Integrated Data Analysis Implicates PLAU, SERPINE1, SPP1, and MMP1 as Prognostic Factors in Head and Neck Squamous Cell Carcinoma

Lixiao Chen¹, Xinjiang Ying², Bin Shen³, Xinwei Chen⁴, Pin Dong⁵

Abstract
Aim: Head and neck squamous cell carcinoma (HNSCC) ranks seventh in incidence among the most common types of cancer worldwide. Although smoking and drinking are risk factors for HNSCC, the specific causes of, and molecular mechanisms underlying, of HNSCC have not been identified. Despite improvements in tumor treatment technology over the last 40 years, the prognosis of patients with HNSCC has not changed significantly. Therefore, to improve HNSCC treatment strategies and for early diagnosis, it is important to study its pathogenesis, and to identify prognostic markers for it.

Materials and methods: We used an integrated bioinformatics approach to identify key pathogenic and prognostic genes involved in HNSCC and to reveal the potential underlying molecular mechanisms. The expression profiles of the GSE6631 and GSE107591 datasets were downloaded from the Gene Expression Omnibus (GEO) database, and the tertiary RNA-sequencing dataset of HNSCC and corresponding clinical information were downloaded from The Cancer Genome Atlas (TCGA). These three datasets were integrated to identify differentially expressed genes (DEGs), and DEGs were analyzed using bioinformatic tools, including R packages.

Results: We identified 83 DEGs among these datasets. Gene Ontology analysis showed that the biological functions of the identified DEGs are primarily associated with regulating extracellular signal cascades, epidermis development, adhesion, and other tumor cytology behaviors. Kyoto Encyclopedia of Genes and Genomes pathway analysis showed that these DEGs were mainly involved in the PI3K-Akt, human papillomavirus infection, and IL-17 signaling pathways. A protein–protein interaction network was constructed to reveal the 20 most closely related genes among the DEGs.

Conclusion: Four key genes in the network (PLAU, SERPINE1, SPP1, and MMP1) were demonstrated to have prognostic relevance based on their significant associations with survival in patients with HNSCC. We verified the high expression of SERPINE1 in hypopharyngeal carcinoma cell lines by qRT-PCR and Western blotting.

Clinical significance: Our findings would help elucidate the molecular mechanisms underlying the development and provide the possibility to improve the prognosis of HNSCC.

Keywords: Differentially expressed gene, Gene expression omnibus, Head and neck squamous cell carcinoma, Prognostic factor, The cancer genome atlas.

Journal of Postgraduate Medicine, Education and Research (2021): 10.5005/jp-journals-10028-1411

Introduction
Head and neck squamous cell carcinomas (HNSCCs) encompass tumors of various subsites, including those of the oral cavity, oropharynx, larynx, and hypopharynx. HNSCC is an aggressive epithelial malignancy, ranking seventh in incidence among the most common cancers worldwide, at approximately 36/100,000 men and 7/100,000 women.¹ Its incidence has increased annually,² with 600,000 new cases and >300,000 deaths caused by HNSCC globally each year.³ Over 90% of HNSCCs are squamous cell carcinomas that are associated with a poor prognosis, despite diverse treatments being available.⁴ The 5-year survival rate is approximately 35%,⁵ and nearly 50% of patients have locoregionally advanced disease.⁶ However, the causes, mechanisms, genetic markers, and underlying molecular events associated with HNSCC are not well defined. The use of alcohol and tobacco, immunosuppression, and exposure to high-risk human papilloma virus (HPV) are the predominant environmental risk factors for developing HNSCC. Nevertheless, not all smokers and drinkers develop cancer. Only 10–15% of smokers develop lung cancer, and fewer develop HNSCC, suggesting that there are other environmental or genetic factors at play.⁷ Several studies have indicated that the occurrence of oral and oropharyngeal HNSCCs, but not cancer of the nasopharynx or

¹-⁵Department of Otolaryngology–Head and Neck Surgery, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Corresponding Author: Pin Dong, Department of Otolaryngology–Head and Neck Surgery, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, e-mail: dongpin64@aliyun.com

How to cite this article: Chen L, Ying X, Shen B, et al. Integrated Data Analysis Implicates PLAU, SERPINE1, SPP1, and MMP1 as Prognostic Factors in Head and Neck Squamous Cell Carcinoma. J Postgrad Med Edu Res 2021;55(4):160–170.

