

# Discovery Potential Therapeutic Drugs for Oral Tongue Squamous Cell Carcinoma Based on Text Mining and Data Analysis

Liuqing Yang

Jilin University

Jue Wang

Jilin University

Jinling Yu

jilin university

Handan Huang

Jilin University

Ying Gao

Jilin University

Lu Tao

jilin university

Jingjing Yao

jilin university

Zhihui Liu ([✉ liu\\_zh@jlu.edu.cn](mailto:liu_zh@jlu.edu.cn))

jilin university

---

## Research Article

**Keywords:** text mining, drug discovery, oral tongue squamous cell carcinoma, GEO, differentially expressed genes (DEGs).

**Posted Date:** October 25th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-970755/v1>

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

[Read Full License](#)

---

## Abstract

**Background:** Oral tongue squamous cell carcinoma (OTSCC) is the most common malignant tumor of the oral cavity. The aim of this study was to use text mining and data analysis to discover some existing drugs that target genes and to explore potential therapeutic drugs for OTSCC.

**Methods:** We used the text mining tool pubmed2ensembl to extract genes associated with OTSCC, and two datasets (GSE30784, GSE23558) from Gene Expression Omnibus (GEO) were used for the data analysis. Then, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed for the intersection of the three gene sets. Protein-protein interaction (PPI) network was constructed by STRING, gene module analysis was performed using the Molecular Complex Detection (MCODE), a plug-in in Cytoscape. Lastly, a database of drug-gene interactions was used to identify significant genes to explore potential drugs for the treatment of OTSCC.

**Results:** We produced 403 unique genes associated with oral tongue squamous cell carcinoma through text mining. GSE23558 and GSE 30784 obtained 1637 and 1159 differentially expressed genes (DEGs) through data analysis, respectively. A total of 28 genes were obtained from the intersection gene sets, including 20 up-regulated genes and 8 down-regulated genes. We screened the most significant modules by using MCODE, among which 8 genes were associated with oral tongue squamous cell carcinoma as core genes. Eventually, nine drugs were found to target eight genes.

**Conclusions:** In this study, PLAU, SERPINE1, MMP1, MMP3, MMP10, CXCL10, CXCL12, and SPP1 were potentially key genes involved in the treatment of OTSCC. Furthermore, 12 drugs were identified as potential therapeutic agents for oral tongue squamous cell carcinoma treatment and management

## Introduction

Oral squamous cell carcinoma (OSCC) is a malignant tumor and poses a serious threat to human health. Oral tongue squamous cell carcinoma (OTSCC) is a malignant tumor that occurs in the anterior 2/3 of the tongue in the oral cavity and is the sixth most common cancer in the world (1, 2). It is an aggressive head and neck malignancy characterized by local invasion and distal metastasis, with a high recurrence rate and no significant improvement in 5-year survival (3–5). The incidence of oral cancer varies globally (6, 7). Typical risk factors are smoking and excessive alcohol consumption (8). Another risk factor is HPV (9). Cervical lymph node metastasis is an important prognostic factor (10). Early symptoms of oral tongue squamous cell carcinoma are mild and patients are usually diagnosed at a later stage (11). Therefore, the prognostic effect is relatively poor. OTSCC is a complex disease. In the clinic, surgery and chemoradiotherapy are considered to be effective methods for the treatment of OTSCC, but research on drug therapy remains limited (12). Obviously, there is an urgent need for a new approach to the treatment and prevention of OTSCC.

In this study, we aimed to find some existing drugs to provide new ideas for the prevention and treatment of oral tongue squamous cell carcinoma. The development of new drugs requires physical

experimentation and drug compound development, which take a long time and costs. However, drug repositioning may be a more cost-saving and shortened drug development time. Further research is needed to provide a theoretical basis for the development of new treatment modalities that maximize functional preservation and minimize recurrence and metastasis rates.

Nowadays, data analysis and text mining have been applied in many aspects such as disease diagnosis, identifying potential key genes and predicting the occurrence of diseases. Compared with bioinformatics in the cancer field, there are fewer researches focused on oral tongue squamous cell carcinomas through text mining and data analysis. Following the repurposing paradigm, in this study, we used bioinformatics strategies to search for existing drugs to provide new ideas for the treatment of OTSCC. First, we obtained a unique list of genes through text mining and data analysis and intersected these genes to obtain common genes. Then, we performed protein-protein interaction analysis (PPI) on these genes and identified the most significant gene modules. Finally, drug candidates were obtained from drug-gene interaction analysis of module genes. Figure 1 illustrates the workflow of this study.

**Figure 1** illustrates the workflow of this study.

## Method And Materials

### Text mining and data analysis

We used pubmed2ensemble (<http://pubmed2ensemble.ls.manchester.ac.uk>), an open website, to perform text mining (13). We first entered the key word “tongue squamous cell carcinoma” into pubmed2ensemble, and then combined all the duplicates to obtain unique gene set.

GSE23558 and GSE30784 were downloaded and obtained from the National Center Biotechnology Gene Expression Omnibus (NCBI-GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) (14). The platforms used were GPL6480 and GPL570. GSE23558 contains 27 patients and 5 healthy controls. GSE30784 includes 229 samples. After the gene expression profile was obtained, R software was used to analyze the original datasets. The gene probes were converted into the corresponding gene symbols. Moreover, for one probe corresponding to multiple genes and non-mRNA probes were removed. GSE30784 and GSE23558 were analyzed using RStudio software. Limma package of RStudio software was used for identify the DEGs between patients and healthy control samples (15). The DEGs were selected for subsequent analysis by setting the adjusted *p*-value<0.01 and  $|\log_2 \text{fold change(FC)}| \geq 1.5$  as cutoff criteria. Veen package (<http://www.molbiotools.com/listcompare.html>) was used to draw the intersection of the three gene sets.

