Gastroenterology and Hepatology From Bed to Bench.

ORIGINAL ARTICLE

Oral squamous cell cancer protein-protein interaction network interpretation in comparison to esophageal adenocarcinoma

Nasibeh Khayer[1], Mona Zamanian-Azodi[2], Vahid Mansouri[3], Mohammad Ghassemi-Broumand[2], Mostafa Rezaei-Tavirani[1] , Mohammad Hossein Heidari[1], Majid Rezaei Tavirani[4]

[1]*Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[2]*Student Research Committee, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[3]*Physiotherapy Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[4]*Faculty of Medicine, Iran University of Medical Sciences, Tehran, Iran*

ABSTRACT

Aim: The aim of this study is to present the oral Squamous Cell Cancer protein-protein interaction network interpretation in comparison to esophageal adenocarcinoma.

Background: Oral squamous cell cancer (OSCC) is a common disease worldwide, with poor prognosis and limited treatment. Thus, introducing molecular markers through network analysis can be helpful.

Methods: STRING database (DB) was applied for network construction through Cytoscape 3.4.0. Clue GO handled the gene annotation for the retrieved clusters. Eight proteins were indicated to be differential in the network constitution.

Results: The centrality and clustering analysis indicate that TP53 plays an over-significant role in network integrity among eight most central proteins including TP53, AKT1, EGFR, MYC, JUN, CDH1, CCND1, and CTNNB1. The suggested biomarker set is very similar to the related biomarker panel of esophageal adenocarcinoma.

Conclusion: The ontology analysis implies that the prominent proteins are involved in regulation of smooth muscle cell proliferation, regulation of fibroblast proliferation, and response to UV-A processes. In conclusion, these proteins and their associated biological processes may be more critical compared to other reported biomarkers for OSCC. Nevertheless, validation studies are required for confirming the pivotal role of potential candidates. Similar biomarker panel of this disease and esophagus adenocarcinoma is corresponded to the origin of the two malignancies.

Keywords: Oral squamous cell cancer (OSCC), Protein-protein interaction network analysis, Clustering analysis, Gene ontology.

(Please cite as: **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(2):118-124).**

Introduction

Oral squamous cell cancer (OSCC) ranks as the sixth prevalent malignancy and the most important oral cancer in the world (1, 2). Patients with this malignancy have less than 60% chance of survival (1). Many cases are not detectable before the cancer reaches its advanced levels (2) when it is mostly not curable (3). The statistics show that about two-thirds of these patients are diagnosed at the severe level. The common diagnostic method is to investigate oral cavity and obtain biopsies, which is not very promising (3). The need for improved diagnostic and treatment approaches prompt different molecular approaches for better understanding of the disease mechanisms (4, 5). One of the most important molecular fields for this purpose is investigating related promising biomarkers for different stages of malignancy management, especially at protein level. The associated proteins are in a systematic interacting pattern that is called protein-protein interaction network (6). Specific elements in this

Received: 12 January 2017 *Accepted*: 18 March 2017

Reprint or Correspondence: **Mona Zamanian-Azodi,** PhD. Research Committee, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

E-mail: mona.azodi@gmail.com

complex interacting profile are more prominent than the others, considering their centrality properties. The central proteins have a vast interacting topology and any small change in them may trigger abnormalities in the whole system (7). These malfunctions result in different types of biological processes and functions of the system that could alter the phenotype of the organism (8). When severe modification occurs in the interactome profile, different diseases can manifest depending on the architecture of sequential modification growth (9). In this light, studying interactome map of diseases, including malignant tumors, may provide a better insight into cancer behavior. Many biomarkers have been identified for OSCC (1); however, highlighting the most promising ones regarding topological features may facilitate molecular approaches and consequently clinical application (10). Therefore, in this study, protein-protein interaction network and its properties have been investigated. The findings are compared to esophageal adenocarcinoma.

