Original Article

Chin Med Sci J
2023; 38(1): 20-28
doi:10.24920/004148

Analysis of Significant Genes and Pathways in Esophageal Cancer Based on Gene Expression Omnibus Database

An-Yi Song[1,2†*], Lan Mu[1†], Xiao-Yong Dai[1], Li-Jun Wang[1], Lai-Qiang Huang[1*]

[1]The Shenzhen Key Laboratory of Gene and Antibody Therapy, State Key Laboratory of Chemical Oncogenomics, Tsinghua-Berkeley Shenzhen Institute (TBSI), Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, Guangdong Province, China
[2]Department of Chemistry, Tsinghua University, Beijing 100084, China

ABSTRACT

Objective To screen antigen targets for immunotherapy by analyzing over-expressed genes, and to identify significant pathways and molecular mechanisms in esophageal cancer by using bioinformatic methods such as enrichment analysis, protein-protein interaction (PPI) network, and survival analysis based on the Gene Expression Omnibus (GEO) database.

Methods By screening with highly expressed genes, we mainly analyzed proteins MUC13 and EPCAM with transmembrane domain and antigen epitope from TMHMM and IEDB websites. Significant genes and pathways associated with the pathogenesis of esophageal cancer were identified using enrichment analysis, PPI network, and survival analysis. Several software and platforms including Prism 8, R language, Cytoscape, DAVID, STRING, and GEPIA platform were used in the search and/or figure creation.

Results Genes *MUC13* and *EPCAM* were over-expressed with several antigen epitopes in esophageal squamous cell carcinoma (ESCC) tissue. Enrichment analysis revealed that the process of keratinization was focused and a series of genes were related with the development of esophageal cancer. Four genes including *ALDH3A1, C2, SLC6A1,* and *ZBTB7C* were screened with significant *P* value of survival curve.

Conclusions Genes *MUC13* and *EPCAM* may be promising antigen targets or biomarkers for esophageal cancer. Keratinization may greatly impact the pathogenesis of esophageal cancer. Genes *ALDH3A1, C2, SLC6A1,* and *ZBTB7C* may play important roles in the development of esophageal cancer.

Key words: GEO; esophageal cancer; antigen; enrichment analysis; survival curve; signaling pathway

INTRODUCTION

With the advancement of nucleic acid *in situ* hybridization (ISH) technology, gene-chip with large-scale assay of gene expression is available from bioinformatics databases such as The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). Information of clinical samples can be downloaded from these databases and analyzed with R language software using special packages such as Bioconductor and Limma. Based on several basic platforms and software, conventional methods including enrichment analysis, survival analysis, and protein–protein interaction (PPI) networks may be carried out to reveal important signaling pathways and mechanisms associated with a specific disease. Generally, as bio-technology develops, analysis of differentially expressed genes from bioinformatics databases will become an increasingly important approach[1-2].

Esophageal carcinoma is a common malignancy in digestive tract and accounts for over 300,000 deaths

Received August 1, 2022; accepted January 28, 2023; published online March 1, 2023.
†Co-first authors.
*Corresponding authors: An-Yi Song, E-mail: sonanyi0422@163.com; Lai-Qiang Huang, huanglq@tsinghua.edu.cn.

annually worldwide. Patients usually present with symptoms including dysphagia, weight loss, cough, hoarseness, and pain in esophagus. Studies have shown that smoking, alcohol, hot beverages, radiation, contaminated food, chronic irritation, genetic mutation, and dystrophia are risk factors for esophageal carcinoma[3]. Esophagogram, computed tomography (CT), endoscopic ultrasonography, and positron-emission tomography (PET) are useful diagnostic tools for this disease. The common treatments include chemotherapy, surgery, radiotherapy, and biological therapy.

In this study, information of gene-chip GSE 92396 was downloaded from GEO database for further analysis considering of sample uniformity[4]. The down-regulated genes underwent enrichment analysis to concentrated with specific pathways and molecular progress via STRING and DAVID website. Potential antigen targets, significant genes, and molecular mechanisms were demonstrated through analysis.

MATERIAL AND METHOD

Information on gene-chip GSE 92396

Considering the experimental conditions and sample uniformity, we chose the gene-chip GSE 92396 for analysis in this study. For GSE 92396, 12 esophageal adenocarcinoma (EAC) samples and 9 normal esophageal tissue samples were analyzed, including GSM2428973, GSM2428975, GSM2428976, GSM2428977, GSM2428978, GSM2428979, GSM2428980, GSM2428981, GSM2428982, GSM2428983, GSM2428984, GSM2428985, GSM2428986, GSM2428988, GSM2428989, GSM2428990, GSM2428991, GSM2428992, GSM2428993, and GSM24289894. Output data included *P* value, logFC (Fold Change), gene names, and gene titles in one list. By setting with the filter of *P* value (*P*>0.05, or *P*<-0.05), 4,316 over-expressed genes and 4,112 down-regulated genes were discriminated.

Websites and platforms

Original data of gene expressions were outputted and visualized with R language software from GEO database. Through STRING and DAVID websites, the remarkably down-regulated genes underwent enrichment analysis and PPI network analysis[5-6]. Survival curve of each gene was analyzed with Gene Expression Profiling Interactive Analysis (GEPIA) website, which was based on clinical data from TCGA database[7].

