application in Cytoscape software. Furthermore, Gene Ontology (GO) analysis of the key genes in combination to the seed nodes was performed.

Results: The network of 17 common differential expressed genes in three esophageal, gastric and colon adenocarcinomas including 1730 nodes and 9188 edges were constructed. Eight crucial genes were determined. Three Clusters of the network were analyzed by GO analysis.

Conclusion: The analyses of common genes of the three cancers showed that there are some common crucial genes including TP53, EGFR, MYC, AKT1, CDKN2A, CCND1 and HSP90AA1 which are tightly related to gastrointestinal cancers and can be predictive biomarkers for these cancers.

Keywords: Colon cancer, Gastric cancer, Esophageal cancer, Gene ontology, Biomarker.

(Please cite as: Vaseghi Maghvan P, Rezaei-Tavirani M, Zali H, Nikzamir A, Abdi S, Khodadoostan M, et al. Network analysis of common genes related to esophageal, gastric, and colon cancers. Gastroenterol Hepatol Bed Bench 2017;10(4):295-302).
in the world. Although the mortality rate of the cancer has been decreased in recent years, the prevalence of gastric and gastroesophageal cancers have risen suddenly (4).

Colorectal cancer (CRC) is the third most frequently diagnosed cancer in males and the second in females. One of the main risk factors of these diseases in the western countries is the lifestyle (5). Despite a wide range of studies about these diseases to determine biomarkers, extensive heterogeneity of patients has led to the failure of target therapy. The late diagnosis in acute stage of the disease will lead to an increase in the mortality rate (6). Association of these cancers with chronic digestive problems such as intestinal polyps and chronic injuries in the esophagus and stomach tissues, as well as the significant relationship between these three cancers implies the common predictive biomarkers to investigate (7-8).

Bioinformatics approaches specially study of protein networks of diseases have attracted attention of many medical scientists. So far, there are many bioinformatics studies on various diseases to determine diagnostic and therapeutic biomarkers. Some of these studies have been done on the neurodegenerative diseases such as MS, Alzheimer's and Huntington's (9-10). In addition, some other studies on diseases like depression and psychiatric disorders and some cancers have been conducted (11-13). This approach can lead to provide new diagnostic protocol for early detection and prognosis of these three diseases (14).

Protein-protein interaction (PPI) is one of the most fundamental underlying mechanisms of life. PPI network analysis is an attractive field in proteomics and bioinformatics which provides deeper understanding of cellular and molecular processes in the case of diseases (15). Any disturbance in protein interactions can cause the onset of a disease. Many human diseases are the result of such disorders (16). PPI network analysis accompanied with GO analysis can be considered as an excellent complementary research for experimental studies about mechanisms and risk factors of diseases, drug resistant mechanism and detection of predictive and therapeutic biomarkers (17). Gastrointestinal cancers are the most common related deaths in the world and early diagnosis can help better targeted therapy and reducing mortality of patients. So bioinformatics analyses can be one of the solutions to these cancers (18).

In this research, the common genes involved in esophageal, gastric and colorectal cancers are interacted to achieve a suitable biomarker panel.

### Methods

Esophageal, gastric and colon cancers-related genes were extracted from String database. Of the 507 genes, 232, 143 and 132 genes were related to colon, gastric and esophageal cancer, respectively. 17 common genes

