Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 410 – 417
Received 25 April 2021; Accepted 08 May 2021.

Protein Gene Enrichment and Network Analysis of Genes in Esophageal Cancer

Devindren Subramaniam[1], Suresh Kumar[2]*

[1,2]Faculty of Health Sciences, Management and Science University, Shah Alam, Selangor, Malaysia

ABSTRACT

Esophageal cancer occurs when the uncontrolled growth of abnormal cells in the food tube (oesophagus). Esophageal cancer is one of the most prevalent malignant tumors in the food tract with worldwide distribution due to late clinical development, rapid growth, and very poor survival. Also, the reason for this poor prognosis is that Esophageal cancer normally indicates widespread local tumor invasion and frequent spread to metastatic locations, especially regional lymph nodes. The objective of this study is to analyse all genes involved in protein-protein interaction through a systems biology approach and to identify key genes involved in Esophageal cancer. There is a total of 108 genes that causes Esophageal cancer and a protein-protein network was constructed and analyzed using these genes. The genes are obtained from the UniProt database using STRING and Cytoscape 3.7.1 tools. Moreover, the functional enrichment analysis was done to identify the key genes by using CentiScape and FUNRICH database. RBM8A, ZWINT, EIF4A3, BRCA2, GNB2L1 were recognized as the main genes in the network based on network topology parameters. Among them, the EIF4A3 gene with the greatest betweenness centrality and node degree was acquired as a super hub gene. The assessment of functional enrichment was carried out using the FUNRICH database. Conclusion From this study, the Eukaryotic initiation factor 4A-3 (EIF4A3) identified as one of the key genes in Esophageal cancer. This result suggests the potential role for EIF4A3 to serve as a diagnostic marker or therapeutic target for Esophageal cancer.

Keywords
Esophageal cancer, Esophageal Neoplasms, Systems Biology, STRING, Cytoscape

INTRODUCTION

Esophageal cancer is a disease that begins in the food pipe. Hence, the food pipe is also known as esophagus or gullet. Esophagus cancer is one of the most common malignant tumors of the food tract, with aglobal distribution. Esophageal cancer is characterized by late clinical presentation, fast progression, and very bad survival. (1).
Many previous studies shown that common people are lack of awareness of cancer symptoms due to iinsufficientnsuccient knowledge.(2, 3).
In Esophageal cancer cells, specific peripheral benzodiazepine receptor (PBR) ligands are known to cause apoptosis and cell cycle arrest. The fundamental mechanisms remain unknown. Here, in reaction to PBR-specific ligands, able to explore the transcriptional changes and activation of protein kinases. FGIN-1-27 caused comprehensive modifications in gene expression engaged in regulating apoptosis and the cell cycle in Esophageal cancer cells. There are many previous computational studies in identification of key genes for disease (4). The objective of this study to identify key genes for causing Esophageal cancer.

METHODOLOGY

2.1 Data Collection

There were 108 protein-protein interactions found to be involved in Esophageal cancer were retrieved from Uniprot database by using Cytoscape 3.7.1 as shown in figure 1.

2.2 Building a network using all the Cytoscape genes involved

Protein-protein interaction was obtained by using STRING. Moreover, the network of Esophageal cancer was constructed with the highest confidence level which is 0.7 to provide functional associations and the interaction level with 40, this is because to obtain more closely related genes to the targeted protein. Furthermore, for a clearer view, the PPI network was constructed by STRING and able to view the network from Cytoscape 3.7.1. The intention of creating this PPI network is to identify the closest gene of the interaction.

2.3 Analysis of PPI network

CentiScape2.2 was downloaded from Cytoscape 3.7.1 to analyze the protein-protein interactions. Hence, the reason for using CentiScape2.2 is mainly to identify the betweenness centrality and the node degree. Moreover, the results

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 410 – 417
Received 25 April 2021; Accepted 08 May 2021.

of betweenness centrality were used to identify the number of shortest paths passing through each node. The node degree, however, was used to define the number of node-related edges.

