Signatures and Prognostic Values of Related Immune Targets in Tongue Cancer

Xi Yu

the Second Hospital of Tianjin Medical University

Xiaofei Lv (✉ lx876@hotmail.com)

the Second Hospital of Tianjin Medical University

Research Article

Keywords: tongue cancer, DEGs, CCL20, SCG5, SPP1

Posted Date: November 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-997544/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Tongue cancer, as one of the most malignant oral cancers, is highly invasive and has a high risk of recurrence. At present, tongue cancer in the advanced stage is not obvious, easy to miss the opportunity of early diagnosis. It is important to find markers that can predict the occurrence and progression of tongue cancer. Bioinformatics analysis plays an important role in the acquisition of marker genes. GEO and TCGA data are very important public databases. In addition to expression data, TCGA database also contains corresponding clinical data. In this study, we screened three GEO datasets included GSE13601, GSE34105 and GSE34106 that met the standard. These data sets were combined using the SVA package to prepare the data for differential expression analysis, and then the LIMMA package was used to set the standard to $p < 0.05$ and $|\log_2(FC)| \geq 1.5$. We got 170 DEGs (104, raised 66 downgrade). Besides, the DEseq package was used for differential expression analysis using the same criteria for samples in TCGA database. It ended up with 1589 DEGs (644 up-regulated, 945 down-regulated). By merging these two sets of DEGs, 5 common up-regulated DEGs (CCL20, SCG5, SPP1, KRT75 and FOLR3) and 15 common down-regulated DEGs were obtained. Further functional analysis of the DEGs showed that CCL20, SCG5 and SPP1 is closely related to prognosis and may be a therapeutic target of TSCC.

Introduction

Head and neck tumors are the fifth most common malignancy in the world, and smoking and alcohol consumption are two of the most common risk factors associated with these lesions [1]. Oral cancer accounts for 2.1% of all new cancer cases worldwide, and squamous cell carcinoma is a known malignancy, accounting for more than 90% of all oral cancers [2, 3]. Tongue squamous cell carcinoma (TSCC) is one of the most common malignancies diagnosed in the oral cavity and is associated with poor prognosis due to its high regional recurrence rate and lymphatic metastasis [4]. It is reported that there are about 500,000 new cases of tongue cancer every year, and the global incidence of tongue cancer is increasing year by year, and the onset of tongue cancer is getting younger and younger [5]. Numerous lymphatic vessels of the tongue are abundant in blood circulation, and the tongue is active frequently. These factors promote the early spread of cancer cells to adjacent tissues and organs, such as lymph nodes, the base of the mouth, throat and neck. And although tongue cancer treatments are improving as technology advances, the five-year survival rate is still poor [3, 6]. Oral squamous cell carcinoma (OTSCC) has a lower survival rate than squamous cell carcinoma elsewhere in the mouth. There is currently a lack of OTSCC markers, so it is of great interest to find reliable predictive markers [7].

Materials And Methods

Data preparation

The expression matrix was download from Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>). In order to improve the accuracy of the study, we screened three eligible datasets. These datasets included tongue cancer and normal tissues. GSE13601, GSE34105 and

GSE34106 contains 57, 78 and 43 samples respectively. The specific information is shown in Supplementary Table 1. The SVA software package in R language is used to combine the samples of these three data sets to form a combined data set with 178 samples. Then we downloaded the expression data of head and neck tumors from the TCGA database, and screened the tongue squamous cell carcinoma samples, a total of 147 samples, including 140 tumor samples and 7 normal samples.

Screening and analysis of differentially expressed genes (DEGs)

The limma package and DESeq package in R language was used for differential expression analysis, and the standard of cut-off was set to False Discovery Rate (FDR) <0.05 and $|\log_2(\text{FC})| \geq 1.5$ for both GEO and for TCGA respectively [8, 9]. The volcano package and heatmap package in R language were used to visualize DEGs. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, Gene Ontology (GO) analysis and Gene Set Enrichment Analysis (GSEA) were performed for DEGs. Metascape website (<https://metascape.org/>) was used for pathway enrichment analysis of DEGs in TCGA. T test was used to compare the expression in the two groups, and ANOVA was used to compare that in multiple groups. We further verified the up-regulated DEGs expression using the GEPIA (<http://gepia.cancer-pku.cn/>).