#### Table 1. Common differential expression genes in esophageal, gastric and colon adenocarcinomas

| R  | Name of gene | Description                                      |
|----|--------------|--------------------------------------------------|
| 1  | AKT1         | RAC-alpha serine/threonine-protein kinase        |
| 2  | CCND1        | G1/S-specific cyclin-D1                         |
| 3  | CDH1         | Cadherin-1                                       |
| 4  | CDKN2A       | Cyclin-dependent kinase inhibitor 2A             |
| 5  | CDX2         | Homeobox protein CDX-2                          |
| 6  | EGFR         | Epidermal growth factor receptor                |
| 7  | ERBB2        | Receptor tyrosine-protein kinase erbB-2          |
| 8  | GAST         | Gastrin                                          |
| 9  | IL1B         | Interleukin-1 beta                              |
| 10 | KRT7         | Keratin, type II cytoskeletal 7                  |
| 11 | MUC2         | Mucin-2                                          |
| 12 | MUC5AC       | Mucin-5AC                                        |
| 13 | MYC          | Myc proto-oncogene protein                      |
| 14 | PTGS2        | Prostaglandin G/H synthase 2                     |
| 15 | SMUG1        | Single-strand selective monofunctional uracil DNA glycosylase |
| 16 | TP53         | Cellular tumor antigen p53                      |
| 17 | TYMS         | Thymidylate synthase                            |
were identified between these cancers (Table 1). Then protein network of these proteins was created using Cytoscape software (version 3.4.0). A main component of the network was selected and analyzed considering centrality parameters including degree, betweenness, closeness and stress. According to numerous literatures, centrality analysis is a useful method for ranking of network elements. This analysis can identify the key players in biological processes (19). Suitable cutoff for degree values was determined by average of degree plus two standard deviations. The top 5% genes based on betweenness centrality (BC), closeness Centrality (CC) and stress were selected for more analyses. The crucial genes were highlighted. Three clusters of the protein network’s main component accompanied with their seed nodes were determined by MCODE application of Cytoscape software. Furthermore, Gene Ontology (GO) analysis of the key genes in combination to the seed nodes was performed. Molecular function, cellular components and biological process were analyzed by ClueGO application of Cytoscape software. In addition, the pathways related to the proteins of each cluster checked out by KEGG database.

**Results**

The network of 17 common differential expressed genes in esophageal, gastric and colon adenocarcinomas including 1730 nodes and 9188 edges was constructed (Figure 1). This network contains four connected components. Network of the main connected components and the related statistical information have been displayed in Figure 2. Closed correlation in degree values of the nodes is corresponded to the scale free network (Figure 3). According to cutoff value, eight proteins with the highest degrees were selected as hub nodes of the network. All of these proteins were bottleneck nodes. All hub-bottleneck nodes were included in the selected nodes based on stress. Seven of hub-bottleneck nodes were presented in the identified nodes considering closeness (see Table 2). The presence

![Figure 1](image1.png)

**Figure 1.** Protein-protein interaction network of esophageal, gastric and colon adenocarcinomas included 1730 nodes and 9188 edges. The network included four connected components characterized by 1297, 212, 160 and 61 nodes, and 8323, 425, 304 and 136 edges, respectively. Statistic of the network has been shown at the bottom of the figure.
of hub-bottlenecks in the determined clusters was investigated. In cluster 1 which has the highest score, there were three hub-bottlenecks (AKT1, TP53 and MYC). Moreover, one of hub-bottlenecks was in cluster 2 (CDKN2A) and three hub-bottlenecks were in cluster 3 (HSP90AA1, CCND1 and EGFR). None of the crucial proteins was common between the analyzed clusters. The clusters 1 and 3 were introduced as important protein complexes related to the studied cancers. The seeds of clusters 1-3 are SP1, TUBA1A and HDAC2.

Gene Ontology analysis of hub-bottleneck nodes and the identified seeds was performed by ClueGO application.
(Figures 4 and S 1). Results of cluster’s GO based on KEGG database have been shown in Figures 5-7.

**Figure 3.** Degree distribution for the nodes of main connected components. The equation and the other statistical parameters of power law have been shown at the bottom of the figure. Closed correlation in degree values of the nodes is corresponded to the scale free network.

**Figure 4.** Nine clustered GO terms (the 43 presented terms in the figure 4) for the elements of main connected components.

**Discussion**

In the biological network, nodes can be genes, proteins, metabolites or diseases, and edges indicate the relationship between nodes. The hubs are nodes that have the most connection with the nodes around them. Hub-bottleneck nodes are the hub nodes that have more control role on the other nodes of network (20). Most biological networks are scale-free networks. That means the notion of scale-free refers to the lack of a characteristic degree or scale. In scale free networks, the most nodes have a degree close to the average (21). In the present study, 507 genes related to esophageal, gastric and colon cancers were extracted. The network was constructed by the common genes and the crucial genes were selected based on their relationship with other nodes. Like many cancers, there is a reasonable possible biomarker panel related to the studied cancers (14, 22). The findings indicate that the network includes three distinguished clusters.

**Figure 5.** KEGG pathways related to cluster 1. Each color indicates a unique pathway. The nodes highlighted by two or more colors are corresponding to several affected pathways.