Methods

Network construction was handled using Cytoscape 3.4.0 and its integrated network query STRING DB. STRING DB V 10.5 (http://string-db.org/) provides different interaction information from protein to disease query (11). Each interacting component (protein) is assigned an association score and a cutoff can be fitted for the edges. Normally, the default cutoff is set to 0.4. Here, the default cutoff was considered for the network construction. First, the name of disease was queried and the related proteins from disease databases were retrieved as a visualized network in the Cytoscape platform. More information was also provided by the network construction for the nodes and edges including disease score, associated tissues, sequence and edge score (12). The network analyzer examined the network for different centrality parameters. Degree, betweenness, stress, and closeness were analyzed with this application. Module analysis was carried out using MCODE 1.4.2 (ftp://ftp.mshri.on.ca/pub/BIND/Tools/MCODE) and Moduland 2.0. The first application detects different communities that are dense regions of protein connections of the network (13). The statistical criteria for MCODE (Molecular Complex Detection) are as follow: K score =2, Degree cutoff= 2, Node score cutoff= 0.2 as the default. Moduland identifies clusters of proteins and additionally represents hierarchical complexes and overlaps (14). GlueGO plug-in conducted further analysis based on ontology evaluation for the top proteins. This plug-in provides functional groups of terms linked to the queried input. The statistical values for this analysis were as follow: kappa score cutoff was set as default= 0.4, the p-Value< 0.05, the number of genes per term was set to 2 and the percentage for the queried terms was 3. Other statistical options such as the GO group level remained as default: min=2 and max=8. The correction method applied for p-Value was Bonferroni step down. In addition, enrichment/depletion two-sided hypergeometric test was also selected as the default option (15, 16).

Results

Cytoscape, STRING DB, analyzed the OSCC protein-protein interaction network. The disease was queried for top 300 contributing proteins with the default confidence score cutoff = 0.4 (Figure 1).

Network centrality analysis was performed with one of the integrated algorithms in Cytoscape called Network analyzer. This plug-in can present a visualization of the centrality changes for each of the parameters. One of the main centrality values is degree (Figure 2).

Table 1. List of nodes. BC, CC, and DC indicate betweenness centrality, closeness centrality and disease score, respectively.

R	Name	Degree	BC	CC	Stress	DS
1	TP53	134	0.12	0.69	43674	2.5
2	AKT1	116	0.05	0.65	28690	1.9
3	EGFR	115	0.07	0.64	30870	1.9
4	MYC	102	0.09	0.62	35410	1.5
5	JUN	99	0.05	0.62	24092	0.9
6	CDH1	96	0.03	0.61	20546	2.4
7	CCND1	92	0.04	0.61	24672	2.2
8	CTNNB1	90	0.04	0.60	20288	1.9

For network centrality analysis, the network analyzer was used. The centeral properties including degree, betweenness, closeness, and stress were calculated for the nodes. The 10% of highest