Based on the immunological platform Immune Epitope Database (IEDB), antigen epitope of specific protein was searched for potential antigen targets[8]. The transmembrane domains of genes *MUC13* and *EPCAM* were also analyzed from the TMHMM website. Genecards and Kyoto Encyclopedia of Genes and Genomes (KEGG) website were used to demonstrate information on gene function and signaling pathways[9].

In this research, we mainly used methods of enrichment analysis, PPI network, and survival analysis to discuss about significant genes and pathways. Software of Prism 8, R language and cytoscape were used to draw figures. The result of enrichment result was plotted with R software through ggplot package. Protein-protein interaction (PPI) network was analyzed from STRING website and reorganized with cytoscape software.

Statistical methods

The experimental sample and the normal sample was compared, with a *P* value of < 0.05 as statistically significant. Value of logFC was used to screen differentially expressed genes in cancer tissue. Survival curve was plotted with patients' lifespan and gene expression level. The significance of survival curve is discriminated by log rank *P* ($P < 0.05$) between normal group and cancer patients. A larger area under the receiver operating characteristic (ROC) curve (AUC) indicates a higher diagnostic value: the result performs a high experimental accuracy when AUC is between 0.7 and 0.9, and there is no diagnostic value when AUC equals to 0.5.

RESULTS

Potential immunological targets

Data of Gene-chip GSE 92396, which contains the total information of differentially expressed genes from cancer samples, were downloaded from GEO database. With a filter of |logFC|>2, 182 significantly over-expressed and 286 down-regulated genes were screened for further analysis. **Fig. 1** is the volcano map of 1790 DEGs (differentially expressed genes) filtered with the criteria of logFC and –log10 (*P* value).

Top 10 over-expressed genes are shown in **Table 1**. Gene *MUC13 (Mucin13)* can encode a group of transmembrane glycoproteins secreted by epithelial cells and it is also associated with the proliferation, migration, apoptosis, and adhesion of cancer cells. Gene

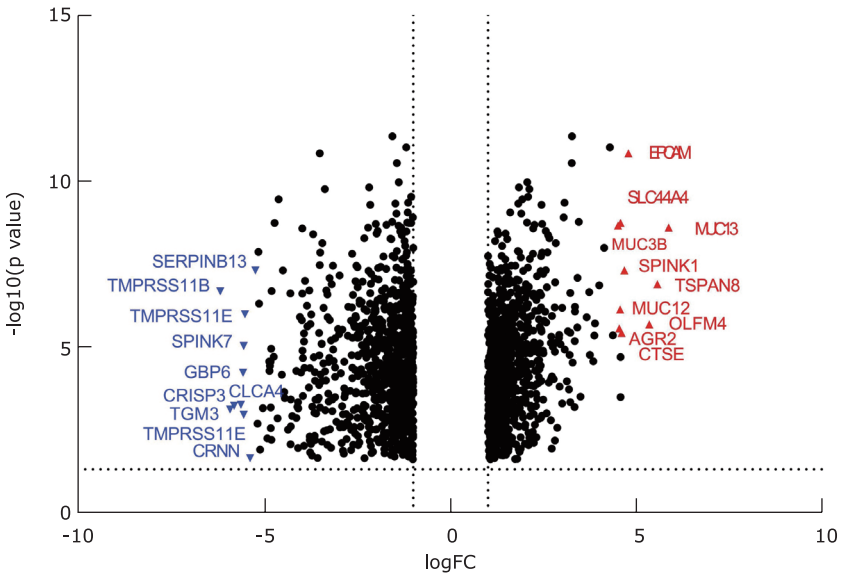

Figure 1. Volcano map of 1790 DEGs filtered with the criteria of log|FC|>1 and -log10 (*P* value)>1. Red dots represent highly expressed genes, and blue dots represent down-regulated genes as noted.

Table 1. Top 10 over-expressed genes from GSE 92396

ID	*P* value	Log FC	Gene	Gene title
8090180	2.57E-09	-5.84111565	*MUC13*	Mucin 13, cell surface associated
7964927	1.31E-07	-5.53654183	*TSPAN8*	Tetraspanin 8
7969288	2.14E-06	-5.3224096	*OLFM4*	Olfactomedin 4
8098439	1.48E-11	-4.7634735	*EPCAM*	Epithelial cell adhesion molecule
8114964	5.03E-08	-4.653026	*SPINK1*	Serine peptidase inhibitor, Kazal type 1
7909164	3.89E-06	-4.58246132	*CTSE*	Cathepsin E
8125149	1.85E-09	-4.55285303	*SLC44A4*	Solute carrier family 44 member 4
8135033	7.60E-07	-4.53832027	*MUC12*	Mucin 12, cell surface associated
8138381	2.84E-06	-4.51515693	*AGR2*	Anterior gradient 2
8135015	2.25E-09	-4.48195073	*MUC3B*	Mucin 3B, cell surface associated

MUC13 can also inhibit immune surveillance in cells and lead to immunological escape during cancer development. Mucins may function as potential targets for immunotherapy[10]. In addition, epithelial cell adhesion molecule *(EPCAM)* mainly encodes epithelial cell adhesion factor, which is involved in cell adhesion and proliferation. It is mostly over-expressed in several types of cancer tissue and can clinically be used as a cancer biomarker. High expression of *EPCAM* has been reported in major malignancies such as colon cancer, gastric cancer, prostate cancer, and lung adenocarcinoma. This pan-carcinoma antigen may be a promising target for immunotherapy[11].