**Figure 6.** KEGG pathways related to cluster 2. The single color indicates a unique pathway. The nodes highlighted by two or more colors are corresponding to several affected pathways.
According to KEGG database analysis, one of the key genes in cluster 1 is TP53 which plays a role in WNT signaling pathway (23). TP53 has an important effect on the cell cycle and prolactin signaling pathway. Many studies have linked prolactin levels with the development of various forms of cancer (24, 25). In cluster 1, AKT1 is effective in WNT signaling pathway, fluid shear stress and atherosclerosis and many studies confirm these roles (26, 27). Infections caused by viruses or changes in the natural flora of the digestive tract are the important topics discussed in advanced digestive diseases (7, 28, 29). In this cluster, the CSK, MAPK11 and PTPN11 genes are involved in the epithelial cell signaling in Helicobacter pylori (H. pylori) infection. In this regard, many studies have identified the association of gastric ulcer disease with gastric cancer (30). According to GO analyses, the seed choline metabolism and GnRH signaling pathway. In recent years, attention has been paid to GnRH signaling pathway and choline metabolism for the diagnosis and treatment of cancer (31, 32). In cluster 2, the key genes CDKN2A, HSP90AB1, TRAF2 and RAF1 are effective agents in cancer pathway. In cluster 3, CCND1 and EGFR play role in MAPK signaling pathway and HSP90AA1 is important in PI3K-Akt pathway. Both of the pathways are related to the cell cycle, apoptosis, invasion and differentiation (33, 34). Moreover, some genes of cluster 3 participate in herpes simplex infection which is another indication of the role of chronic infections in the development of cancers and many studies have confirmed these evidences (35, 36).

By GO analysis, nine main processes including 43 GO terms were identified. The dominant process is
“response to UV” that consists of cell transfusion, ligand binding, cell growth, and response to epidermal growth factor. The next dominant term group are focused on apoptotic processes and the regulation of the cell aging process. Regulation of fibroblastic cells and apoptosis process of them are the third and the fourth processes, respectively. Tight relationship between the components of this process and the studied cancers is confirmed by literatures (18, 37-39). From many years ago, epidermal growth factor receptor is considered as a therapeutic target for gastrointestinal cancers (40). The relationship between the cell aging and cancer has been proven by many studies (41).

Aging and H. pylori inflammation as two risk factors of gastric cancer have been highlighted in the previous studies (42). Since the regulations of fibroblastic cells and their apoptotic process were introduced as the concern terms, the relationship between gastrointestinal cancers and fibroblast cells has been discussed repeatedly (43, 44).

In the present study, the analyses showed that some common crucial genes such as TP53, EGFR, MYC, AKT1, CDKN2A, CCND1 and HSP90AA1 are tightly related to gastrointestinal cancers. More analyses indicated that it is possible to introduce a biomarker panel which can be used for prognosis and early detection of gastrointestinal cancers however more validation is required.

Acknowledgment
This project was supported by Shahid Beheshti University of Medical Sciences.

Conflict of interests
The authors declare that they have no conflict of interest.