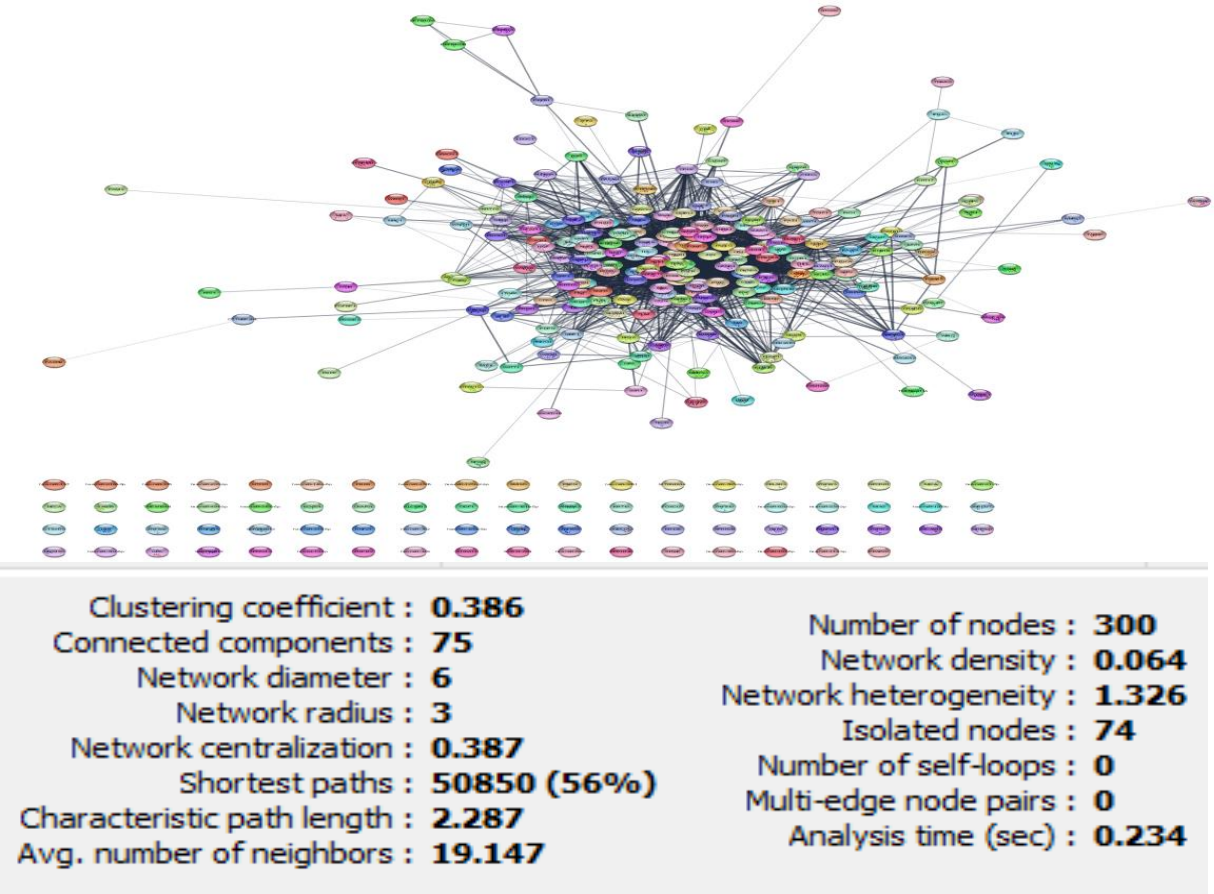

Figure 1. Statistical information related to the network and its centrality values. This network consists of 74 isolated nodes. The number of query nodes are 300. Confidence score = 0.4

amounts for these indices (including 23 nodes in each category) were considered for setting a cutoff. If the rank of a node among the four groups was less than 11, the node was selected as crucial node (Table 1). Degree distribution for the crucial nodes is presented in figure 3.

Module analysis is one of the methods for network protein clustering identification. Here, the first component of the main network was analyzed by two applications separately (Figures 4 and 5). The biological process analysis for the eight proteins was done using ClueGO (Figure 6). Due to low resolution, this figure is presented in figure 7 with more details.

Discussion

Oral cancer is a common neoplasm in the world (17). OSCC has the highest rate among this type of cancer (18). Identification of protein signature has aided with understanding the disease molecular basis and can be valuable for clinical approaches (19). Network medicine is one of the new disciplines that provides further insight of molecular concept of any kind of diseases such as malignant cancers (20). A network query of OSCC showed that the network consists of a big component with some isolated nodes, which are about one fourth of the whole network as indicated in figure 1. The first component was designated for further analysis and the isolated nodes were omitted. At first, the centrality study found some central nodes based on corresponding parameters. One of the most known centrality criteria is degree, and its visualization for the related nodes is presented in figure 2. In order to choose the most important proteins, four centrality parameters were selected. Network analysis identified eight prominent proteins as shown in table 1, including TP53, AKT1, EGFR, MYC, JUN, CDH1, CCND1, and CTNNB1 that can be explored for more