2.4. Topology Analysis of PPI network

PPI data provided a chance to systematically evaluate the topology of such a big network for functional information using multiple graph theory-based methods and use this to build models to predict essentiality, genetic interactions, function, protein complexes, and cellular processes. Gene represents the node in the network and edges represent the node-to-node interaction. The degree indicates no edges linked to nodes; the highest level of nodes is an important biological function. Betweenness centrality defines the node's significance based on the number of shorter paths passing through each node (5).

2.5 Functional Enrichment Analysis

FunRich software was downloaded and used to carry out the functional enrichment analysis of the nodes in the protein-protein interaction network. Hence, the selected five genes were transferred to FunRich software while the FunRich software will analyze the enriched nodes (6).

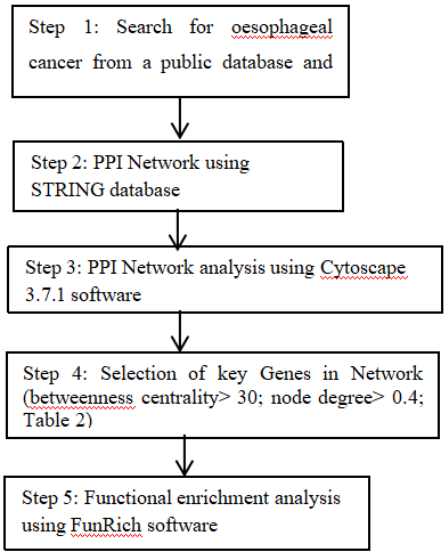

Figure 1: Flow chart of the methodology.

RESULT & DISCUSSION

To know a node is a connection point or not, the parameters for individual nodes must be calculated. Hence, the entire parameter of the network will be calculated, the degree indicates the number of nodes linked to the single node. When there is a higher degree it shows characteristics of a hub which means connection point. Similarly, the nearer a node is to all other nodes, the more central it is (proximity centrality). Higher centrality of proximity represents a node's tendency to be a connection point.

3.1 Protein-protein interaction network

The purpose of the STRING database is to provide critical evaluation and integration of protein-protein interactions, including direct (physical) and indirect (functional) associations. Each node includes another element of the functional role of the genes concerned. Hence, there were 108 genes of Esophageal cancer and by using the STRING database, the PPI network was constructed where it forms a total of 134 nodes and 1978 interactions including the parameters where the confidence score and the maximum additional interactors were set into 0.7 and 40 respectively.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 410 – 417
Received 25 April 2021; Accepted 08 May 2021.

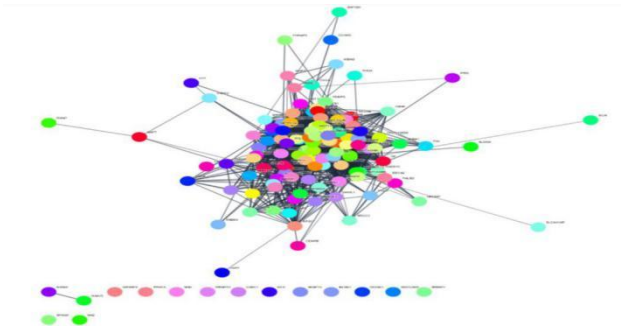

Figure 2:PPI Network overview built using STRING through the Cytoscape database. The network consists of 1978 edges (interaction) between 134 nodes based on a confidence score of 0.7 and the maximum additional interactors at 40 respectively.

3.2 Hub gene identification

The network was analyzed using the CentiScape app in CYTOSCAPE 3.7.1 software after creating a network of protein-protein interactions (Figure 2). Moreover, by implementing this step we can identify the degree and betweenness of the genes involved in the undirected network. Furthermore, the top five genes were chosen which full fill the requirements by having the betweenness centrality of more than 30 and node degree of more than 0.70 as shown in the table 1 below. These genes are considered to be important genes in Esophageal cancer protein. Next, based on the result obtained, EIF4A3 was obtained as a super hub gene with the highest Betweenness and Node degree.

Table 1: Table shows the key genes with "betweenness centrality" and "node degree".