References
1. Xin Z, Jiang S, Jiang P, Yan X, Fan C, Di S, et al. Melatonin as a treatment for gastrointestinal cancer: a review. J Pineal Res 2015;58:375-87.
2. Zali H, Ahmadi G, Bakhshandeh R, Rezaei-Tavirani M. Proteomic analysis of gene expression during human esophagus cancer. J Paramed Sci 2011; 2:37-44.
3. Visser E, Franken IA, Brosens LA, Ruurda JP, van Hillegersberg R. Prognostic gene expression profiling in esophageal cancer: a systematic review. Oncotarget 2017;8:5566-77.
4. Shokrgozar MA, Zali H, Rezaei-Tavirani M, Amanzadeh A. Comparison of Two Staining Assays Trypan Blue and MTT in vitro Evaluation of Human Calprotectin Proliferation Inhibition on Human Gastric Cancer Cells. Trauma Mon 2007;(2):127-37.
5. Siegel R, Desantis C, Jemal A. Colorectal cancer statistics, 2014. CA Cancer J Clin 2014;64:104-17.
6. Zamanian Azodi M, Peyvandi H, Rostami-Nejad M, Safaei A, Rostami K, Vafaee R, et al. Protein-protein interaction network of celiac disease. Gastroenterol Hepatol Bed Bench 2016;9:268-77.
7. Wojtulewski LE, Peek RM Jr, Coburn LA. The Role of the Microbiome in Gastrointestinal Cancer. Gastroenterol Clin North Am 2016;45:543-56.
8. Nam SY, Park BJ, Ryu KH, Nam JH. Effect of Helicobacter pylori infection and its eradication on the fate of gastric polyps. Eur J Gastroenterol Hepatol 2016;28:449-54.
9. Ragnedda G, Disanto G, Giovannoni G, Ebers GC, Sotgiu S, Ramagopalan SV. Protein-protein interaction analysis highlights additional loci of interest for multiple sclerosis. PLoS One 2012;7:e46730.
10. Tourette C, Li B, Bell R, O'Hare S, Kaltenbach LS, Mooney SD, et al. A large scale Huntington protein interaction network implicates Rho GTPase signaling pathways in Huntington disease. J Biol Chem 2014;289:6709-26.
11. Rakshit H, Rathi N, Roy D. Construction and analysis of the protein-protein interaction networks based on gene expression profiles of Parkinson's disease. PLoS One 2014;9:e103047.
12. Moreno-Gonzalez I, Edwards Iii G, Salvadores N, Shahmawaz M, Diaz-Espinoza R, Soto C. Molecular interaction between type 2 diabetes and Alzheimer's disease through cross-seeding of protein misfolding. Mol Psychiatry 2017;22:1327-34.
13. Safaei A, Rezaei-Tavirani M, Sobhi S, Akbari ME. Breast cancer biomarker discovery: Proteomics and genomics approaches. Iran J Cancer Prev 2013;6:45-53.
14. Zamanian-Azodi M, Rezaei-Tavirani M, Rahmati-Rad S, Hasanzadeh H, Rezaei Tavirani M, Seyyedi SS. Protein-Protein Interaction Network could reveal the relationship between the breast and colon cancer. Gastroenterol Hepatol Bed Bench 2015;8:215-24.
15. Kuzmanov U, Emili A. Protein-protein interaction networks: probing disease mechanisms using model systems. Genome Med 2013;5:37.
16. Gonzalez MW, Kann MG. Chapter 4: Protein interactions and disease. PLoS Comput Biol 2012;8:e1002819.
17. Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, Goliaei B, Peyvandi AA. Protein-protein interaction networks (PPI) and complex diseases. Gastroenterol Hepatol Bed Bench 2014;7:17-31.
18. Zamanian-Azodi M, Rezaei-Tavirani M, Hasanzadeh H, Rahmani Rad S, Dalilian S. Introducing biomarker panel in esophageal, gastric, and colon cancers: a proteomic approach. Gastroenterol Hepatol Bed Bench 2015;8:6-18.

19. Koschützki D, Schreiber F. Centrality analysis methods for biological networks and their application to gene regulatory networks. Gene Regul Syst Bio 2008;2:193-201.

20. Ryan DP, Matthews JM. Protein-protein interactions in human disease. Curr Opin Struct Biol 2005;15:441-6.

21. Goh KI, Oh E, Jeong H, Kahng B, Kim D. Classification of scale-free networks. Proc Natl Acad Sci U S A 2002;99:12583-8.

22. Rezaei-Tavirani M, Rezaei-Tavirani M, Mansouri V, Mahdavi SM, Valizadeh R, Rostami-Nejad M, et al. Introducing crucial protein panel of gastric adenocarcinoma disease. Gastroenterol Hepatol Bed Bench 2017;10:21-8.

23. Anastas JN, Moon RT. WNT signalling pathways as therapeutic targets in cancer. Nat Rev Cancer 2013;13:11-26.

24. Tan D, Tang P, Huang J, Zhang J, Zhou W, Walker AM. Expression of a constitutively active prolactin receptor causes histone trimethylation of the p53 gene in breast cancer. Chin Med J (Engl) 2014;127:1077-83.

25. Yin Y, Tainsky MA, Bischoff FZ, Strong LC, Wahl GM. Wild-type p53 restores cell cycle control and inhibits gene amplification in cells with mutant p53 alleles. Cell 1992;70:937-48.

26. Pandurangan AK. Potential targets for prevention of colorectal cancer: a focus on PI3K/Akt/mTOR and Wnt pathways. Asian Pac J Cancer Prev 2013;14:2201-5.