Finally, we intersected the three gene sets, and used the intersection of the three gene sets for further analysis.

### Functional and Signal Pathway enrichment Analysis

DAVID Bioinformatics Resources 6.8 (<https://david.ncifcrf.gov/>), an online website, provides researchers with a comprehensive set of functional annotation tools to understand the biological significance behind

a large number of genes (16). The gene ontology (GO) analysis annotates genes and gene products from the molecular to the organism level in three categories: biological process (BP), cellular component (CC), and molecular function (MF) (17). The Kyoto Encyclopedia of Genes and Genomes (KEGG) is an open access informatic and a systematic analysis database for linking high-level function information and the biological systems (18), based on the gene-chip and high-throughput experimental technologies.  $P<0.05$  was considered as statistically significant as the cutoff criterion.

## **Protein-Protein interaction Analysis and Gene Module Analysis**

The database STRING (<http://string-db.org>) is a global open resource for predicting and analyzing protein-protein associations, including text mining, experiments, databases, co-expression, neighborhood, gene fusion, co-occurrence (19). We submitted 28 genes to STRING to build PPI networks. The medium confidence score (0.400) was set as the minimum required interaction score and we set up to hide disconnected nodes in the network, the other parameters in STRING were set by default. The Cytoscape plugin MCODE and STRING app were then used to screen the most significant gene modules (20, 21). Besides, we set the parameter k-core=5 in MCODE as the standard and the other parameters as default.

## **Drug-gene interaction and functional analysis of potential genes**

The drug-gene interaction database (DGIdb: <http://www.dgidb.org>) is an open website that consolidates and organizes information on drug-gene interactions and gene druggability from papers, databases and online resources (22). The most significant gene modules were entered into the gene-drug database as potential targets to explore existing drug or compounds, and these candidate genes were then performed enrichment analysis.

## **Results**

### **Text mining and microarray data analysis**

GSE23558 and GSE30784 were analyzed singly to identify differential genes, which expressed in oral squamous cell carcinomas. There were 1637 DEGs, including 1094 down-regulated genes and 543 up-regulated genes in GSE23558, which were differentially expressed between patients and healthy control. There were 1159 DEGs(585 down-regulated and 574 up-regulated) in GSE30784, which were differentially expressed between cancer and healthy. We extracted 403 unique genes related to tongue squamous cell carcinoma from text mining. A total of 28 common genes were obtained from the intersection of the three gene sets (Figure 1). Among these common genes, 20 genes were upregulated genes and 8 genes were downregulated genes (Table1).

### **Functional and Signal Pathway enrichment Analysis**

We input a list of 28 genes into DAVID for GO and KEGG enrichment analysis(Figure 2). Furthermore, as shown in Figure 2, it showed that the top six significant most enrichment terms for BP, CC, MF, and KEGG signal pathways of common genes. In BP category, it was mainly enriched in the regulation of cellular component movement, extracellular matrix organization and extracellular structure. In CC category, it was enriched in the extracellular region part, extracellular region and cell surface. In MF category, it was enriched in the receptor binding, cell adhesion molecule binding and serine-type endopeptidase activity. In KEGG category, it was ECM-receptor interaction, focal adhesion and leukocyte transendothelial migration.

## Protein-Protein interaction Analysis and Gene Module Analysis

We pasted 28 genes into STRING to make PPI networks and downloaded files in TSV format, and then input the files into Cytoscape for analysis. There were a total of 23 nodes with 49 edges in the PPI networks, but 5 genes were not in the PPI networks (Figure 3A). MOCDE and STRING in Cytoscape were used to screen the most significant gene modules. Module 1 was composed of 8 nodes with 24 edges (Figure 3B). We finally have eight hub genes, namely *PLAU*, *SERPINE1*, *MMP1*, *CXCL10*, *SPP1*, *CXCL12*, *MMP3*, and *MMP10*. These genes were subjected to drug-gene interaction analysis.

## Drug-gene interaction and functional analysis of potential genes

We selected 8 genes from the most meaning gene modules as potential gene targets for drug-gene interaction analysis, and finally 12 drugs were found to target 6 of the 8 genes (the exceptions were *MMP10* and *CXCL12*), and then these 12 drugs were classified (Figure 4, **Table 2**). In addition, as shown in **Table 3**, the enrichment analysis of 8 genes using David website showed that the GO terms were mainly enriched in extracellular matrix disassembly (BP), extracellular space (CC), serine-type endopeptidase activity (MF).

## Discussion

OTSCC is a highly malignant tumor, which causes harm to human health. Recently, the incidence of OTSCC is on the rise. The tongue has abundant blood circulation and is prone to lymph node metastasis, which often leads to a poor prognosis. Currently, OTSCC is usually treated with a combination of surgery, radiation therapy, and chemotherapy. However, the surgical treatment of OTSCC will inevitably cause oral dysfunction. Therefore, in order to protect the oral function of patients as much as possible, drug therapy is adopted to improve the quality of OTSCC patients.

In our study, we identified 28 genes related to OTSCC through text mining and data analysis. Therefore, our purposes were to discover new drug indications. Finally, as we expected, the PPI network and enrichment analysis identified 8 core genes, namely *PLAU*, *SERPINE1*, *MMP1*, *MMP3*, *MMP10*, *CXCL10*, *CXCL12*, and *SPP1*, target to 9 existing potential drugs which might be used for the treatment of OSTCC.