Gene	Betweenness centrality	Node degree
RBM8A	57.31	25.00
ZWINT	61.73	20.00
EIF4A3	**90.89**	**26.00**
BRCA2	79.71	26.00
GNB2L1	69.01	8.0

3.3 Functional annotation gene

FunRich instrument enables both functional enrichment and network interaction analysis to be performed by users. Furthermore, the FunRich database is purely human-specific and users can use human genes/proteins to conduct functional enrichment analyzes against it. Hence, by using FunRich, users can recognize enhanced and depleted biological process conditions, cellular components, molecular functions, biological pathways, protein domains, expression sites, clinical phenotypes, and transcription factors.

Moreover, based on the results, Figure 3 shows about 20% of genes involved for the cellular component are stored secretory granules. The secretory granules are aligned to argyrophilic cells or argyrophilic granules which are usually discovered in the Esophageal mucosa basal cells and constitute a separate histopathological entity distinguishable from other kinds of Esophageal carcinomas (7). Nucleolus plays a significant role in the control of protein synthesis and cell proliferation. The Nucleolar Organiser Region's (NORs), strongly linked to nucleoli, are loop DNA encoded for the manufacturing of rRNA (8). The argyrophilicnucleolar organizer region (AgNOR) is formed from associating with proteins, where the number of AgNOR was proposed to correlate with both cellular kinetics and tumor malignancy (9). Next, the exon-exon junction complex with 20% genes involved. Homozygous deletion in 16 cell lines of the p15 gene. In another cell line, exon 1a of the p16 gene associated with the absence of a transcript, KYSE 1250, harbored a 32-bp deletion (10). BRCA2 is a protein that has a variety of functions. Thus, BRCA2 and MAGE-D1 expression synergistically suppress cell proliferation regardless of the p53 pathway. Tumorigenesis may involve the suppression of MAGE-D1 expression (11). Polo-like kinase 1 (Plk1) is a kinase in centrosome duplication or maturation; consequently, its deregulation results in enhanced centrosome size and/or centrosome numbers, indicating significant aneuploidy and chromosome instability correlation.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 410 – 417
Received 25 April 2021; Accepted 08 May 2021.

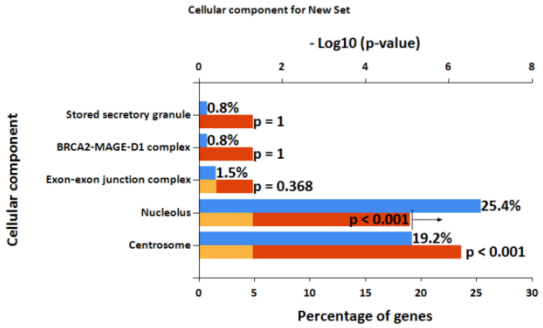

Figure 3: Cellular components of the key genes for Esophageal cancer.

Furthermore, as shown in Figure 4 will display the molecular function of the key genes. A variety of distinct mutant p53 proteins are generated in human cancers; this is especially true for Esophageal cancer, where mutations are widespread across the p53 coding region, including missense, absurdity, big and tiny deletions, and splicing mistakes arising from intronic mutations, and happen in both evolutionarily preserved and non-conserved areas (13, 14). Scaffolding proteins are characterized as proteins that combine various proteins to form functional complexes and therefore modulate cellular signaling pathways (15).

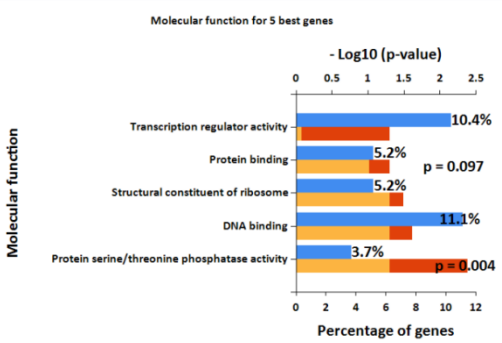

Figure 4: Molecular function of the key genes for Esophageal cancer.