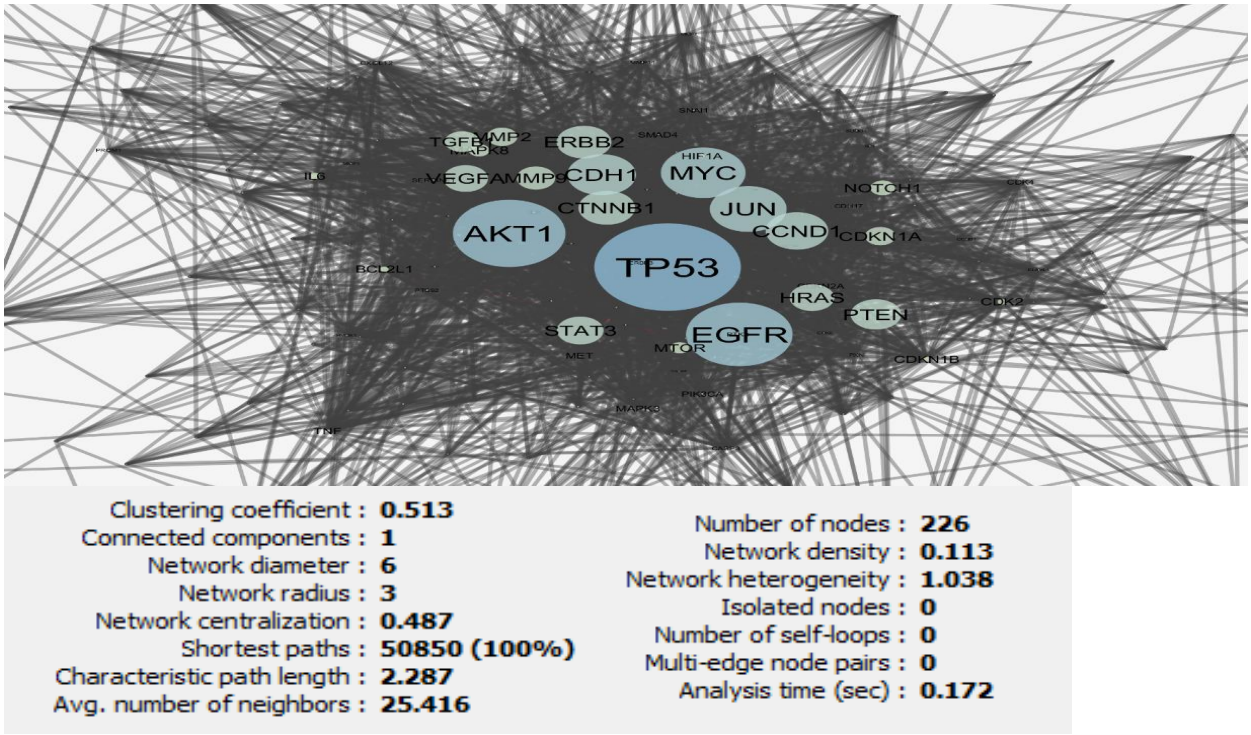

Figure 2. Network analysis for the main connected component (network) using network analyzer. The network consists of 226 nodes. The central nodes in this network are zoomed to yield a better view. The bigger and darker the nodes, the higher the degree centrality

investigations such as gene ontology examinations. The highest ranked protein among those identified is TP53,

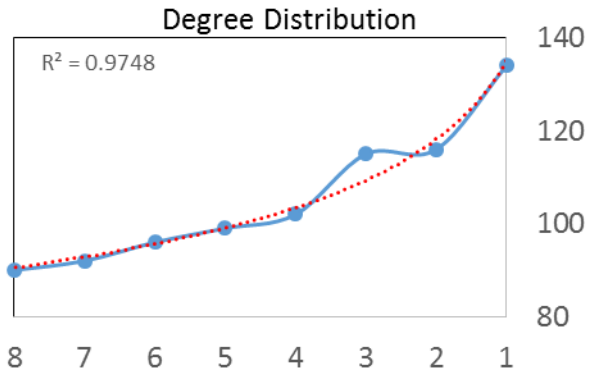

Figure 3. Trend of degree distribution among central nodes. The horizontal axis represents rank of node in table 1.

which is a famous protein for any kind of cancer (21, 22). The prominent roles of AKT1, EGFR, MYC, JUN, CDH1, CCND1, and CTNNB1 in colon, esophageal, gastric and some other cancers are discussed in more details (23-28). This gene panel is 87% similar to the related biomarker panel of