Protein interaction network

The interaction between proteins corresponding to each gene was analyzed using the website String (<https://string-db.org/>), and then the results were visualized using CytoScope. Genemania website (<http://genemania.org/>) was used for functional enrichment analysis of DEGs.

Identify the Hub gene

The DEGs obtained from combined dataset were crossed with those obtained from TCGA, including 5 up-regulated genes and 15 down-regulated genes.

Survival analysis

Kaplan-Meier Plotter website (<https://kmplot.com/>) (citation) was used for survival analysis of the screened HUB genes.

Results

Differential expression analysis of GEO merged dataset

We downloaded three datasets GSE13601, GSE34105 and GSE34106 containing tongue squamous cell carcinoma (TSCC) in GEO database. Each dataset contains tumor samples and normal tissue samples. We used SVA package to merge these datasets and obtain a combined dataset. To identify differences between TSCC samples and normal tongue samples, we performed principal component analysis (PCA) and found that the two groups could be separated in the combined dataset (Figure 1A). We used limma package ($|\log_2(\text{FC})| \geq 1.5$, $p < 0.05$) in R to analyze the 178 samples and screened for DEGs. We identified 170 DEGs, including 104 up-regulated genes and 66 down-regulated gene which were shown in the

volcano diagram, where the red on the right represents the up-regulated differentially expressed genes and the blue one represents the down-regulated differentially expressed genes (Figure 1B). In the heatmap, groups were made according to datasets and sample characteristics to explore the expression of different genes. The top annotation represents the different datasets. DEGs are marked on the right side of the figure (Figure 1C). In order to further explore the functions of the screened DEGs, we use R language for further functional enrichment analysis of these genes in KEGG database, and enriched some up-regulated pathways such as Cytokine-Cytokine receptor interaction, Viral protein interaction with cytokine and cytokine receptor and primary immunodeficiency (Figure 1D). Then GO analysis was performed on these genes, and the pathways enriched in up-regulated genes were mainly cartilage development, connective tissue development, skeletal system development and chondrocyte differentiation in the biological process (BP) module (Figure 1E). In cellular component (CC) module, the enrichment pathways were mainly collagen-containing extracellular matrix, basement membrane, apical part of cell and laminin complex. In Molecular Function (MF) module, mainly enriched in receptor ligand activity, signaling receptor activator activity and metalloendopeptidase activity. GSEA analysis was also performed on the first six up-regulated pathways, and the results were consistent with the previous analysis (Figure 2). The KEGG correspondence is in Supplementary Table2.

Differential expression analysis of TSCC in TCGA database

DEGs of TSCC vs normal samples in TCGA database was performed using DESeq package in R language ($|\log_2(FC)| \geq 1.5$, $p < 0.05$). 644 DEGs up-regulated and 945 down-regulated DEGs were found as shown in the volcano diagram (Figure 3A). We used Metascape website to analyze these up-regulated DEGs and found that the mainly enriched GO pathway was aminoglycoside antibiotic metabolic process, proximal/distal pattern formation and negative regulation of blood coagulation (Figure 3B). KEGG enrichment analysis of up-regulated DEGs showed that the up-regulated pathways were mainly Folate biosynthesis, Steroid hormone biosynthesis, Complement and coagulation cascades, Staphylococcus aureus infection and Cytokine-cytokine receptor interaction. The main down-regulated pathways were Salivary secretion, Calcium signaling pathway and Pancreatic secretion (Figure 3C).

Acquisition and analysis of hub genes

DEGs acquired in GEO were crossed with those in TCGA, and screened 5 common up-regulated genes (CCL20, SCG5, SPP1, FOLR3 and KRT75) and 15 common down-regulated genes (MAOB, HLFSELENBP1, CRISP3, TMPRSS2, TLX1, TGFB3, APOL1, PPP1R3C, SLITRK5, MFAP4, STATH, GSTM5, RNASE4, COX7A1, FHL1, SLC25A4, EFHD1, MYH11, MGP, SCGB2A1, LPIN1, ELF5, ADH1B, CKMT2) (Figure 4A). The protein-protein interaction (PPI) network analysis of these five up-regulated genes found that proteins expressed by other genes were all functionally related except FOLR3 (Figure 4B). Then, we conducted interaction network analysis of these genes individually (Figure 4C).