These 8 core genes are involved in the NF-kappa B signaling pathway, Toll-like receptor signaling pathway, TNF signaling pathway, Complement and coagulation cascades and Rheumatoid arthritis.

Plasminogen Activator Urokinase (*PLAU*) converts the pro-enzyme into serine protease plasmin. Urokinase-type plasminogen activator (uPA) is involved in the proteolysis of basement membrane and extracellular matrix structure, and activates fibroblast growth factor, vascular endothelial growth factor and transforming growth factor- $\beta$ , which plays an important role in tumorigenesis process. In the colorectal, stomach, and oral cavity, increased expression of uPA was observed. Marianna et al. reported that uPA is highly expressed in high-grade tumors and in the worst invasive mode, and is involved in the progression of OSCC to poor prognosis (23). Serpin family E member 1 (*SERPINE1*), a serine proteinase inhibitor (serpin), is the main inhibitor of tissue plasminogen activator (tPA) and urokinase (uPA). Defects in this gene lead to plasminogen activator inhibitor-1 deficiency (PAI-1 deficiency). *SERPINE1* is upregulated in OTSCC and plays an oncogenic role (24). Dhanda J et al. have found that the combination of Serpin family E member 1 (*SERPINE1*) and alpha smooth muscle actin (SMA) immunohistochemistry offer potential as prognostic biomarkers in OSCC (25). Matrix metalloproteinases (MMPs), a family of zinc-dependent endopeptidases, can degrade various proteins in the extracellular matrix (ECM). MMPs include *MMP1*, *MMP3* and *MMP10*. Therefore, MMPs may play an important role in development of cancer. Matrix metallopeptidase 1 (*MMP1*) is the interstitial collagen, which is activated in aggressive oral squamous cell carcinoma. Jordan et al. have shown that *MMP1* was overexpressed and the mRNA level of *MMP1* was significantly elevated in OSCC (26). Overexpression of *MMP3* can lead to ECM degradation and overexpression of *MMP3* appears to kill epithelial cells (27). Chemokines and their receptors are involved in tumor development and metastasis. In addition, chemokines also promote the anti-tumor activity by guiding and mobilizing the mobilization and colonization of cells. The ligand for CXCL10 is CXCR3, which inhibits tumor growth. In different human cancers, high expression of CXCL10 at the tumor site is associated with a favorable prognosis (28–30). C-X-C motif chemokine ligand 12 (CXCL12), also known as SDF-1 (stromal cell-derived factor-1), belongs to the chemokine family. CXCL12 includes the chemokine receptors CXCR4 and CXCR7. The combination of CXCL12 and CXCR4 is responsible for regulating a variety of biological and pathological processes, including hematopoiesis and apoptosis, immune and mitotic activity, cancer cell growth, migration, and neovascularization (31). Therefore, the CXCL12/CXCR4 axis is considered as a new drug target for the treatment of oral cancer. Secreted phosphoprotein 1 (SPP1), also known as osteopontin, a secreted and chemokine-like glycol-phosphoprotein is involved in immune responses, cancer progression and cell signaling (32). Feng et al. confirmed that SPP1 plays an important role in oral cancer invasion (33). Additionally, Liu et al. found that SPP1 was significantly associated with the survival rate and the survival rate of patients with high SPP1 expression was significantly reduced. They also demonstrated that the regulation of SPP1 expression influenced proliferation, migration, and invasion, and inhibited apoptosis in cell lines (34).

In total, we obtained 12 drugs from the drug-gene interaction database. These drugs are classified as antineoplastic agent, small molecule, thrombolytic agents and protein (**Table 3**). Although these existing drugs are likely to be of further help in the treatment of OTSCC, the lack of experimental validation is a limitation of this study, so further experimental studies are needed to verify the results.

# **Conclusions**

According to the text mining conception (keyword: tongue squamous cell carcinoma) and microarray data analysis (dataset: GSE30784 and GSE23558), we found 20 existing drugs, approved by FDA, target to nine genes, which involved in the inflammatory pathway. These genes might be used for OTSCC, as well as its original drug indications.

# **Declarations**

## **Ethics approval and consent to participate**

Not applicable

## **Consent for publication**

Not applicable

## **Availability of data and materials**

Not applicable

## **Competing interests**

The authors declare that they have no competing interests

## **Funding**

This project was supported and funded by Scientific and Technological Talent Project (20200301008RQ).

## **Authors' contributions**

LY analyzed the data regarding the oral tongue squamous cell carcinomas and was a major contributor in writing the manuscript. JW modified the grammar of the article. JY, HH, GY, LT, and JY collected data. All authors read and approved the final manuscript.