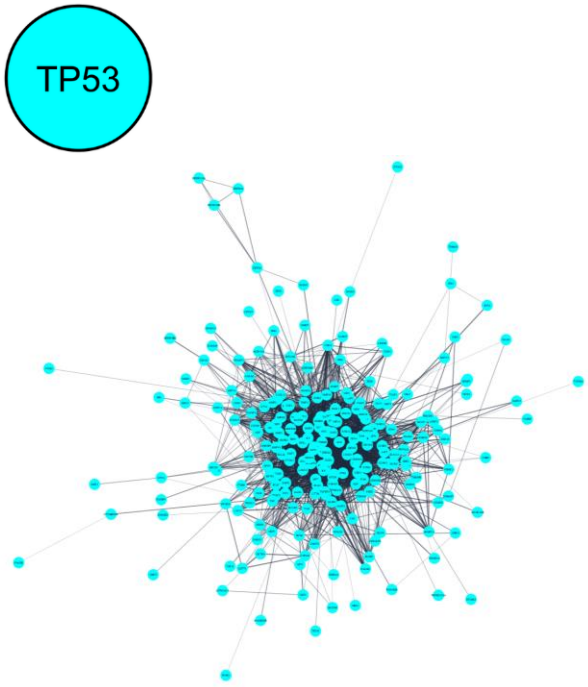

Figure 4. Moduland analysis of OSCC showed that the first component of the network consisting of 226 nodes and 2875 edges is a cluster. This cluster closely depends on TP53.