Based on the TMHMM website, we also analyzed the transmembrane domains of proteins MUC13 and EPCAM. It was found that these genes could encode transmembrane proteins with long extracellular fragments, which could be identified as potential antigenic targets. Besides, genes *MUC13* and *EPCAM* were significantly over-expressed in esophageal cancer compared with normal group according to the GEPIA website ($p<0.05$). The original data of each gene was derived from TCGA database on the GEPIA website. **Fig. 2** is the boxplot showing the expression level of each gene in 182 cancer samples and 286 normal samples. The AUC values for *MUC13* and *EPCAM* were 0.99 and 1.0, respectively, suggesting the good diagnostic values of these two genes[12].

We also analyzed with antigen epitopes of proteins MUC13 and EPCAM from the IEDB website (http://www.iedb.org). Several antigen epitopes of *MUC13* and *EPCAM* that had been researched previously are listed by amino acid sequence in **Table 2**. Thereby these epitope sites may be identified by antibodies as prom-

ising targets for immunotherapy. Further analysis of over-expressed genes may help screen antigen targets of specific cancer for immunotherapy.

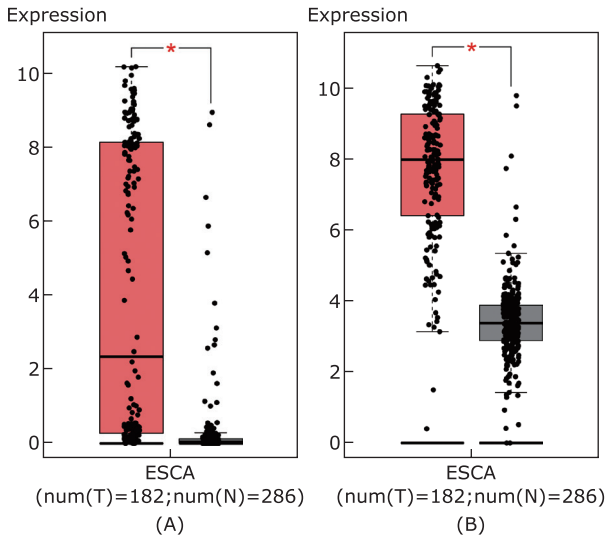

Figure 2. Expression levels of genes *MUC13* and *EPCAM* from GEPIA website. (A) The expression level of gene MUC13. (B) The expression level of gene EPCAM.

Red asterisk represents a significant *P* value between cancer tissue and normal tissue ($P<0.05$). ESCA: esophageal carcinoma.

Enrichment analysis

Enrichment analysis and PPI networks of 286 remarkably down-regulated genes were conducted via the STRING and DAVID websites to explore the potential impacts of mutations and functional defects of significant genes on the pathogenesis of esophageal cancer (**Fig. 3**). Genes *SPRR1A, SPRR2B, CSTA, KRT1, KRT6A, CNFN, SCEL, FLG, DSC2, DSG2, SERPINB3, SPINK5,* and *KLK12* were found to be associated with the process of keratinization[13]. Studies revealed that proteins encoded by above genes might be important components of keratin, desmosome, and filamentary bundle, or proteolysis enzyme which are related with keratinization. Down-regulation of these genes may lead to functional defects in skin barrier, which is crucial for the development of ESCC.

Table 2. Antigen epitopes of genes *MUC13* and *EPCAM*

Gene	Antigen epitopes
MUC13	(1) ACDCKSDLQRP; (2) GLACDCKSDLQRP; (3) DLHSEITSLFK; (4) INKAIRSSSSNFLNYD
EPCAM	(1) KPYDSKSLRTAL; (2) AATATFAAA; (3) DVDIADVAY; (4) RAKPEGALQNNDGLYD; (5) IIIELKHKAREK