Furthermore, Figure 5 shows the biological process for the key genes. For the majority of human cancers analyzed by immunohistochemistry, intratumoral hypoxia and genetic alterations that dysregulate signal transduction pathways resulting in drastic overexpression of HIF-1 alpha which causes Esophageal cancer and many other cancers (16). Besides, the following will be regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism. Hence, tumor suppressor gene mutations encoding p53 eliminate the control point of the cell cycle and increase the frequency of genomic rearrangements. The cancer cell genome reorganization, which allows tumor cells to overcome ordinary strictures against excessive multiplication and metastasis, may, therefore, be due to abnormal cell cycle control (17).

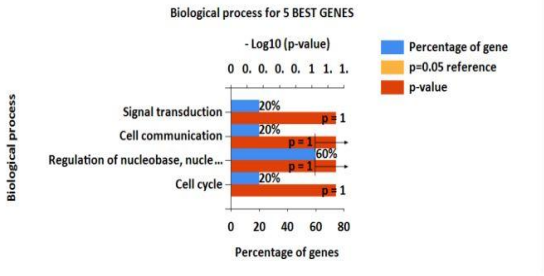

Figure 5: Biological process of the key genes for Esophageal cancer.

Besides, Figure 6 shows the biological pathways for the top 5 key genes. Meanwhile, P63 over-expression seems to involve autonomous processes of genomic amplification. In any case, countless trials have shown p63

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 410 – 417
Received 25 April 2021; Accepted 08 May 2021.

overexpression in up to 80% of main head and neck squamous cell carcinomas (HNSCCs), and p63 overexpression is also common in other squamous epithelial diseases, including lung, nasopharyngeal, esophageal, and cervical diseases (18). Molecular suppression of NMD across a range of cell lines has surprisingly disclosed that a range of non-mutated transcripts is also upregulated with NMD suppression, including cell cycle genes, differentiation, signaling, and RNA splicing (4). Moreover, NMD objectives include a broad range of transcripts, including tumorigenesis-relevant ones. With the recognition that NMD is a physiologically controlled method, we argued that identifying extra transcripts stabilized by NMD inhibition could suggest biological features that are affected by NMD regulation (1). The next one will be, p73 transcription factor network (18).

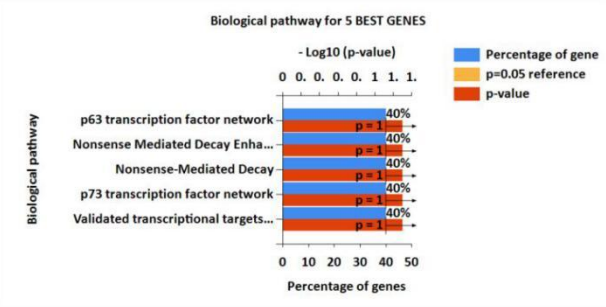

Figure 6: Biological pathway for the key genes of Esophageal cancer.

On the other hand, Figure 7 is about to show the site of expression of the key genes. Thus, there are 5 sites in total which has a different percentage of gene involvements. Moreover, because of its similarities to ordinary Esophageal mucosa, oral mucosa is an alternative cell source. Also, the cultivated oral mucosal epithelial cell sheets resemble the indigenous Esophageal epithelium with a stratified, squamous, non-keratinized framework, and comparable cytokeratin expression.

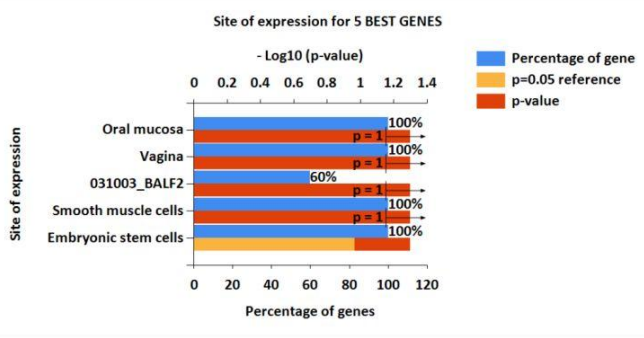

Figure 7: Site of expression for the key genes of Esophageal cancer

Furthermore, there is 5 transcription factor as shown in Figure 8. Hence, the first one will be GABPA with 50% of genes involved (Figure 8). GABPA may be the main transcriptional variables that mediate PGC-1 muscle action. Moreover, GABPA genes indicate that each contains prospective binding sites within the vicinity of their promoters for both transcription factors. Next, STRA13 with 25% of Esophageal cancer key genes. Stra13, a fundamental transcription factor for helix-loop-helix, is up-regulated when CD4+T cells are activated. ELK1 is a key regulator of immediate-early genes, such as FOS, quickly, and transiently caused after exposure to extracellular ligands activating the MAP kinase pathways (19).