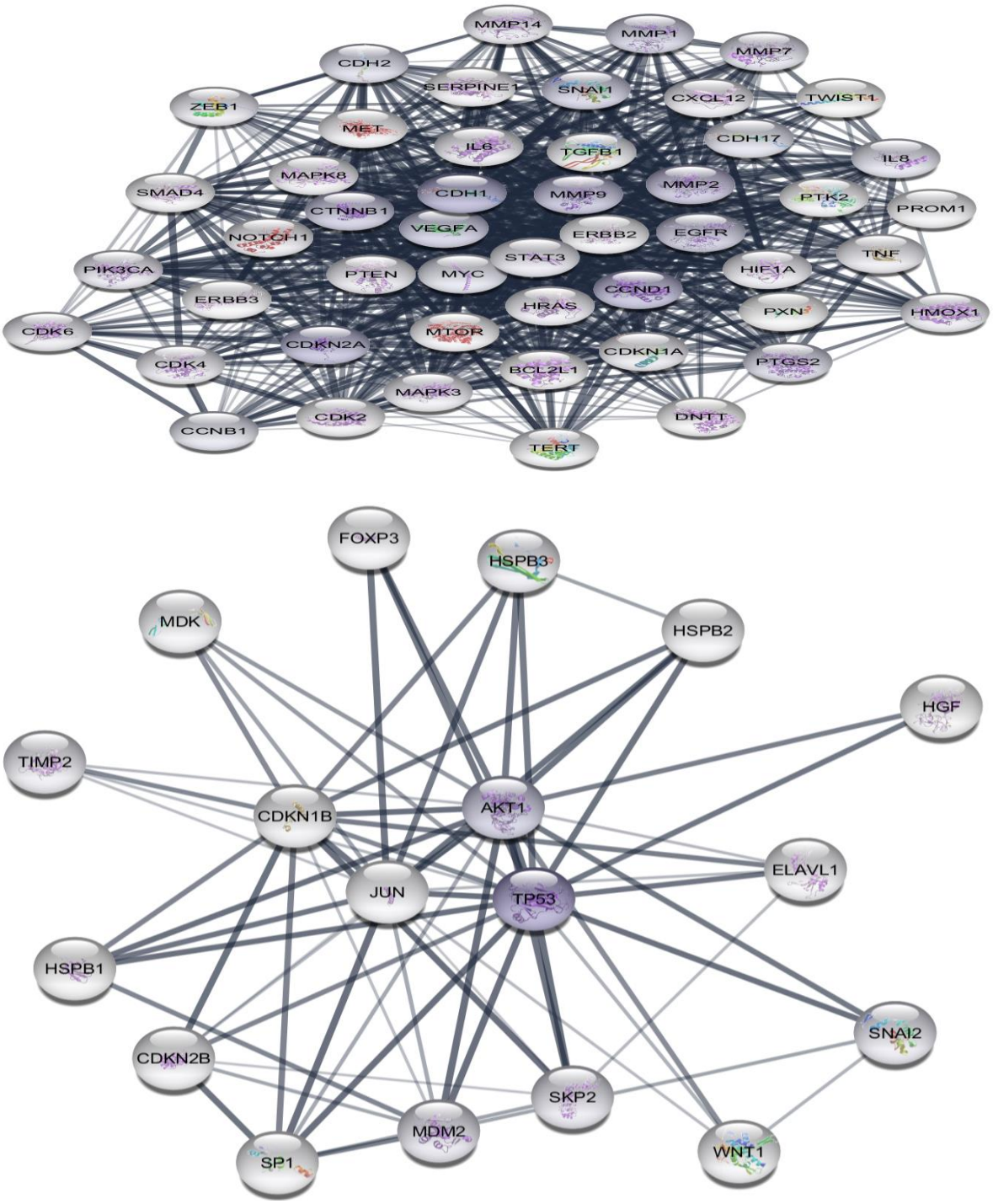

Figure 5. The two first ranked clusters obtained by MCODE. The color changes indicate the disease score of each node, as the association of that specific protein with oral squamous cell cancer increases from light to dark. The first complex score is 34 and the second complex score is 7. The seed protein in the first cluster is HIF1A and in the second cluster is TP53. As the node colors gets darker, their disease score increases.

esophageal adenocarcinoma (unpublished data). This finding possess some difficulties in differentiating between the two cancers. In the other hand, similar biomarker panels of the two diseases refer to the near origin of the two malignancies. It can be interpreted that the colon, gastric, esophagus adenocarinoma and

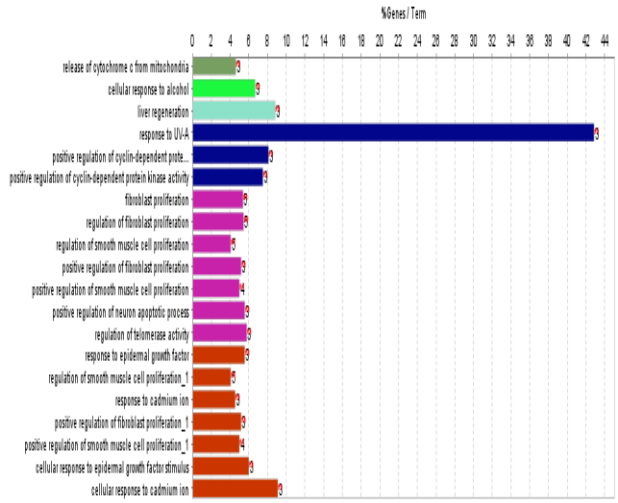

Figure 6. The associated terms are shown with different colors. Terms within each groups are colored similarly. The number and percentage of contributing genes for each term can be inferred. The importance of groups is assigned in a way that the groups with the highest number of terms and related genes are the most associated. This ranking is from bottom to up of the figure. The asterisk indicates the significance of the related terms and groups based on corrected p-Values. P-Value < 0.001 is assigned with ** and 0.001< P-Value < 0.05 is assigned with *. Kappa score = 0.4.