Further expression verification of these hub genes

We further verified the expression of up-regulated DEGs using GSE13601 dataset. This dataset includes tumor samples and their paired normal samples. After remove the individual samples in the database, we

obtained 20 pairs of paired samples and then we compared the expressions of CCL20, SCG5, SPP1, FOLR3 and KRT75 genes in tumor tissues vs normal tissues. It was shown that the expression of these 5 genes was higher in TSCC than in normal tissue both in the whole sample and the most paired samples (Figure 5A). The same results can be obtained using TCGA dataset (Figure 5B).

To explore the relationship between the expression of these Hub genes and clinical stage, we used TCGA database to explore and found that the expression level of CCL20 was significantly increased in patients with stage III and IV compared with stage I ($P < 0.05$), the expression of SPP1 and SCG5 in stage IV was significantly higher than that in stage I patients ($P < 0.05$) (Figure 5C).

Multivariate COX regression of TSCC in TCGA database for each clinical feature and prognosis

To further explore the clinical data of TCGA, Multivariate Cox regression was performed using clinical data from TCGA database (Figure 6). We found that $\text{ajcc_pathologic_T} > 2$ vs $\text{ajcc_pathologic_T} \leq 2$ the risk increased. Similarly, years of smoking also has an impact on the survival of patients, when the smoking time is more than 29 years, the risk of patients is significantly increased. Besides, we found that when TSCC stage is higher than grade III, age > 59 , N stage > 1 and the number of cigarettes smoked per day > 2.5 increased the risk, but P value of multivariate Cox did not meet the statistical standards. Kaplan-meier analysis was performed on clinical data using SPSS software, and we found that patients with stage higher than III had a significantly worse prognosis than patients with stage lower than III. AJCC_pathologic_T had a poor prognosis when it was higher than T2. Similarly, AJCC_pathologic_N stage higher than N1 and smoking years higher than 29 years remained risk factors (Figure 7).

Hub genes analysis in HNCS

These up-regulated hub genes were further verified by GEPIA, and the results showed that the expression of these 5 up-regulated genes in Head and Neck squamous cell carcinoma (HCNC) was also higher than that in normal tissues (Figure 8). Kaplan – Meier Plotter was used to analyze the relationship between the expression of these genes and prognosis, and the results showed that high expression of these genes was strongly associated with poor prognosis in HCNC ($P < 0.05$) (Figure 9).

Discussion

Key molecular alterations need to be identified to identify effective therapeutic targets [10–12]. In order to explore the expression genes related to the occurrence and progression of tongue cancer, we used GEO database and TCGA database for exploration. To increase the sample size, we used the SVA package in R language for GSE13601, GSE34105 and GSE34106. Three datasets were combined to obtain a dataset containing 178 samples (121 TSCC and 57 normal samples). Use Limma package for DEGs analysis and set $|\text{Log}_2\text{FC}| > 1.5$, $p < 0.05$. Finally, we get 104 up-regulated genes, 66 down-regulation genes. Then, DEGs of TCGA were obtained by using DEseq package in R language, and the cut-off value was set as same before. DEGs obtained from GEO database and TCGA dataset were crossed, obtaining 5 up-regulated DEGs and 15 down-regulated DEGs. The 5 up-regulated DEGs are CCL20, SCG5, SPP1, FOLR3 and KET75.

Chemokine (C-C motif) Ligand 20 (CCL20) is a substance involved in tissue validation and homeostasis with a specific receptor C-C Chemokine receptor 6 (CCR6). It is expressed in various tissues and immune cells in human body, and the CCL20-CCR6 axis is closely associated with inflammation and infectious diseases. It has been reported that the CCL20-CCR6 axis is closely related to a variety of cancers which can directly promote the progress of cancer by enhancing the migration and proliferation of cancer cells, and can also regulate the tumor microenvironment by immune cells.