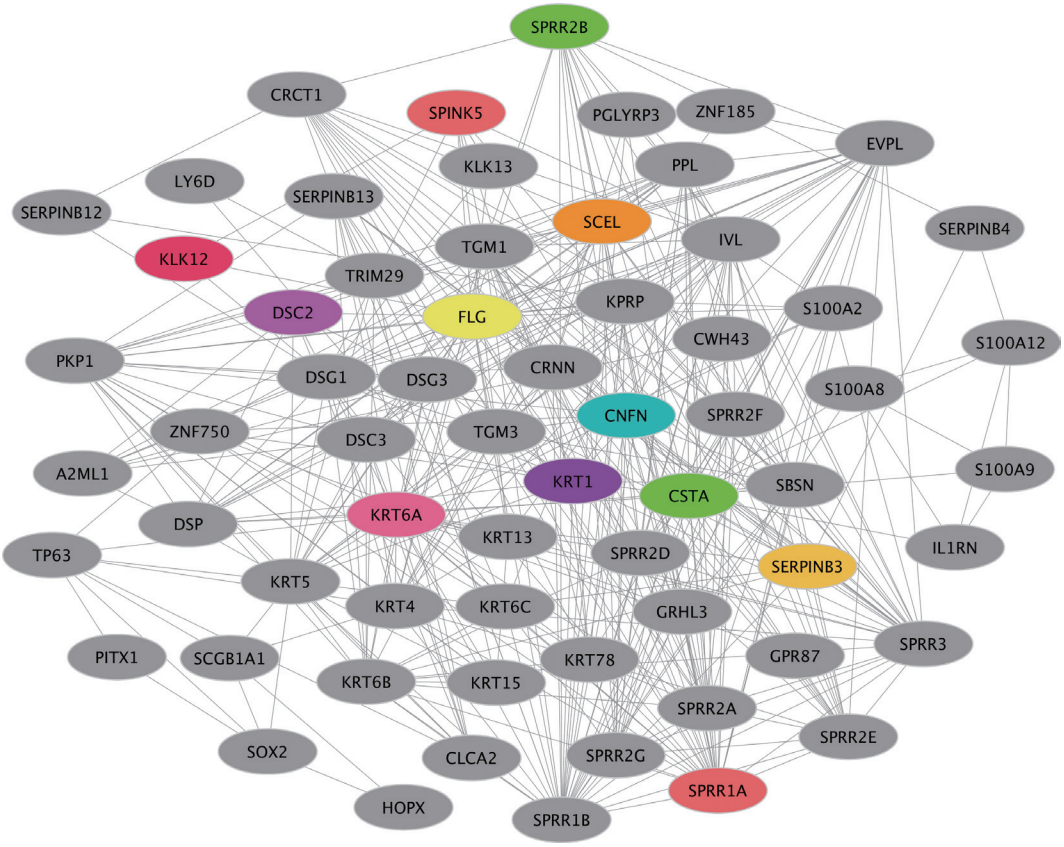

Figure 3. PPI network of significantly down-regulated genes (reorganized with Cytoscape software)

Skin is a barrier that can protect tissues and organs from damage caused by radiation as well as chemical and physical agents. In addition, skin barrier is important for the maintenance of water and electrolyte balance. Skin cuticle mainly consists of keratinocyte, angular desmosome, and extracellular matrix. Keratin is the main structural protein in keratinocyte and is rich with amino acid cysteine. It is usually combined with filaggrin to form cuticular membrane, which acts as a supportive structure in skin tissue. Abnormal expression of keratin will result in abnormal differentiation of keratinocytes. Enrichment analysis has suggested that functional defects in some important genes associated with keratinization are crucial for the development of esophageal cancer[14].

The results of our enrichment analysis of down-regulated genes are shown in **Fig. 4** and **Table 3**. Several processes such as keratinocyte differentiation, keratinization, establishment of skin barrier, and development of epidermis were significant for the pathogenesis of esophageal cancer, consistent with the results of enrichment analysis from the STRING website. Therefore, defects in keratinization and dysfunction in skin barrier might be crucial for the development of esophageal cancer.

From the PPI network of significantly down-regulated genes and enrichment analysis, many genes were found to be associated with keratinization. For instance, gene *CSTA* encodes a cysteine protease inhibitor and is important for the development and maintenance of epidermal tissue. Evidence also indicates that *CSTA* is closely related to psoriasis. Gene *FLG* can encode an intermediate silk protein, which forms filamentary bundles with keratin in mammalian epidermis[15]. Mutations in these genes may cause ichthyosis vulgaris. Gene *SPINK5* has great impact on the morphogenesis of skin and hair follicles and its mutation may result in Netherton syndrome, defective cornification, and atopy[16]. Proteins encoded by genes *DSC3* and *DSG3* are the main components of desmosome and consist of vital structure in cell-cell junction. Mutations in *DSC3* may cause hypotrichosis and recurrent disorders in skin vesicles[17]. Moreover, gene *DSG3* is a natural antigen of pemphigoid, and it is involved in the pathogenesis of squamous cell carcinoma. Thus, mutations in these genes can cause epithelial malignancies. Esophageal cancer may derive from tissue damage, infection, and gene mutations due to the impairments of the associated signaling pathways.

Survival analysis of significant genes

Survival curves were plotted from clinical records and indicated the relationships between gene expressions and survival time. A log rank *P* of <0.05 indicates that the gene may have significant impact on patients' lifespan, in addition to being crucial for the pathogenesis of esophageal cancer. Data of patients' survival time were downloaded from the GEPIA website, which was originated from the TCGA database[18]. By screening with significant log rank *P* values of survival curves, we selected four genes including *ALDH3A1, C2, SLC6A1* and *ZBTB7C* for further discussion (Figure 5).