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 410 – 417
Received 25 April 2021; Accepted 08 May 2021.

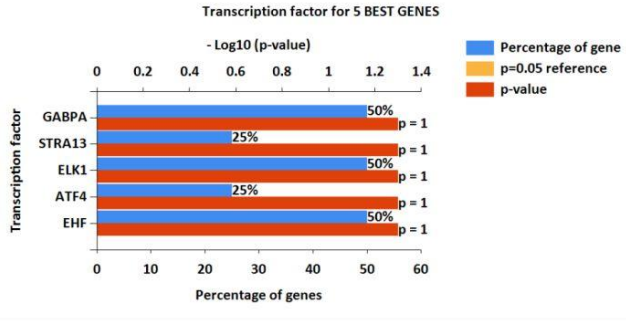

Figure 8: Transcription factor for the key genes of Esophageal cancer

Figure 9 shows the clinical phenotype for the key genes, p-value, and fold where they show the same percentage of the key genes and value of the p-values and fold present in 5 different phenotypes. Hypoplastic fibula (rare) has a key gene of 33.3%, p-value 0, and fold 604.7. Fibular hypoplasia happens with fibular hemimelia as its most serious type in varying degrees of severity (20). The distal epiphysis plate of the fibula is normally the same as the distal end of the distal tibial epiphysis, whereas the tip of the proximal fibular epiphysis is the same as the proximal tibial epiphysis plate (21). Hypoplastic fifth finger the phalanges concerned was generally both shortened and narrowed. However, some phalanges were normal but abnormally thin in length. Then, tibial torsion is defined as any twist on the longitudinal axis of the tibia that results in a shift in the alignment of the proximal and distal articulations. The knee and ankle mainly work as joints of the hinge.

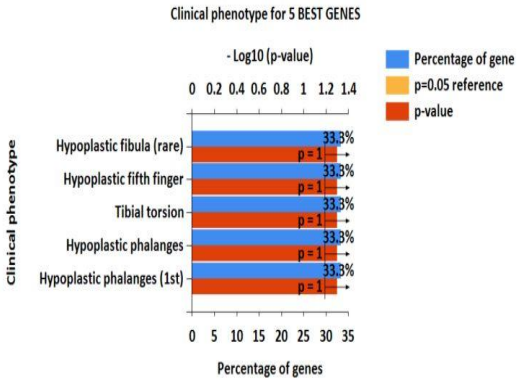

Figure 9: Clinical phenotype for the key genes of Esophageal cancer

Moreover, Figure 10 shows the COSMIC of the key genes involved. Hence, the pancreas has the highest which is 100% key genes involved, p-value 0 and fold 1.5. Besides, the second highest will be ovary with 80% of genes involved, p-value 0 and fold 1.2. In contrast, the salivary gland and cancer gene census gene list having the same percentage which is 20%. The salivary gland has a p-value of 0 and folds 6.8. Meanwhile, the census gene list has a p-value of 0 and fold 7.4.

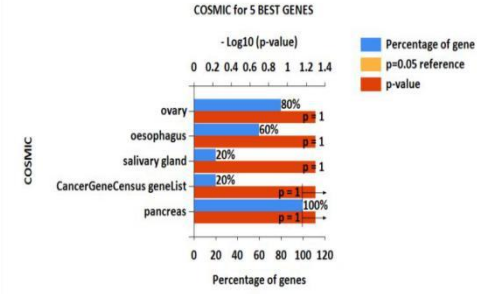

Figure 10: COSMIC for the key genes of Esophageal cancer

Furthermore, based on the result obtained protein Eukaryotic Initiation Factor 4A3 which has a gene EIF4A3 will be clarified as the hub gene or the highly interconnected molecule nodes that have the highest value of betweenness Centrality and degree node.