Source of support: This work was supported by grants from the National Key R&D Program of China (Grant No. 2017YFB1102900), National Natural Science Foundation of China (Grant Nos. 81570902, 81902756, 81900952, 81800893), the Foundation of Science and Technology Commission of Shanghai Municipality (Grant No. 18511108200), Clinical Research Plan of SHDC (Grant No. 16CR2029B), Shanghai Jiao Tong University Foundation on Medical and Technological Joint Science Research (Grant No. ZH2018QNA25, YG2019QNA55) and a grant from the Institute of Medical Robotics, Shanghai Jiao Tong University (Grant No. IMR2018ZD02).

Conflict of interest: None

© The Author(s). 2021 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted use, distribution, and non-commercial reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Integrated Data Analysis Implicates PLAU, SERPINE1, SPP1, and MMP1 as Prognostic Factors in HNSCC

The molecular pathogenesis of HNSCC is gradually being revealed through genomic, proteomic, and transcriptomic approaches. Epidermal growth factor receptor (EGFR), nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), signal transducers and activators of transcription (STATs), and cyclin-dependent kinase inhibitor 2A (CDKN2A) have all been identified to play a key role in HNSCC. The roles of these proteins in clinical applications and prognosis are also being continuously studied. Bioinformatics, which combines molecular biology and information technology, has emerged as a valuable tool to study the underlying molecular mechanisms of various diseases and to identify tumor biomarkers.11 With advances in molecular biology techniques, abundant gene expression data are publicly available for a diverse set of neoplasms. The Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA), two important databases, provide information regarding the expression of coding or non-coding RNAs and clinical information submitted by research groups using different microarray platforms for analysis.12,13 They are widely used for the discovery of tumor biomarkers and mechanisms. Here, to identify the key genes involved in HNSCC, we applied an integrated bioinformatics approach using mRNA expression data from the GEO and TCGA datasets. These findings would help reveal the molecular mechanisms underlying the development of HNSCC and associated tumor biomarkers.

Materials and Methods

Microarray Datasets

The keyword “HNSCC” was used to search the GEO DataSets database (https://www.ncbi.nlm.nih.gov/geo/), and gene expression profiles from the GSE10759114 and GSE663115 datasets were obtained. The GPL6244 [HuGene-1.0-st] Affymetrix Human Gene 1.0 ST Array platform was used for the GSE107591 dataset, which includes 23 normal and 24 tumor tissues. The GPL8300 [HG_U95Av2] Affymetrix Human Genome U95 Version 2 Array platform was used for the GSE6631 dataset, which includes 22 normal and 22 tumor tissues. Detailed information about the GEO datasets is provided in Table 1. We also downloaded an RNA-sequencing dataset of mRNA level 3 and corresponding clinical information for 544 patients from the TCGA, including detailed biological data from 500 tumor tissue samples and 44 samples of adjacent normal tissue. R v. 3.6.0 was used to process the downloaded files for conversion and filtering of unqualified data with the GEOquery, dplyr, limma, heatmap, ggplot2, and clusterProfiler packages, along with analysis tools from the STRING, Cytoscape, and Gene Expression Profiling Interactive Analysis (GEPIA) databases.

Data Processing and Screening of Differentially Expressed Genes Associated with HNSCC

We employed R language software and annotation packages to handle the downloaded platform data and series of matrix files. Specifically, GEO data were calibrated, standardized, and log2-transformed to obtain expression levels using the R packages limma, dplyr, and tibble. The platform IDs corresponding to the probe names were converted to standard recognized gene symbols and preserved in an R data file. TCGA data were calibrated and standardized, and differential gene expression analysis and annotation were performed using the R packages dplyr, lapply, tidyr, DESeq2, and edgeR. DEGs (genes that were differentially expressed between HNSCC and normal tissue) were identified according to the criteria |log2 fold change (FC)| ≥ 1 and adjusted p value < 0.05. Heatmaps and volcano plots were created using the R package ggplot2. Subsequently, we used the R venn package to identify DEGs common to the three datasets.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Pathway Enrichment Analyses of Differentially Expressed Genes