## **Acknowledgements**

Not applicable

# **References**

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7-30.  
doi:10.3322/caac.21442
2. Warnakulasuriya S. Causes of oral cancer—an appraisal of controversies. Br Dent J. 2009;207(10):471-475. doi:10.1038/sj.bdj.2009.1009

3. Ng JH, Iyer NG, Tan MH, Edgren G. Changing epidemiology of oral squamous cell carcinoma of the tongue: A global study. *Head Neck.* 2017;39(2):297-304. doi:10.1002/hed.24589
4. Franceschi D, Gupta R, Spiro RH, Shah JP. Improved survival in the treatment of squamous carcinoma of the oral tongue. *Am J Surg.* 1993;166(4):360-365. doi:10.1016/s0002-9610(05)80333-2
5. Amit M, Yen TC, Liao CT, et al. Improvement in survival of patients with oral cavity squamous cell carcinoma: An international collaborative study. *Cancer.* 2013;119(24):4242-4248. doi:10.1002/cncr.28357
6. Zhang YY, Wang DC, Su JZ, Jia LF, Peng X, Yu GY. Clinicopathological characteristics and outcomes of squamous cell carcinoma of the tongue in different age groups. *Head Neck.* 2017 Nov;39(11):2276-2282. doi: 10.1002/hed.24898. Epub 2017 Aug 26. PMID: 28842932.
7. Patel SC, Carpenter WR, Tyree S, et al. Increasing incidence of oral tongue squamous cell carcinoma in young white women, age 18 to 44 years. *J Clin Oncol.* 2011;29(11):1488-1494. doi:10.1200/JCO.2010.31.7883
8. Jamal A, King BA, Neff LJ, Whitmill J, Babb SD, Graffunder CM. Current Cigarette Smoking Among Adults - United States, 2005-2015. *MMWR Morb Mortal Wkly Rep.* 2016;65(44):1205-1211. Published 2016 Nov 11. doi:10.15585/mmwr.mm6544a2
9. Rooper LM, Windon MJ, Hernandez T, et al. HPV-positive Squamous Cell Carcinoma of the Larynx, Oral Cavity, and Hypopharynx: Clinicopathologic Characterization With Recognition of a Novel Warty Variant. *Am J Surg Pathol.* 2020;44(5):691-702. doi:10.1097/PAS.0000000000001433
10. Yuasa-Nakagawa K, Shibuya H, Yoshimura R, et al. Cervical lymph node metastasis from early-stage squamous cell carcinoma of the oral tongue. *Acta Otolaryngol.* 2013;133(5):544-551. doi:10.3109/00016489.2012.748988
11. Almangush A, Bello IO, Coletta RD, et al. For early-stage oral tongue cancer, depth of invasion and worst pattern of invasion are the strongest pathological predictors for locoregional recurrence and mortality. *Virchows Arch.* 2015;467(1):39-46. doi:10.1007/s00428-015-1758-z
12. Camisasca DR, Silami MA, Honorato J, Dias FL, de Faria PA, Lourenço Sde Q. Oral squamous cell carcinoma: clinicopathological features in patients with and without recurrence. *ORL J Otorhinolaryngol Relat Spec.* 2011;73(3):170-176. doi:10.1159/000328340
13. Yu S, Tranchevent LC, De Moor B, Moreau Y. Gene prioritization and clustering by multi-view text mining. *BMC Bioinformatics.* 2010;11:28. Published 2010 Jan 14. doi:10.1186/1471-2105-11-28
14. Barrett T, Troup DB, Wilhite SE, et al. NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.* 2009;37(Database issue):D885-D890. doi:10.1093/nar/gkn764
15. Larriba Y, Rueda C, Fernández MA, Peddada SD. Microarray Data Normalization and Robust Detection of Rhythmic Features. *Methods Mol Biol.* 2019;1986:207-225. doi:10.1007/978-1-4939-9442-7\_9
16. Huang DW, Sherman BT, Tan Q, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.*

- 2007,8(9):R183. doi:10.1186/gb-2007-8-9-r183
17. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 2010,11(2):R14. doi:10.1186/gb-2010-11-2-r14
18. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000,28(1):27-30. doi:10.1093/nar/28.1.27
19. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* 2003,31(1):258-261. doi:10.1093/nar/gkg034
20. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics.* 2011,27(3):431-432. doi:10.1093/bioinformatics/btq675
21. Bandettini WP, Kellman P, Mancini C, et al. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. *J Cardiovasc Magn Reson.* 2012,14(1):83. Published 2012 Nov 30. doi:10.1186/1532-429X-14-83
22. Cotto KC, Wagner AH, Feng YY, et al. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* 2018,46(D1):D1068-D1073. doi:10.1093/nar/gkx1143
23. Serpa MS, Mafra RP, Queiroz SIML, Silva LPD, Souza LB, Pinto LP. Expression of urokinase-type plasminogen activator and its receptor in squamous cell carcinoma of the oral tongue. *Braz Oral Res.* 2018,32:e93. doi:10.1590/1807-3107bor-2018.vol32.0093
24. Zhao C, Liu Z. MicroRNA 617 Targeting SERPINE1 Inhibited the Progression of Oral Squamous Cell Carcinoma. *Mol Cell Biol.* 2021,41(6):e0056520. doi:10.1128/MCB.00565-20
25. Dhanda J, Triantafyllou A, Liloglou T, et al. SERPINE1 and SMA expression at the invasive front predict extracapsular spread and survival in oral squamous cell carcinoma. *Br J Cancer.* 2014,111(11):2114-2121. doi:10.1038/bjc.2014.500
26. Jordan RC, Macabeo-Ong M, Shibuski CH, et al. Overexpression of matrix metalloproteinase-1 and -9 mRNA is associated with progression of oral dysplasia to cancer. *Clin Cancer Res.* 2004,10(19):6460-6465. doi:10.1158/1078-0432.CCR-04-0656
27. Liu SY, Liu YC, Huang WT, Huang GC, Su HJ, Lin MH. Requirement of MMP-3 in anchorage-independent growth of oral squamous cell carcinomas. *J Oral Pathol Med.* 2007,36(7):430-435. doi:10.1111/j.1600-0714.2007.00524.x
28. Karin N. Chemokines and cancer: new immune checkpoints for cancer therapy. *Curr Opin Immunol.* 2018,51:140-145. doi:10.1016/j.coim.2018.03.004
29. Bronger H, Singer J, Windmüller C, et al. CXCL9 and CXCL10 predict survival and are regulated by cyclooxygenase inhibition in advanced serous ovarian cancer. *Br J Cancer.* 2016,115(5):553-563. doi:10.1038/bjc.2016.172
30. Toiyama Y, Fujikawa H, Kawamura M, et al. Evaluation of CXCL10 as a novel serum marker for predicting liver metastasis and prognosis in colorectal cancer. *Int J Oncol.* 2012,40(2):560-566. doi:10.3892/ijo.2011.1247