27. Melchior B, Frangos JA. Distinctive subcellular Akt-1 responses to shear stress in endothelial cells. J Cell Biochem 2014;115:121-9.

28. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. Nat Rev Microbiol 2014;12:661-72.

29. Fazeli Z, Alebouyeh M, Rezaei Tavirani M, Azimirad M, Yadegar A. Helicobacter pylori CagA induced interleukin-8 secretion in gastric epithelial cells. Gastroenterol Hepatol Bed Bench 2016;9:S42-6.

30. Wroblewski LE, Peek RM Jr, Wilson KT. Helicobacter pylori and gastric cancer: factors that modulate disease risk. Clin Microbiol Rev 2010;23:713-39.

31. Limonta P, Montagnani Marelli M, Mai S, Motta M, Martini L, Moretti RM. GnRH receptors in cancer: from cell biology to novel targeted therapeutic strategies. Endocr Rev 2012;33:784-811.

32. Glunde K, Penet MF, Jiang L, Jacobs MA, Bhujwalla ZM. Choline metabolism-based molecular diagnosis of cancer: an update. Expert Rev Mol Diagn 2015;15:735-47.

33. Kim EK, Choi EJ. Pathological roles of MAPK signaling pathways in human diseases. Biochim Biophys Acta 2010;1802:396-405.

34. Fresno Vara JA, Casado E, De Castro J, Cejas P, Belda-Iniesta C, González-Barón M. PI3K/Akt signalling pathway and cancer. Cancer Treat Rev 2004;30:193-204.

35. Hennessy BT, Smith DL, Ram PT, Lu Y, Mills GB. Exploiting the PI3K/akt pathway for cancer drug discovery. Nat Rev Drug Discov 2005;4:988-1004.

36. Hsueh YS, Yen CC, Shih NY, Chiang NJ, Li CF, Chen LT. Autophagy is involved in endogenous and NVP-AUY922-induced KIT degradation in gastrointestinal stromal tumors. Autophagy 2013;9:220-33.

37. Khayer N, Zamanian-Azodi M, Mansouri V, Ghassemi-Broumand M, Rezaei-Tavirani M, Heidari MH, et al. Oral squamous cell cancer protein–protein interaction network interpretation in comparison to esophageal adenocarcinoma. Gastroenterol Hepatol Bed Bench 2017;10:118-24.

38. Mukai S, Oue N, Oshima T, Mukai R, Tatsumoto Y, Sakamoto N, et al. Overexpression of Transmembrane Protein BST2 is Associated with Poor Survival of Patients with Esophageal, Gastric, or Colorectal Cancer. Ann Surg Oncol 2017;24:594-602.

39. Inoue Y, Matsuura S, Yoshimura K, Iwashita Y, Kahyo T, Kawase A, et al. Characterization of V-set and immunoglobulin domain containing 1 exerting a tumor suppressor function in gastric, lung, and esophageal cancer cells. Cancer Sci 2017;108:1701-14.

40. Hecht JR, Bang YJ, Qin SK, Chung HC, Xu JM, Park JO, et al. Lapatinib in Combination With Capecitabine Plus Oxaliplatin in Human Epidermal Growth Factor Receptor 2-Positive Advanced or Metastatic Gastric, or Gastrolesophage Adenocarcinoma: TRIO-013/LOGiC–A Randomized Phase III Trial. J Clin Oncol 2016;34:443-51.

41. Lasry A, Ben-Neriah Y. Senescence-associated inflammatory responses: aging and cancer perspectives. Trends Immunol 2015;36:217-28.

42. Levi E, Sochacki P, Khoury N, Patel BB, Majumdar AP. Cancer stem cells in Helicobacter pylori infection and aging: Implications for gastric carcinogenesis. World J Gastrointest Pathophysiol 2014;5:366-72.

43. Hasegawa T, Yashiro M, Nishii T, Matsuoka J, Fuyuhiro Y, Morisaki T, et al. Cancer-associated fibroblasts might sustain the stemness of scirrhous gastric cancer cells via transforming growth factor-β signaling. Int J Cancer 2014;134:1785-95.

44. Yang TS, Yang XH, Chen X, Wang XD, Hua J, Zhou DL, et al. MicroRNA-106b in cancer-associated fibroblasts from gastric cancer promotes cell migration and invasion by targeting PTEN. FEBS Lett 2014;588:2162-9.