CCL20 has been reported to indirectly promote tumor progression by recruiting Treg, Th17, and Th22 cells to maintain the development and immunosuppressive microenvironment [13]. Studies have shown that TSCC cells produce CCL20 after interacting with macrophages, and the cells can express CCR6 in the TSCC microenvironment. The CCL20-CCR6 axis may be associated with OSCC progression by inducing CD163 expression in macrophages [14]. SPP1 is secreted glycoprophosphoprotein, which is involved in a variety of functions, including cell adhesion, migration and invasion [15–18]. FOLR3 is highly overexpressed on several tumor cells, including ovarian, non-small cell lung cancer, kidney, brain, endometrial, colorectal, breast, pancreatic, gastric, prostate, testicular, and bladder, thus it was known as therapeutic target for cancer treatment [19]. It has been reported that these genes are associated with the progression of a variety of tumors. In addition, according to the results of survival analysis, we found that CCL20, SPP1 and SCG5 are closely related to the prognosis of TSCC and HCSN. We have reason to believe that these genes can be developed into prognostic indicator and are expected to become a therapeutic target of TSCC.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the [NAME] repository(<https://github.com/Wangyouzhi88/-/commit/b99e2fb3f4340a78ae80ff3704dd56cc00f569a2>).

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

Conceptualization, Xi Yu, Xiaofei Lv; material preparation, Xi Yu; writing-original draft preparation, Xi Yu, Xiaofei Lv; data analysis and interpretation, Xi Yu, Xiaofei Lv; writing editing and review, All authors; All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Diaz EM KM, Sturgis EM et al. In Kufe DW, Bast RC, Hait WN et al.. **Neoplasms of the head and neck.** *Holland-Frei Cancer Medicine*: **1131–78**.
2. Furukawa A, Shinoda M, Kubo A, Honda K, Akasaka R, Yonehara Y *et al*. **Endothelin Signaling Contributes to Modulation of Nociception in Early-stage Tongue Cancer in Rats.** *Anesthesiology* 2018; **128**: **1207-19**.
3. Huang L, Luo EL, Xie J, Gan RH, Ding LC, Su BH *et al*. **FZD2 regulates cell proliferation and invasion in tongue squamous cell carcinoma.** *Int J Biol Sci* 2019; **15**: **2330-9**.
4. Bagan J, Sarrion G, Jimenez Y. **Oral cancer: clinical features.** *Oral Oncol* 2010; **46**: **414-7**.
5. Ng JH IN, Tan MH, Edgren G. **Changing epidemiology of oral squamous cell carcinoma of the tongue: A global study.** *Head Neck* 2017 Feb; **39**(2): **297-304**.
6. Zhang Z PJ, Li L, Wang Z, Xiao W, Li N. **Survey of risk factors contributed to lymphatic metastasis in patients with oral tongue cancer by immunohistochemistry..** *J Oral Pathol Med* 2011 Feb; **40**(2): **127-34**.
7. Heikkinen I, Bello IO, Wahab A, Hagstrom J, Haglund C, Coletta RD *et al*. **Assessment of Tumor-infiltrating Lymphocytes Predicts the Behavior of Early-stage Oral Tongue Cancer.** *Am J Surg Pathol* 2019; **43**: **1392-6**.
8. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W *et al*. **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015; **43**: **e47**.
9. Anders S, Huber W. **Differential expression analysis for sequence count data.** *Genome Biol* 2010; **11**: **R106**.
10. Hussein AA, Forouzanfar T, Bloemena E, de Visscher J, Brakenhoff RH, Leemans CR *et al*. **A review of the most promising biomarkers for early diagnosis and prognosis prediction of tongue squamous cell carcinoma.** *Br J Cancer* 2018; **119**: **724-36**.
11. Mithani SK, Mydlarz WK, Grumbine FL, Smith IM, Califano JA. **Molecular genetics of premalignant oral lesions.** *Oral Dis* 2007; **13**: **126-33**.
12. Vigneswaran N, Williams MD. **Epidemiologic trends in head and neck cancer and aids in diagnosis.** *Oral Maxillofac Surg Clin North Am* 2014; **26**: **123-41**.
13. Walch-Ruckheim B, Mavrova R, Henning M, Vicinus B, Kim YJ, Bohle RM *et al*. **Stromal Fibroblasts Induce CCL20 through IL6/C/EBPbeta to Support the Recruitment of Th17 Cells during**

Cervical Cancer Progression. <bi>Cancer Res</bi> 2015; **75**: 5248-59.