Analysis based on the Genecards website indicated that gene Complement Component 2 *(C2)* was associated with defective clearance of apoptotic cells. Cell autophagy is an important biological process that maintains cellular homeostasis. Apoptotic cells resulting from the effects of UV light and infection must be cleared quickly. Defective clearance of apoptotic cells can lead to inflammatory reactions and autoimmune

Table 3. Enrichment analysis of down-regulated genes

Term	Count	%	*P* value	Benjamini
Epidermis development	23	9.5	3.10E-23	2.70E-20
Keratinocyte differentiation	22	9.1	6.80E-23	3.00E-20
Peptide cross-linking	16	6.6	2.80E-17	8.30E-15
Keratinization	15	6.2	5.10E-16	1.10E-13
Establishment of skin barrier	7	2.9	6.00E-08	1.10E-05
Negative regulation of endopeptidase activity	9	3.7	1.50E-04	2.20E-02
Negative regulation of peptidase activity	4	1.7	1.20E-03	1.50E-01
Wound healing	6	2.5	3.30E-03	3.70E-01
Proteolysis	15	6.2	4.40E-03	4.10E-01
Defense response to fungus	4	1.7	4.60E-03	4.10E-01

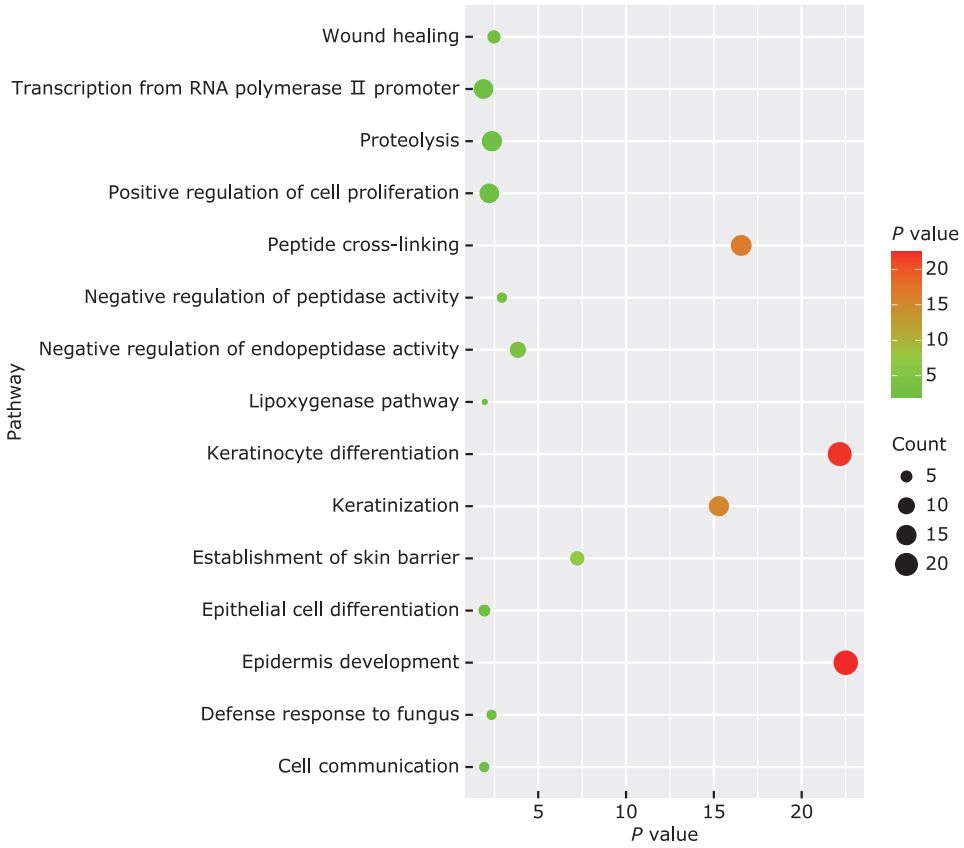

Figure 4. Enrichment analysis of 286 remarkably down-regulated genes with DAVID platform.

Significantly down-regulated genes were imported into DAVID website for enrichment analysis, and the results were plotted with R language software.