This gene encodes a protein family member of the DEAD-box. DEAD-box proteins are putative RNA helicases, defined by the preserved motif Asp-Glu-Ala-Asp (DEAD). They are involved in several cellular procedures involving altering the secondary structure of RNA, such as initiation of translation, nuclear and mitochondrial splicing, and assembly of ribosome and spliceosome.

Furthermore, this gene in a multitude of cancers including lung cancer, breast cancer, and gastric cancer, H19 expression is involved in solid tumors. To encourage the development of CRC, H19 recruits eukaryotic translation initiation factor 4A3 (eIF4A3). Besides, tumor differentiation and tumor node metastasis (TNM) is linked with elevated expression of H19. Moreover, this gene also involves many other mutations in our human body such as non-coding expansion in EIF4A3 can cause Richieri-Costa-Pereira Syndrome, a Limb-related craniofacial disorder.

EIF4A3 changed into overexpressed in commonplace malignancies on the transcription tiers. High incidences of breast, lung, and urinary cancers had been closely associated with the prognostic index for survival. Thus, EIF4A3 additionally has high incidences of Esophageal cancer. The most familiar mutation in EIF4A3 became E59K/Q. The tumor necrosis aspect-α (TNF-α)/nuclear aspect-κB (NF-κB) signaling pathway was suffering from those mutations. In a nutshell, a conclusion can be made that this gene is the hub gene for Esophageal cancer that plays a major role in causing many deformations and diseases in our human body.

CONCLUSION

The aim is to use protein-protein network analysis to define the targeted gene and the pathway against Esophageal cancer. The hub genes identified in this study will be helpful for treatment and prognosis of Esophageal cancer patients.

REFERENCES

[1] Nonsense-Mediated RNA Decay by the Tumor Microenvironment Promotes Tumorigenesis. *Molecular and Cellular Biology*,*31*(17), 3670-3680. doi:10.1128/mcb.05704-11

[2] Al-Naggar, R. A., & Chen, R. (2012). Practice and barriers towards cervical cancer screening among university staff at a Malaysian university. Journal of Community Medicine & Health Education, 2(1), 120.

[3] Dubai, S. A. R. A., Ganasegeran, K., Alabsi, A. M., Alshagga, M. A., & Ali, R. S. (2012). Awareness and knowledge of oral cancer among university students in Malaysia. Asian Pacific Journal of Cancer Prevention, 13(1), 165-168.

[4] Wee, J. J., & Kumar, S. (2020). Prediction of hub genes of Alzheimer's disease using a protein interaction network and functional enrichment analysis. Genomics & Informatics, 18(4).

[5] Przulj, N., Wigle, D., &Jurisica, I. (2004). Functional topology in a network of protein interactions. *Bioinformatics*,*20*(3), 340-348. doi:10.1093/bioinformatics/btg415

[6] Pathan, M., Keerthikumar, S., Ang, C., Gangoda, L., Quek, C. Y., Williamson, N. A., . . .Mathivanan, S. (2015). FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics*,*15*(15), 2597-2601. doi:10.1002/pmic.201400515

[7] Kuwano, H., Sumiyoshi, K., Nozoe, T., Yasuda, M., Watanabe, M., &Sugimachi, K. (1995). The prognostic significance of the cytophotometric DNA content and its relationship with the argyrophilicnucleolar organizer regions (AgNOR) and proliferating cell nuclear antigen (PCNA) in Esophageal cancer. *European Journal of Surgical Oncology (EJSO)*,*21*(4), 368-373. doi:10.1016/s0748-7983(95)92384-5

[8] Luo, J., Qu, J., Wu, D., Lu, Z., Sun, Y., &Qu, Q. (2017). Long non-coding RNAs: A rising biotarget in colorectal cancer. *Oncotarget*,*8*(13). doi:10.18632/oncotarget.14728