14. Shigeoka M, Koma YI, Kodama T, Nishio M, Akashi M, Yokozaki H. **Tongue Cancer Cell-Derived CCL20 Induced by Interaction With Macrophages Promotes CD163 Expression on Macrophages.** <bi>Front Oncol</bi> 2021; **11**: 667174.
15. Packer L, Pavey S, Parker A, Stark M, Johansson P, Clarke B <bi>et al</bi>. **Osteopontin is a downstream effector of the PI3-kinase pathway in melanomas that is inversely correlated with functional PTEN.** *Carcinogenesis* 2006; **27**: 1778-86.
16. Zhou Y, Dai DL, Martinka M, Su M, Zhang Y, Campos EI <bi>et al</bi>. **Osteopontin expression correlates with melanoma invasion.** <bi>J Invest Dermatol</bi> 2005; **124**: 1044-52.
17. Deng G, Zeng F, Su J, Zhao S, Hu R, Zhu W <bi>et al</bi>. **BET inhibitor suppresses melanoma progression via the noncanonical NF-kappaB/SPP1 pathway.** <bi>Theranostics</bi> 2020; **10**: 11428-43.
18. Elias EG, Hasskamp JH, Sharma BK. **Cytokines and growth factors expressed by human cutaneous melanoma.** <bi>Cancers (Basel)</bi> 2010; **2**: 794-808.
19. Bin-Alee F, Arayataweegool A, Buranapraditkun S, Mahattanasakul P, Tangjaturonrasme N, Hirankarn N <bi>et al</bi>. **Transcriptomic analysis of peripheral blood mononuclear cells in head and neck squamous cell carcinoma patients.** <bi>Oral Dis</bi> 2021; **27**: 1394-402.

Figures

Figure 1

Differential expression analysis of GEO merged dataset (GSE13601, GSE34105 and GSE34106). (A) PCA analysis was performed to identify differences between tumor tissue and normal tissue in the combined dataset. (B) The volcano diagram represents the DEGs obtained by differential expression analysis using limma packages, where red represents up-regulated DEGs and blue represents down-regulated DEGs. (C) The heatmap shows the differentially expressed genes in three datasets, with the annotations above representing each dataset and the annotations in the second row representing different types of samples. The differentially expressed genes are shown on the right, where red represents up-regulated genes and blue represents down-regulated genes. (D) KEGG pathway enrichment analysis of DEGs, in which red represents up-regulation gene enrichment pathway and gray represents down-regulation gene enrichment pathway. (E) GO enrichment analysis of DEGs.

Figure 2

The DEGs enrichment pathway obtained from the combined database was further verified by GSEA.

Figure 3

Differential expression analysis of TSCC in TCGA database. (A) The volcano map represents DEGs of TSCC in TCGA database, with red representing up-regulated genes and blue representing down-regulated DEGs. (B) Results of pathway enrichment analysis of up-regulated differentially expressed genes using Metascape. (C) Pathway enrichment analysis of DEGs—the left represents the enrichment of up-regulated gene KEGG pathway, and the right represents the enrichment of down-regulated gene KEGG pathway.

Figure 5

Further expression verification of these 5 common up-regulated genes (CCL20, SCG5, SPP1, FOLR3 and KRT75) (A) Box plots show the expression levels of these 5 genes in tumor tissue compared with matched normal tissue in the GSE13601 database, with lines connected to a pair of paired samples. (B) The violin diagram shows the comparison of the expression levels of these 5 genes in TCGA tongue cancer tissue and normal tissue. (C) The box diagram shows the expression of these 5 genes at various stages in the TCGA database.

Figure 6

Multivariate COX regression of TSCC in TCGA database for each clinical feature and prognosis.

Figure 7

SPSS software was used to analyze the individual clinical features and prognosis.

Figure 8

Use GEPIA software to analysis the expression of CCL20, SCG5, SPP1, FOLR3 and KRT75 these five genes in each cancer species.

Figure 9

Using Kaplan – Meier Plotter website to identify the relationship between the expression of CCL20, SCG5, SPP1, FOLR3 and KRT75 genes and prognosis in HNCS.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryTable.docx](#)