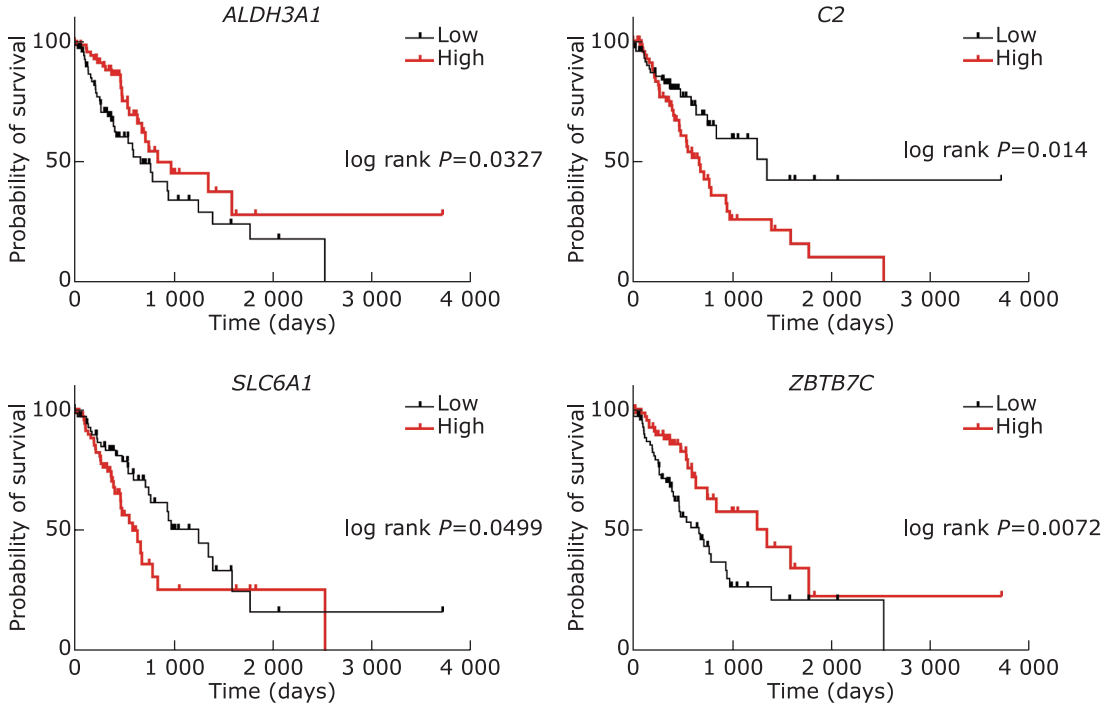

Figure 5. Survival curves of genes *ALDH3A1, C2, SLC6A1,* and *ZBTB7C.*

Red curve represents survival curve of cancer patients with high expression of a specific gene, and black curve for low expressed group. *P* value is the result of significance test between high expressed group and low expressed group.

disease. In patients with system lupus erythematosus, *C2, C1q*, and *C4* have great function for the mechanism of defective clearance. High expression of gene *C2* may deteriorate the condition (**Fig. 5**).

Survival analysis also showed that high expression level of gene *ALDH3A1* might extend the survival period of patients owing to its role in metabolic detoxification. In contrast, high expression level of gene *C2* might lead to defective clearance of apoptotic cells and deterioration of patients' conditions. Defects in gene *SLC6A1* might lead to neural disorders, and *SLC6A1* might be involved in the pathogenesis of colorectal cancer. Besides, it was also found that high expression of gene *ZBTB7C* (a tumor suppressor gene involved in the p53 signaling pathway) might lead to improvement in patients' condition. Thereby further analysis of significantly expressed genes and signaling pathways might be meaningful in revealing possible molecular mechanisms and screening targets for esophageal cancer treatment.

DISCUSSION

Recently, CAR-T (Chimeric Antigen Receptor T-Cell) therapy has been increasingly applied in cancer treatment. The fundamental principle of CAR-T therapy is to construct chimeric T-cell with specific antibody. The combination of chimeric T-cell with tumor antigen may trigger the downstream immunological effects, thus resulting in the elimination of tumor cells. The most significant step for experimental design is to select appropriate antigen targets with specificity and accuracy. Tumor antigens are proteins over-expressed on cancer cell membrane, with specific antigen epitopes. With the use of gene-chip, the expression level of each gene and ROC analysis curve can be plotted. In fact, the transmembrane fragment of each protein can be revealed based on the TMHMM website, and the epitope sequences of specific proteins can be searched from the IEDB website. Consequently, these proteins can be clinically applied as biomarkers and/or immunological antigens. Analysis of significantly over-expressed genes from GEO database may provide an effective method to screen potential antigen targets. Many new databases and software have been adopted in bioinformatic studies on the regulation of protein transcription and gene expression, which will further promote research in this filed.

Several studies have reported that *MUC13* and *EPCAM* are over-expressed on esophageal cancer, and therapeutic antibodies towards these proteins have already been applied in clinical settings. From previous research, analysis of gene expression indicated that *MUC13* and *EPCAM* were significantly over-expressed in esophageal cancer tissue. Analysis of ROC curve revealed that these proteins had good specificity and accuracy for ESCC diagnosis. Besides, MUC13 and EPCAM are transmembrane proteins with several antigen epitopes that can be identified and combined by antibodies. Thereby, these proteins may be promising antigen targets for immunotherapy, although further experiments are warranted.

As for the mechanism of *ALDH3A1*, aldehyde dehydrogenases *(ALDHs)* mainly have function to oxidize various aldehydes to corresponding acids. They are associated with the detoxification of alcohol-derived acetaldehyde and metabolism process of corticosteroids, biogenic amines, neurotransmitters, and lipid peroxidation. The enzyme encoded by *ALDH3A1* can form a cytoplasmic homodimer that preferentially oxidizes aromatic and medium-chain saturated and unsaturated aldehyde substrates. The gene *ALDH3A1* has impact on promoting resistance to ultraviolet (UV) and 4-hydroxy-2-nonenal-induced oxidative damage in the cornea. It is located within the Smith-Magenis syndrome region on chromosome 17. Analysis from the Genecards website reveals that Sjogren-Larsson Syndrome and conjunctival degeneration are associated with gene *ALDH3A1*. Evidence indicates that *ALDH3A1* has huge impact on human epithelial malignancies. Since the development of esophageal cancer may closely relate with the stimulation from carcinogens, tissue damage, and accumulated toxins caused by smoking and drinking, the relation of *ALDH3A1* with esophageal cancer mainly relies on its involvement in the metabolism and detoxification of toxins[19]. According to the KEGG website, eight pathways including glycolysis/gluconeogenesis, histidine metabolism, tyrosine metabolism, phenylalanine metabolism, metabolism of xenobiotics by cytochrome P450, drug metabolism-cytochrome P450, and metabolic pathways are related with gene *ALDH3A1*. Furthermore, gene *ALDH3A1* is important for β-alanine metabolism and production .