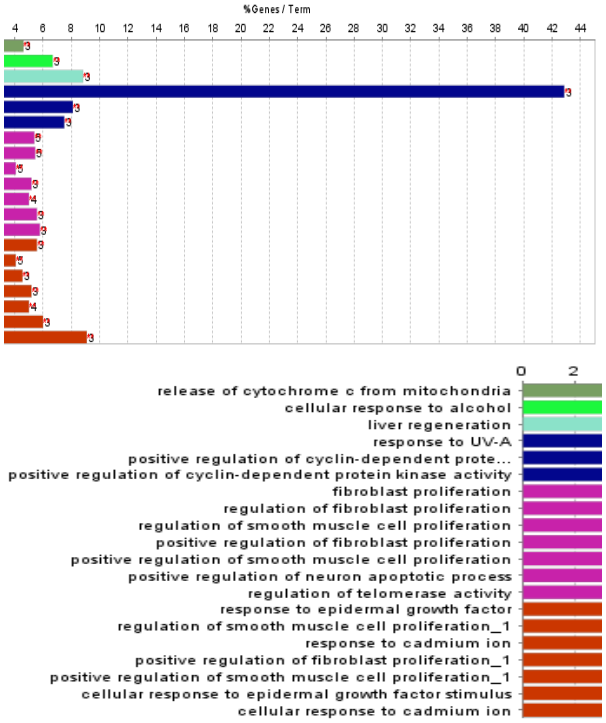

Figure 7. The detailed presentation of figure 6.

OSCC diseases may have a common set of biomarkers. The distribution of degree value is depicted in figure 3 and the trend shows distance of the TP53 from other central components and its significant place in the network. Furthermore, module analysis with ModuLand considered the network as an individual dependent on the presence of TP53, as can be inferred from figure 4. In addition, clustering analysis by MCODE explained the two highest complexes. The central nodes are distributed in the identified clusters. CDH1, CCND1, MYC, RGFR, and CTNNB1 are present in cluster 1 wherase TP53, AKT1, and JUN are present in the second cluster. This analysis validates the vital role of the eight central proteins. Gene ontology analysis showed that there are 20 biological processes related to crucial genes of the OSCC network. These genes are organized in six groups. Smooth muscle cell proliferation, regulation of fibroblast proliferation, and response to UV-A processes are highlighted as correlated biological processes to the essential proteins. It can be inferred that malfunction of the identified critical nodes can dysregulate the underlying processes. TP53 showed to be promising as it is the most central protein in the evaluated network. However, a need for considering other biomarkers for better oral cancer management suggests that the other seven proteins are worth examining as a panel for this malignancy.

In summary, the critical gene panel of oral squamous cell cancer was introduced. This panel is 87% similar to esophageal adenocarcinoma. It seems that any changes in this interacting profile may trigger massive modification in phenotype.

Acknowledgment

This study is related to the project NO.1395/76833 "from Student Research Committee, Shahid Beheshti University of Medical Sciences, Tehran, Iran. We also appreciate the "Student Research Committee" and "Research & Technology Chancellor" in Shahid Beheshti University of Medical Sciences for their financial support of this study.

Conflict of interests

The authors declare that they have no conflict of interest.

References

1. Pauls DL, Abramovitch A, Rauch SL, Geller DA. Obsessive-compulsive disorder: an integrative genetic and

neurobiological perspective. Nat Rev Neurosci 2014; 15: 410-24.

2. Yakob M, Fuentes L, Wang MB, Abemayor E, Wong DT. Salivary biomarkers for detection of oral squamous cell carcinoma - current state and recent advances. Curr Oral Health Rep 2014; 1: 133-141.

3. Gallo C, Ciavarella D, Santarelli A, Ranieri E, Colella G, Lo Muzio L, et al. Potential Salivary Proteomic Markers of Oral Squamous Cell Carcinoma. Cancer Genomics Proteomics 2016; 13: 55-61.

4. Nooshinfar E, Bashash D, Safaroghli-Azar A, Bayati S, Rezaei-Tavirani M, Ghaffari SH, et al. Melatonin promotes ATO-induced apoptosis in MCF-7 cells: Proposing novel therapeutic potential for breast cancer. Biomed Pharmacother 2016; 83: 456-465.

5. Dave JH, Vora HH, Ghosh NR, Trivedi TI. Mediator of DNA damage checkpoint protein 1 (MDC1) as a prognostic marker for patients with oral squamous cell carcinoma. J Oral Pathol Med 2017; 46: 253-258.