[9] Tateishi, R., Taniguchi, K., Horai, T., Iwanaga, T., Taniguchi, H., Kabuto, T., . . . Wada, A. (1976). Argyrophil cell carcinoma (apudoma) of the esophagus. *VirchowsArchiv A Pathological Anatomy and Histology*,*371*(4), 283-294. doi:10.1007/bf00428049

[10] Tanaka, H., Shimada, Y., Imamura, M., Shibagaki, I., &Ishizaki, K. (1997). Multiple types of aberrations inthep16 (INK4a) and thep15(INK4b) genes in 30 esophageal squamous-cell-carcinoma cell lines. *International Journal of Cancer*,*70*(4), 437-442. doi:10.1002/(sici)1097-0215(19970207)70:43.0.co;2-c

[11] Rai, D., Li, J., Zou, C., Bai, Y., Wazer, D., Band, V., &Gao, Q. (2005, June 01). BRCA2 Suppresses Cell Proliferation via Stabilizing MAGE-D1. Retrieved from https://cancerres.aacrjournals.org/content/65/11/4747.short

[12] Sánchez-Pérez, I., Alonso, P. G., &Iniesta, C. B. (2009). Clinical impact of aneuploidy on gastric cancer patients. *Clinical and Translational Oncology,11*(8), 493-498. doi:10.1007/s12094-009-0393-z
[13] Horan, K., Jang, C., Bailey-Serres, J., Mittler, R., Shelton, C., Harper, J. F., . . .Girke, T. (2008, May). Annotating genes of known and unknown function by large-scale coexpression analysis. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2330292/
[14] Unger, T., Nau, M., Segal, S., &Minna, J. (1992). P53: A transdominant regulator of transcription whose function is ablated by mutations occurring in human cancer. *The EMBO Journal,11*(4), 1383-1390. doi:10.1002/j.1460-2075.1992.tb05183.x
[15] Walther, Cornelia, Ferguson, & G., S. S. (2015, June 01). Minireview: Role of Intracellular Scaffolding Proteins in the Regulation of Endocrine G Protein-Coupled Receptor Signaling. Retrieved from https://academic.oup.com/mend/article/29/6/814/2556266
[16] Ohki, T., Yamato, M., Maeda, M., Murakami, D., Okano, T., & Yamamoto, M. (2007). A Novel Inflatable Surgical Device Used in Endoscopic Transplantation of Tissue Engineered Cells Sheets for Regenerative Treatments of Esophageal Ulcerations: A Preliminary Report. *Gastrointestinal Endoscopy,65*(5). doi:10.1016/j.gie.2007.03.655
[17] Hartwell, L. (1992). Defects in a cell cycle checkpoint may be responsible for the genomic instability of cancer cells. *Cell,71*(4), 543-546. doi:10.1016/0092-8674(92)90586-2
[18] Deyoung, M. P., &Ellisen, L. W. (2007). P63 and p73 in human cancer: Defining the network. *Oncogene,26*(36), 5169-5183. doi:10.1038/sj.onc.1210337
[19] Boros, J., Donaldson, I. J., Odonnell, A., Odrowaz, Z. A., Zeef, L., Lupien, M., . . .Sharrocks, A. D. (2009). Elucidation of the ELK1 target gene network reveals a role in the coordinate regulation of core components of the gene regulation machinery. *Genome Research,19*(11), 1963-1973. doi:10.1101/gr.093047.109
[20] Minagi, H., & Steinbach, H. L. (1966). Roentgen Appearance Of Anomalies Associated With HypoplasticAnemias Of Childhood: Fanconis Anemia And Congenital Hypoplastic Anemia (ErythrogenesisImperfecta). *American Journal of Roentgenology,97*(1), 100-109. doi:10.2214/ajr.97.1.100
[21] Favaro, F., Alvizi, L., Zechi-Ceide, R., Bertola, D., Felix, T., De Souza, J., . . .Passos-Bueno, M. (2014). A Noncoding Expansion in EIF4A3 Causes Richieri-Costa-Pereira Syndrome, a Craniofacial Disorder Associated with Limb Defects. *The American Journal of Human Genetics,94*(1), 120-128. doi:10.1016/j.ajhg.2013.11.020