Search in the KEGG website has shown that six pathways including complement and coagulation cascades, alcoholic liver disease, pertussis, staphylococcus aureus infection, coronavirus disease, and systemic lupus erythematosus are related to gene *C2*. Diseases

such as systemic lupus erythematosus, classic complement pathway component defects, and age-related macular degeneration are associated with complement component 2 (C2).

According to the KEGG website, two pathways including synaptic vesicle cycle and GABAergic synapse are related to gene *SLC6A1*. Myoclonic-atonic epilepsy is associated with this gene. *SLC6A1* encodes for gamma-aminobutyric acid (GABA) transporter type 1 (GAT1) which is involved in the reuptake of neurotransmitter GABA. Expression of *SLC6A1* has an important relationship with diseases associated with mental and nervous system dysfunction, such as seizures, autism, mental retardation, and epilepsy. Functional defects of *SLC6A1* may result in reduced sensation and neurasthenia. Major risk factors for ESCC may relate to tissue damage, genetic mutations, and dysfunction in immune system[20].

Gene *ZBTB7C* can encode a Pokemon protein which is also known as human immunodeficiency virus short transcription inducer linking factor 1 (FBI-1). It is an important suppressor gene that can inhibit ARF-specific transcription. *ZBTB7C* can also regulate cellular function and carcinogenesis. Evidence shows that regulation of cell growth by *ZBTB7C* relies mainly on the ARF-p53 pathway[21].

To sum up, analysis of gene-chip from bioinformatics databases may be valuable for identifying over-expressed antigens. In this study, we found genes *MUC13* and *EPCAM* are significantly overexpressed in esophageal cancer tissue, which may be promising antigen targets for immunological therapy. Analytic methods such as enrichment analysis, PPI networks, and survival analysis can be useful approaches for discovering significant genes and molecular pathways related with specific diseases. The result of enrichment analysis indicates that the process of keratinization is important for the pathogenesis of esophageal cancer. Genes *ALDH3A1, C2, SLC6A1* and *ZBTB7C* were screened with significant *P* values of survival curve, which may have great impact on cancer development. Bioinformatics databases can offer information on gene expressions in cancer tissue in terms of mRNA, miRNA, lncRNA, and ssRNA. Visualization of the analytic results with R language software also makes bioinformatics research simpler and more understandable. A wider application of the gene-chip-based bioinformatic studies based on databases and software can be expected in the near future.

Acknowledgments

Many thanks to my colleague Fahim for his assistance in the research and Kewei Liu for his valuable suggestions. Also great thanks to Chunhui Sun and Yunmei Peng for their help and support.

Conflict of interests

The authors declared no conflict of interests.

Authors' contributions

Huang LJ proposed the research direction and revised the manuscript. Mu L guided the use of resources including databases, platforms, websites, and tools. Dai XY contributed to software application and manuscript revision. Song AY was responsible for manuscript writing, revision, statistical analysis, figure creation, and background survey. All authors confirmed and agreed to the final revision of this article.

REFERENCES

1. Liu JX, Zhao L, Zhang XE, et al. Examination of the expression and prognostic significance of DLGAPs in gastric cancer using the TCGA database and bioinformatics analysis. Mol Med Rep 2018; 18 (6): 5621-9. doi: 10.3892/mmr.2018.9574.
2. Luo Q, Cui M, Deng QF, et al. Comprehensive analysis of differentially expressed profiles and reconstruction of a competing endogenous RNA network in papillary renal cell carcinoma. Mol Med Rep 2019; 19 (6): 4685-96. doi: 10.3892/mmr.2019.10138.
3. Enzinger PC, Mayer RJ. Esophageal Cancer. N Engl J Med 2003; 349: 2241-52. doi: 10.1056/NEJMra035010.
4. Li M, Wang K, Pang Y, et al. SPP1 and FN1 associated with progression and prognosis of esophageal cancer identified by integrated expression profiles analysis. Med Sci Monit 2020; 26: e920355. doi: 10.12659/MSM.920355.
5. Li Y, Shi R, Zhu G, et al. Construction of a circular RNA-microRNA-messenger RNA regulatory network of hsa_circ_0043256 in lung cancer by integrated analysis. Thorac Cancer 2022; 13:61-75. doi: 10.1111/1759-7714.14226.
6. Andrea F, Damian S, Sune F, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013; 41: D808-15. doi: 10.1093/nar/gks1094.
7. Yang R. Application of BiNGO and DAVID in Biological Enrichment Analysis of miR-155 Target Genes. J Fujian Med Univ 2012; 46 (6):408-14.
8. Minoru K, Yoko S, Masayuki K, et al. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 2016; 44: D457-62. doi:10.1093/nar/gkv1070.
9. Randi V, Overton J A, Greenbaum J A, et al. The immune epitope database (IEDB) 3.0. Nucleic Acids Res 2015; 43: D405-12. doi: 10.1093/nar/gky1006.
10. Liyan Huang, Sheng Zhang. Research progress in MUC13. J Clin Pathol Res 2014; 34 (3):312-6. doi: 10.11714/j.issn.2095-6959.2014.03.018.
11. Osta, W. A. EpCAM Is overexpressed in breast cancer and is a potential target for breast cancer gene therapy. Cancer Res 2004; 64 (16): 5818-24. doi: 10.1158/0008-5472.CAN-04-0754.
12. Yao QG. Analysis of transmembrane structure of rice aquaporin os-pip2: 6 using TMHMM software. Henan Agr 2017; 000 (029): 59. doi:

CNKI:SUN:NYHN.0.2017-29-037.
13. Wei W, Wang K. Screening and bioinformatics analysis of myocardial infarction markers Based on GEO database chip. J Clin Pathol Res 2019; 39 (1): 27-32.
14. Elias PM. Epidermal lipids, membranes, and keratinization. Int J Dermatol 1982; 20 (1): 1-19.
15. Li CY, Jiang FB, Liang Y. Mutations in the filaggrin gene in Chinese patients with ichthyosis vulgaris. Chin J Mod Med 2010; 20 (7): 113-8. doi: 10.1111/j.1365-2133.2010.09740.x.
16. Raghunath M. SPINK5 and Netherton syndrome: novel mutations, demonstration of missing LEKTI, and differential expression of transglutaminases. J Invest Dermatol 2004; 123 (3): 474-83. doi: 10.1111/j.0022-202X.2004.23220.x.
17. Cuil T, Chen Y, Yang L, et al. DSC3 expression is regulated by p53, and methylation of DSC3 DNA is a prognostic marker in human colorectal cancer. Brit J Cancer 2011; 104: 1013-9. doi: 10.1038/bjc.2011.28.
18. Anaya J. OncoLnc: Linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. Peer J Computer Science 2016; 2(2): e67. doi: 10.7287/PEERJ.PREPRINTS.1780.
19. Patel M, Lu L, Eander DS, et al. ALDH1A1 and ALDH3A1 expression in lung cancers: correlation with histologic type and potential precursors. Lung Cancer 2008; 59: 340-9. doi: 10.1016/j.lungcan.2007.08.033.
20. Thoeringer CK, Ripke S, Unschuld PG. The GABA transporter 1 (SLC6A1): a novel candidate gene for anxiety disorders. J Neural Transm 2009; 116 (6): 649-57. doi: 10.1007/s00702-008-0075-y
21. Jeon BN, Kim MK, Choi WI, et al. KR-POK Interacts with p53 and Represses Its Ability to Activate Transcription of p21WAF1/CDKN1A. Cancer Res 2012; 72 (5): 1137-48. doi: 10.1158/0008-5472.CAN-11-2433.

(Edited by Liang-Jun Gu)

论著

基于 GEO 数据库食管癌关键基因和信号通路分析

宋安忆[1, 2*]，木兰[1]，代小勇[1]，王丽君[1]，黄来强[1*]

1 深圳市基因与抗体治疗重点实验室，清华大学深圳国际研究生院，深圳 518055，广东，中国
2 清华大学化学系，北京 100084，中国

摘要

目的 基于 GEO 数据库，通过分析癌症组织的高表达基因，筛选用于免疫治疗的抗原靶点，并通过富集分析、PPI 网络和生存分析等方法，探讨癌症相关的关键通路和分子机制。

方法 通过筛选高表达基因，借助于 TMHMM 和 IEDB 平台，分析蛋白质的跨膜域和抗原表位。基于富集分析、PPI 网络和生存分析的方法，对癌症发展相关的基因和信号通路进行分析。分析和绘图涉及的软件和平台包括 Prism 8、R 语言、Cytoscape、DAVID、STRING 和 GEPIA 网站等。

结果 *MUC13* 和 *EPCAM* 基因在食管癌组织中高表达，并具有多个抗原识别位点。根据富集分析的结果，一系列基因与角质化过程相关。生存分析结果表明，基因 *ALDH3A1*、*C2*、*SLC6A1* 和 *ZBTB7C* 的生存曲线具有显著差异性。

结论 *MUC13* 和 *EPCAM* 可能作为食管癌免疫治疗的抗原靶点和生物标志物。角质化过程可能在食管癌的发生发展过程中发挥重要作用。基因 ALDH3A1, C2, SLC6A1 和 ZBTB7C 可能与食管癌患者生存周期密切相关。

关键词：GEO；食管癌；抗原；富集分析；生存曲线；信号通路

[*] **通讯作者：**宋安忆，sonanyi0422@163.com；黄来强，huanglq@tsinghua.edu.cn