31. Zhou W, Guo S, Liu M, Burow ME, Wang G. Targeting CXCL12/CXCR4 Axis in Tumor Immunotherapy. *Curr Med Chem.* 2019;26(17):3026-3041. doi:10.2174/0929867324666170830111531
32. Castello LM, Raineri D, Salmi L, et al. Osteopontin at the crossroads of inflammation and tumor progression. *Mediators Inflamm.* 2017;2017:4049098. <https://doi.org/10.1155/2017/4049098>
33. Zhang X, Zhang L, Tan X, et al. Systematic analysis of genes involved in oral cancer metastasis to lymph nodes. *Cell Mol Biol Lett.* 2018;23:53. Published 2018 Nov 15. doi:10.1186/s11658-018-0120-2
34. Liu K, Hu H, Jiang H, et al. Upregulation of secreted phosphoprotein 1 affects malignant progression, prognosis, and resistance to cetuximab via the KRAS/MEK pathway in head and neck cancer. *Mol Carcinog.* 2020;59(10):1147-1158. doi:10.1002/mc.23245

## Tables

**Table I.** There were a total of 28 intersection genes from oral tongue squamous cell carcinoma, including 20 up-regulated genes and 8 down-regulated genes.

Common genes	Gene Names
Up-regulated genes	<i>MMP1,PLAU,ITGA3,PTHLH,MSN,SPP1,CDC6,PDPN,SERPINE1, MMP3,PXN,ITGB4,TPM1,MMP10,DSG2,IFI44,CXCL10,WARS, EIF5A2,TNC</i>
Down-regulated genes	<i>CXCL12,KRT4,GATM,TGM3,KRT13,CYP3A5,KRT19,CLDN7</i>

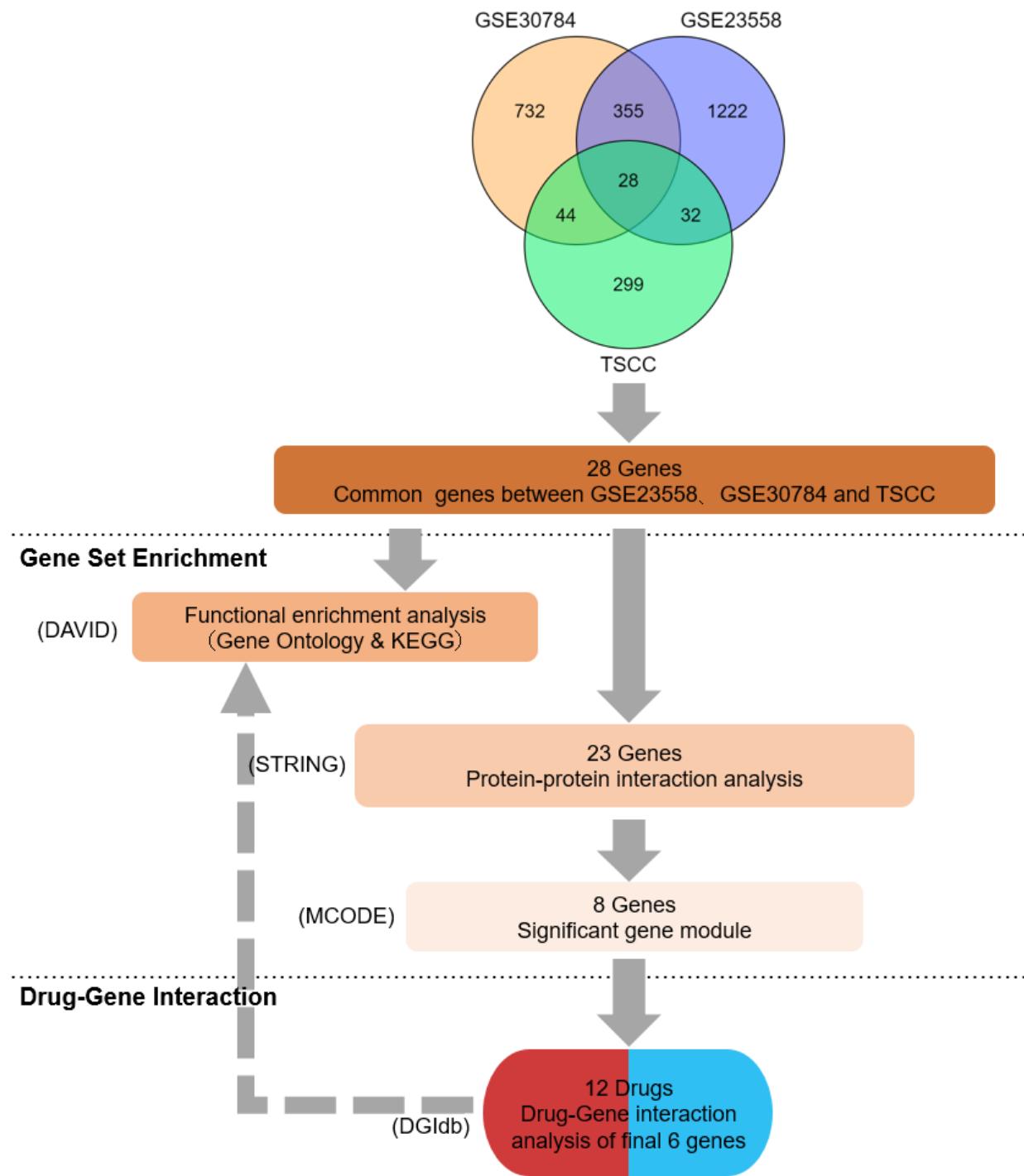
**Table II.** The detailed information of the 12 drugs and its target genes.