Gene ontology (GO) annotation includes three categories: biological process, cellular component, and molecular function. Therefore, GO enrichment analysis can offer a crude understanding of the biological functions, pathways, or cell localization of DEGs. GO analysis was performed using the R package clusterProfiler.16 Kyoto encyclopedia of genes and genomes is a database for associating molecular-level information with the high-level functions and utilities of a biological system at the cell, organism, and ecosystem levels and is widely used for analyzing large-scale molecular datasets generated by genome sequencing and other high-throughput techniques. We also used the R clusterProfiler package to search the key pathways associated with DEGs via the KEGG database. A p value <0.05 was considered to indicate statistically significant enrichment of a gene with a given KEGG pathway or GO term.

Protein–protein Interaction Network Integration

A protein–protein interaction (PPI) refers to the process in which two or more proteins form complexes through noncovalent bonds; thus, a PPI network can be used to identify interactions between known and predicted proteins. We used the STRING database (http://string-db.org) to identify PPIs for the DEGs found to be associated with HNSCC. The hub genes in the network represent candidate key or core genes with important physiological regulatory functions. We exported the results as a simple tab-separated value (TSV) table, which were inputted to Cytoscape v. 3.7.2 for network visualization analysis to select the hub genes.

Association of Patient Survival with Hub Genes

We used GEPIA (http://gepia.cancer-pku.cn/) to analyze the associations of the identified hub genes with overall survival of patients with HNSCC. The GEPIA server has been running for the last 2 years and has processed 280,000 analysis requests for 110,000 users from 42 countries. For this analysis, TCGA data were classified into two groups of high and low expression based on the median value. Genes with expression levels significantly (p value < 0.05) associated with overall survival were considered as candidate markers of HNSCC prognosis.

Table 1: Details of the head and neck squamous cell cancer (HNSCC) datasets for Homo sapiens in the GEO DataSets database

| GEO      | Platform | Total | Normal | Tumor | Series published |
|----------|----------|-------|--------|-------|-----------------|
| GSE107591| GPL6244  | 47    | 23     | 24    | 2017/12/10      |
| GSE6631  | GPL8300  | 44    | 22     | 22    | 2007/01/20      |

hypopharynx, is linked to HPV infection.8,9 Despite improvements in tumor treatment technology over the last 40 years, the prognosis of patients with HNSCC has not changed significantly.10
Cell Lines and Cell Culture
The human HNSCC cell line FaDu and a human immortalized normal mucosal cell line (DOK) were purchased from the Chinese Academy of Science Cell Bank (http://www.cellbank.org.cn/). FaDu cells were maintained in Minimum Essential Medium (MEM; Gibco, California, USA) supplemented with 10% fetal bovine serum (FBS; Gibco) and 1% penicillin-streptomycin liquid (Solarbio). DOK cells were grown in complete medium (RPMI-1640) supplemented with 10% FBS and 1% penicillin-streptomycin liquid (Solarbio). Both cell lines were grown at 37°C in a humidified chamber supplemented with 5% CO₂.

Quantitative Real-Time Polymerase Chain Reaction
Total RNAs were extracted from cells using Trizol Reagent (Takara Biotechnology, Dalian, China) in accordance with the manufacturer’s instructions. Approximately 1 μg of RNA was used for cDNA synthesis using a reverse transcription kit (RR036A; Takara), and quantitative real-time polymerase chain reaction (qRT-PCR) assays were conducted using a SYBR Green Premix Ex Taq kit (RR820A; Takara) in an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA) using the 2−ΔΔCt method. The primers used were as follows: SERPINE1, Forward primer 5′-AGAGCGCTGTCAAGAAGACC-3′; Reverse primer 5′-AGTTCTCAGAGGTGCTTGC-3′; GAPDH, Forward primer 5′-CGGATTTGGTCGTATTGGGC-3′; Reverse primer 5′-TTCCGGTCTCAGCCATGTAG-3′.