6. Zamanian–Azodi M, Rezaei–Tavirani M, Hasanzadeh H, Rahmati Rad S, Dalilan S. et al. Introducing biomarker panel in esophageal, gastric, and colon cancers; a proteomic approach. Gastroenterol Hepatol bed bench 2015; 8: 6-18.

7. Rezaei-Tavirani M, Rezaei-Tavirani M, Mansouri V, Rostami-Nejad M, Valizadeh R, Mahdavi SM, et al. Introducing crucial protein panel of gastric adenocarcinoma disease. Gastroenterol Hepatol bed bench 2017; 10: 21-28.

8. Ardakani MJE, Safaei A, Oskouie AA, Haghparast H, Haghazali M, Shalmani HM, et al. Evaluation of liver cirrhosis and hepatocellular carcinoma using Protein-Protein Interaction Networks. Gastroenterol Hepatol Bed Bench 2016; 9: 14–22.

9. Safari-Alighiarloo N, Rezaei-Tavirani M, Taghizadeh M, Tabatabaei SM, Namaki S. Network-based analysis of differentially expressed genes in cerebrospinal fluid (CSF) and blood reveals new candidate genes for multiple sclerosis. Peer J 2016; 4: 1-22.

10. Okhovatian F. Towards understanding topological features of protein interactome map of muscle cancer. Arvand Journal Health Medical Sciences 2017;1.

11. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res 2016;45:362-368.

12. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2014; 43: 447-452.

13. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC bioinformatics 2003; 4: 2. 1-27.

14. Kovács IA, Palotai R, Szalay MS, Csermely P. Community landscapes: an integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics. PloS one. 2010;5(9):1-14.

15. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 2009 ; 25: 1091-3.

16. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics 2013; 29: 661-3.

17. Jiang S, Dong Y. Human papillomavirus and oral squamous cell carcinoma: A review of HPV-positive oral squamous cell carcinoma and possible strategies for future. Curr Probl Cancer 2017: S0147-0272(16)30203-3.

18. Minhas S, Kashif M, AItaf W, Nagi A. Oral Squamous Cell Carcinoma Epidemiological. RMJ 2016;41:81-5.

19. Khanna V, Karjodkar F, Robbins S, Behl M, Arya S, Tripathi A. Estimation of serum ferritin level in potentially malignant disorders, oral squamous cell carcinoma, and treated cases of oral squamous cell carcinoma 2017.

20. Zali H, Rezaei Tavirani M. Meningioma protein-protein interaction network. Arch Iranian med 2014; 17: 262-72.

21. Joerger AC, Fersht AR. The p53 Pathway: Origins, Inactivation in Cancer, and Emerging Therapeutic Approaches. Annu Rev Biochem 2016; 85: 375-404.

22. Coates P. p53–The Gene That Cracked the Cancer Code. J Pathol; 2015. 236: 395.

23. Carneiro A, Isinger A, Karlsson A, Johansson J, Jönsson G, Bendahl P-O, et al. Prognostic impact of array-based genomic profiles in esophageal squamous cell cancer. BMC Cancer 2008; 8: 1-9.

24. Essack M. Transcription regulation and candidate diagnostic markers of esophageal cancer. Western Cape: University of the Western Cape: 2009.

25. Moroni M, Veronese S, Benvenuti S, Marrapese G, Sartore-Bianchi A, Di Nicolantonio F, et al. Gene copy number for epidermal growth factor receptor (EGFR) and clinical response to antiEGFR treatment in colorectal cancer: a cohort study. Lancet Onco 2005; 6: 279-86.

26. Vivanco I, Sawyers CL. The phosphatidylinositol 3-kinase–AKT pathway in human cancer. Nat Rev Cancer 2002; 2: 489-501.

27. Dehm S, Senger M-A, Bonham K. SRC transcriptional activation in a subset of human colon cancer cell lines. FEBS lett 2001; 487: 367-71.

28. Dunn EF, Iida M, Myers RA, Campbell D, Hintz K, Armstrong EA, et al. Dasatinib sensitizes KRAS mutant colorectal tumors to cetuximab. Oncogene 2011; 30: 561-74.