Number	Gene Names	Drug	Interaction	Drug class	FDA
1	<i>PLAU</i>	Upamostat	Inhibitor	Antineoplastic agent	Not approved
2	<i>SERPINE1</i>	Urokinase	Inducer, substrate	Thrombolytic agents, protein	Approved
3	<i>SERPINE1</i>	Aleplasinin	Inhibitory	Not available	Not approved
4	<i>MMP1</i>	Cipemastat	Inhibitor	Not available	Not approved
5	<i>MMP1</i>	Marimastat	Inhibitor	Antineoplastic agents, small molecule	Not approved
6	<i>MMP1</i>	Apratastat	Inhibitor	Not available	Not approved
7	<i>MMP1</i>	Doxycycline Calcium	Inhibitor	Not available	Not approved
8	<i>MMP1</i>	Doxycycline	Inhibitor	Not available	Not approved
9	<i>MMP1</i>	Batimastat	Inhibitor	Antineoplastic agent	Approved
10	<i>MMP3</i>	Marimastat	Inhibitor	Antineoplastic agent	Approved
11	<i>MMP3</i>	ChembI117225	Inhibitor	Not available	Not approved
12	<i>MMP3</i>	Batimastat	Inhibitor	Antineoplastic agent	Not approved
13	<i>CXCL10</i>	NI-0801	Inhibitor	Not available	Not approved
					Not approved
					Not approved
					Not approved
14	<i>SPP1</i>	ASK-8007	Inhibitor	Not available	Not approved