Western Blotting
Cells were lysed in RIPA protein extraction reagent (Sangon Biotech, Shanghai, China) with a protease inhibitor. Total protein concentration was quantified using the Bio-Rad Protein Assay Kit. Approximately 50 μg of each protein extract was separated by SDS-PAGE on a 7.5% acrylamide gel; the resultant protein bands were subsequently transferred to PVDF membranes. Immunoblotting was performed using rabbit anti-SERPINE1 antibodies (1 μg/mL; ab66705, Abcam, China) and anti-GAPDH antibodies (2 × 10⁻⁴ μg/mL; ab181602, Abcam, China).

Results
Identification of Differentially Expressed Genes
The standardized data of the HNSCC expression microarray datasets (GSE107591 and GSE6631) are summarized in Figure 1. DEGs (400), comprising 254 upregulated and 146 downregulated genes, were identified using the GSE107591 dataset, which contains 18,832 genes. Under the same criteria, 162 DEGs (64 downregulated and 98 upregulated) were identified using the GSE6631 dataset. In
addition, 4800 DEGs (2434 downregulated and 2366 upregulated) were identified using the TCGA dataset. The DEGs identified in the three microarray datasets are shown in Figure 2. For simplicity, gene symbols are shown only for those with |log₂ FC| > 3 (Figs 2A and B) or |log₂ FC| > 9 (Fig. 2C). Hierarchical clustering of DEGs is shown using heatmaps (Fig. 3). A Venn diagram revealed 83 intersecting DEGs from three datasets (Fig. 4 and Supplementary Table S1).

GO Term Enrichment Analysis of DEGs
The top 10 enriched GO terms associated with the identified DEGs are presented in Figure 5. In the biological process group, the DEGs were mainly enriched in extracellular matrix organization, extracellular structure organization, epidermis development, skin development, cell–substrate adhesion, and regulation of cell–substrate adhesion, which are all processes related to the occurrence and development of HNSCC. In the cellular component group, the DEGs were mainly enriched in extracellular matrix, collagen-containing extracellular matrix, endoplasmic reticulum lumen, extracellular matrix component, basement membrane, and collagen trimer, which are mainly associated with cell proliferation. In the molecular function group, the DEGs were mainly enriched in the extracellular matrix structural constituent, endopeptidase activity, metalloproteinase activity, serine-type peptidase activity, serine hydrolase activity, and structural constituent conferring tensile strength, which are largely focused on regulating enzyme activity, cell proliferation, migration, and invasion.

KEGG Pathway Analysis of DEGs
KEGG pathway analysis demonstrated that the DEGs were mainly involved in the extracellular matrix (ECM)–receptor interaction, PI3K-Akt signaling pathway, HPV infection, interleukin (IL)-17 signaling pathway, amoebiasis, AGE-RAGE signaling pathway in diabetic complications, focal adhesion, relaxin signaling pathway, and protein digestion and absorption (Fig. 6). The specific results for the enrichment of DEGs in the ECM-receptor interaction pathway are shown in Supplementary Figure S1.

PPI Analysis of DEGs
The PPI network based on the DEGs has 89 edges, with 83 nodes and an average node degree of 2.14 (Fig. 7). The 20 most significantly upregulated genes in the network (FN1, MMP13, MMP3, CXCL8, COL4A1, COL1A1, SPP1, COL6A3, MMP9, COL4A2, MMP1, POSTN, PTHLH, PLAU, MMP12, COL5A2, SERPINE1, LUM, COL3A1, and MMP10) were screened as hub genes (Fig. 8), and their co-expression homologs in humans and other species.
Integrated Data Analysis Implicates PLAU, SERPINE1, SPP1, and MMP1 as Prognostic Factors in HNSCC

Patient Survival Associated with PLAU, SERPINE1, SPP1, and MMP1

GEPIA survival analysis identified the hub genes PLAU, SERPINE1, SPP1, and MMP1 as significantly associated with the overall survival of patients with HNSCC (Fig. 9).

Higher Expression of SERPINE1 in FaDu Cell Lines than in DOK Cell Lines

SERPINE1 was selected from the four key genes associated with overall survival, and was screened for cytological verification of its upregulation. We used a head and neck tumor cell line (FaDu) for the analysis; a human immortalized normal mucosal cell line (DOK) was used as a control. qPCR and Western blot results showed that SERPINE1 mRNA and protein expression levels were higher in the FaDu tumor cell line than in the DOK control cell line (Fig. 10).