**Table III.** Gene ontology enrichment of identified eight genes.

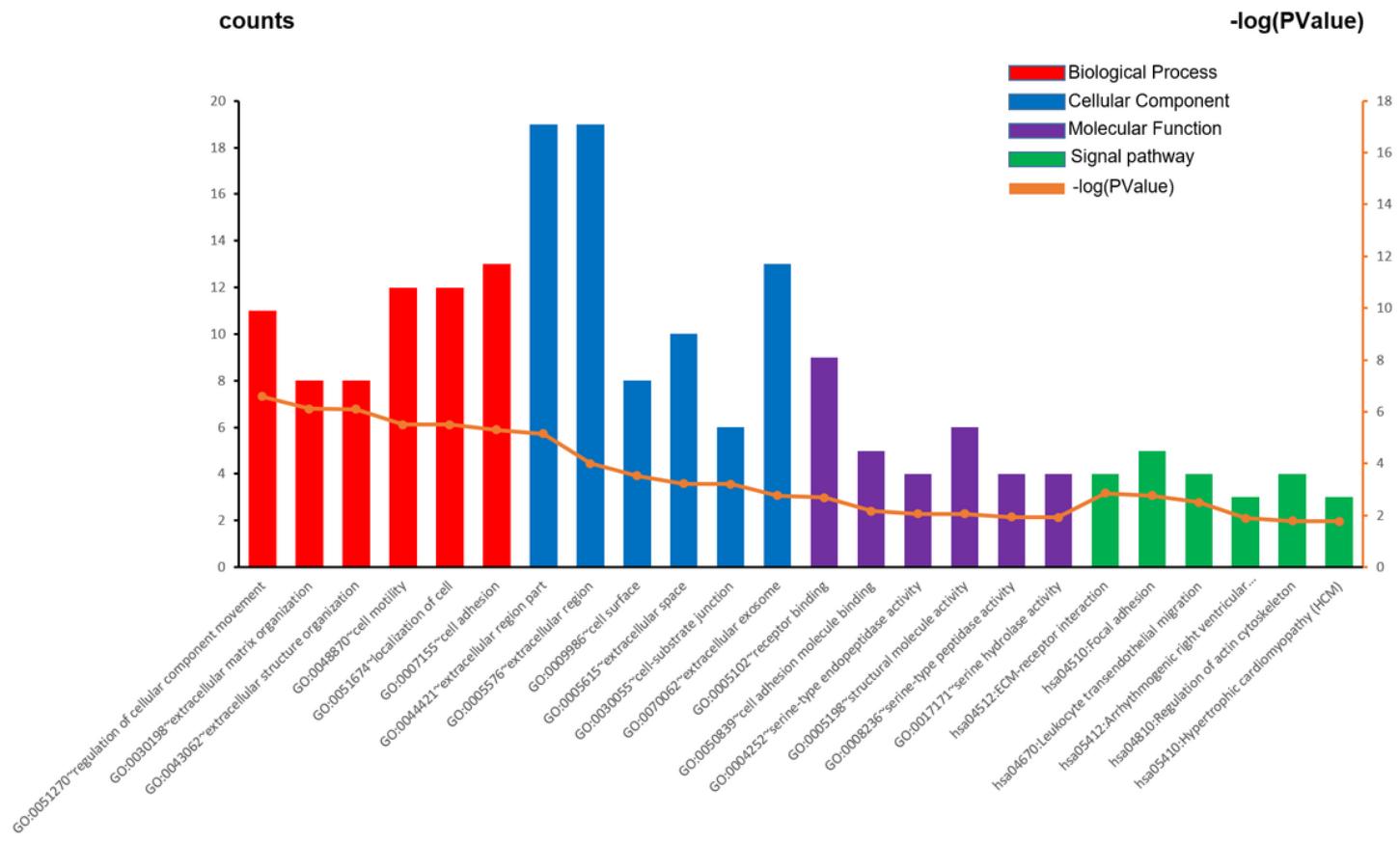
Category	ID	Term	Genes	Adj_pval
BP	GO:0022617	Extracellular matrix disassembly	<i>MMP1, MMP3, SPP1, MMP10</i>	4.92E-06
BP	GO:0030198	Extracellular matrix organization	<i>MMP1, MMP3, SERPINE1, SPP1, MMP10</i>	5.31E-06
BP	GO:0043062	Extracellular structure organization	<i>MMP1, MMP3, SERPINE1, SPP1, MMP10</i>	5.37E-06
BP	GO:0090026	Positive regulation of monocyte chemotaxis	<i>CXCL10, CXCL12, SERPINE1</i>	1.81E-05
BP	GO:0090025	Regulation of monocyte chemotaxis	<i>CXCL10, CXCL12, SERPINE1</i>	3.49E-05
BP	GO:0048870	Cell motility	<i>CXCL10, CXCL12, SERPINE1</i>	5.91E-05
CC	GO:0005615	extracellular space	<i>CXCL10, CXCL12, PLAU, MMP1, SERPINE1, MMP10</i>	6.12E-06
CC	GO:0044421	extracellular region part	<i>CXCL10, CXCL12, PLAU, MMP3, SERPINE1, SPP1, MMP10</i>	9.62E-05
CC	GO:0005576	extracellular region	<i>CXCL10, CXCL12, PLAU, MMP1, MMP3, SERPINE1, SPP1, MMP10</i>	3.30E-04
CC	GO:0031012	extracellular matrix	<i>CXCL10, CXCL12, PLAU, MMP1, MMP3, SERPINE1, SPP1, MMP10</i>	0.001514251
CC	GO:0005578	cell surface	<i>CXCL10, CXCL12, PLAU, MMP1, MMP3, SERPINE1, SPP1, MMP10</i>	0.011906334
CC	GO:0009986	serine-type endopeptidase activity	<i>MMP1, MMP3, MMP10</i>	0.048747842
MF	GO:0004252	serine-type peptidase activity	<i>CXCL10, CXCL12, PLAU, MMP1, MMP3, SERPINE1, SPP1, MMP10</i>	1.47E-04
MF	GO:0008236	serine hydrolase activity	<i>PLAU, MMP1, MMP3, MMP10</i>	1.98E-04
MF	GO:0017171	endopeptidase activity	<i>PLAU, MMP1, MMP3, MMP10</i>	2.05E-04
MF	GO:0004175	metalloendopeptidase activity	<i>PLAU, MMP1, MMP3, MMP10</i>	0.001033625
MF	GO:0004222	peptidase activity, acting on L-amino acid peptides	<i>PLAU, MMP1, MMP3, MMP10</i>	0.001083222
MF	GO:0070011		<i>MMP1, MMP3, MMP10</i>	0.002904593

## Figures



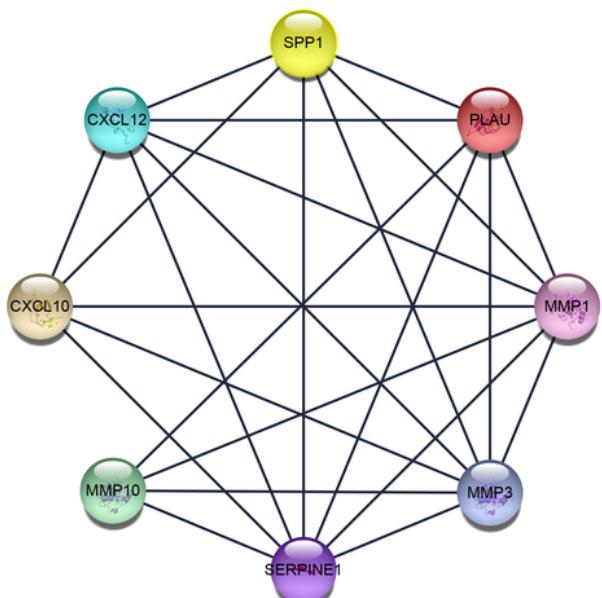
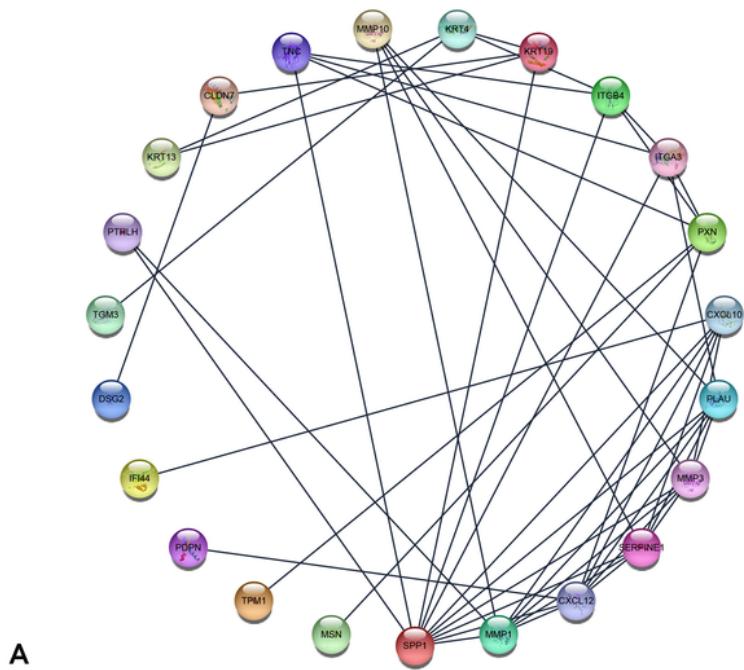
**Figure 1**

Summary of research process design. Text mining was conducted using pubmed2ensembl to identify genes associated with TSCC. DAVID was then used to analyze the gene ontology (GO) biological process terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway for the intersection genes. Gene enrichment was analyzed using STRING and MCODE. Drug-gene interaction database (DGIdb) was used to obtain the final list of enriched genes and their corresponding drugs.



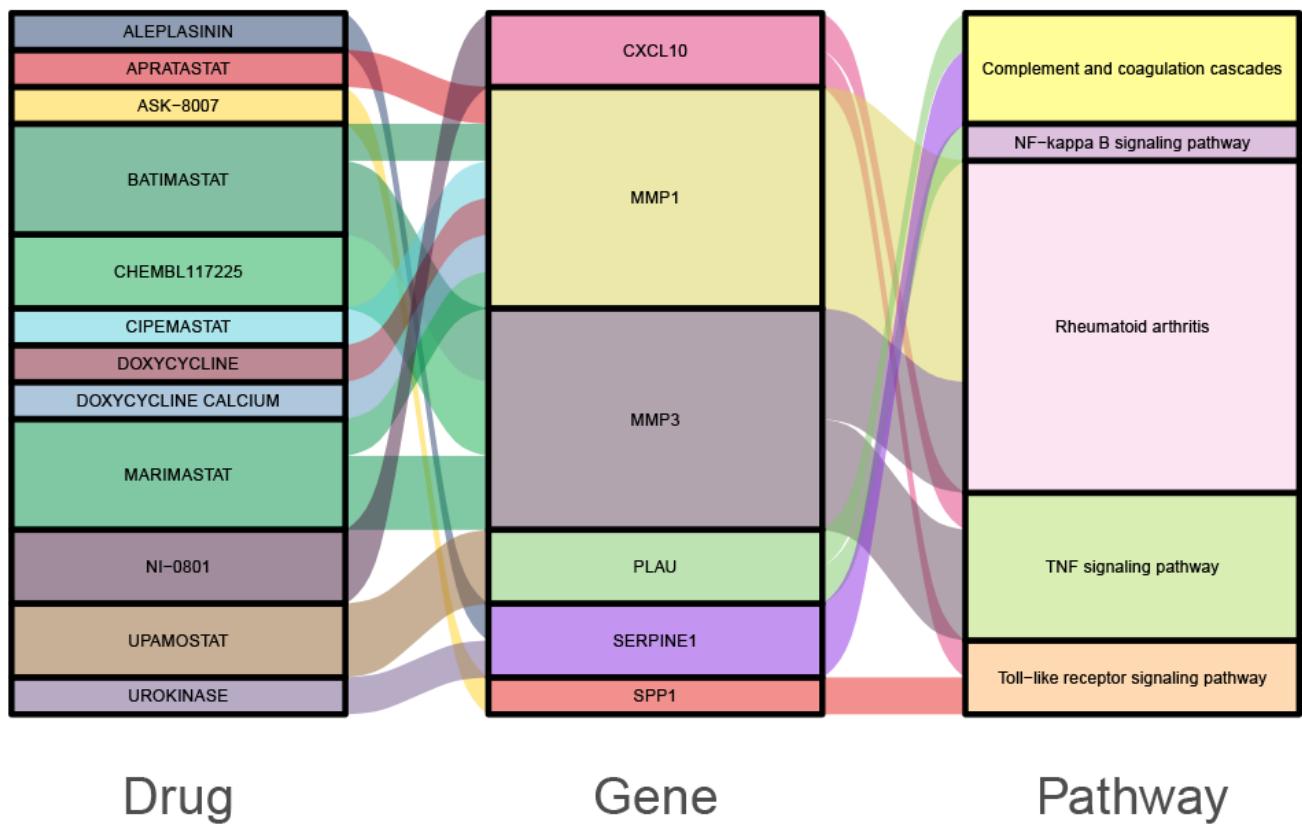
**Figure 2**

The top six significant gene ontology terms and signal pathway enrichments. The bar chart shows that the common genes enrichment in the BP , CC, MF and KEGG respectively, the yellow line chart represents the significance of enrichment terms. GO, gene ontology, BP , biological process, CC, cellular component, MF , molecular function, KEGG, Kyoto Encyclopedia of Genes and Genomes.



**Figure 3**

The protein-protein interaction (PPI) networks construction and significant gene modules analysis. (A) The entire PPI networks of 23 genes, (B) the significant gene module, including 8 genes.



**Figure 4**

The Sankey diagram shows the main links between drugs, genes and pathways.