**Discussion**

We used an integrated bioinformatics approach to identify key pathogenic and prognostic genes, which is an effective method to explore the underlying molecular mechanisms of HNSCC. Alterations in three types of genes—oncogenes, tumor-suppressor genes, and stability genes—are responsible for tumorigenesis. Mammalian cells have several safeguards to protect against the potentially lethal effects of cancer gene mutations. Therefore, mutations in several genes are generally required for the development of invasive
Integrative Data Analysis Implicates PLAU, SERPINE1, SPP1, and MMP1 as Prognostic Factors in HNSCC

Although all genes can undergo changes in mutation rates, only mutations in oncogenes and tumor-suppressor genes affect net cell growth and thereby confer a selective growth advantage on the mutant cells. Thus, the DEGs screened from the three databases are likely to be key factors in the cell proliferation network, and thus in the occurrence and development of HNSCC.

Three studies have sought to elucidate the mechanisms underlying the occurrence of HNSCC and two of these studies focused on a single gene or single cohort.18,19 These studies were unsuccessful in establishing clear and effective treatment options or targets for HNSCC, and the 5-year survival rate remains relatively low. Therefore, additional genes must be researched for their roles in the tumorigenesis, development, and prognosis of HNSCC. A good starting point is to methodically explore the contents of an existing database. Using three separate databases, the present study included the gene expression profiles from 635 samples, which were integrated for bioinformatic analysis, revealing 83 DEGs as candidate markers associated with HNSCC. Functional annotation showed that most of the DEGs associated with HNSCC are significantly enriched in functions related to cell proliferation and transcriptional activity.

Several matrix metalloproteinase (MMP) genes, which have been widely linked to promoting tumor invasion and metastasis when overexpressed,20,21 were upregulated in HNSCC. MMP1 was previously shown to play an important role in initiating tumor growth and promoting HNSCC cell invasion.22 However, another study revealed that MMP1 was not clinically useful as a diagnostic marker for HNSCC.23 The PPI network of these DEGs identified 20 hub genes, four of which (PLAU, SERPINE1, SPP1, MMP1) were identified by GEPIA as significantly associated with survival and are candidate prognostic markers in HNSCC patients. Urokinase-type plasminogen activator (PLAU) and serpin family E member 1 (SERPINE1) were previously identified as candidate biomarkers for HNSCC using an integrated bioinformatic analysis.24 We verified

![Fig. 5: Top 10 enriched gene ontology terms associated with the differentially expressed genes, comparing head and neck squamous cell carcinoma tissue and normal tissue](image)
Integrated Data Analysis Implicates PLAU, SERPINE1, SPP1, and MMP1 as Prognostic Factors in HNSCC

our findings for SERPINE1 by evaluating its mRNA expression using qPCR and its protein expression using Western blotting. SERPINE1 encodes a member of the serine proteinase inhibitor (serpin) superfamily, which is the principal inhibitor of tissue plasminogen activator and urokinase and hence an inhibitor of fibrinolysis. PLAU and SERPINE1 were identified as independent prognostic factors for predicting overall survival in HNSCC patients from the TCGA cohort; however, their combination had better prognostic value than either gene alone. 25–27 Secreted phosphoprotein 1 (SPP1) is involved in the attachment of osteoclasts to the mineralized bone matrix and binds to hydroxyapatite with high affinity. Moreover, SPP1 may also interact with the vitronectin receptor located in the cell membrane. Another study suggested that SPP1 protein and mRNA expression levels were upregulated in metastatic castration-resistant prostate cancer, and that SPP1 may regulate the androgen receptor signaling pathway. Therefore, SPP1 may be a new target for effective therapeutic approaches for the prevention of, and intervention in, drug resistance and metastasis, and also as a potential biomarker and target for the diagnosis and treatment of metastatic castrate-resistant prostate cancer. 28

Although our integrated bioinformatics approach revealed some key genes associated with HNSCC, the interpretation of their clinical value and prospects is limited without additional cytological and gene signal pathway research. Thus, further cytological verification is needed to better understand the role of these DEGs in HNSCC. However, the genes identified herein can serve as a starting point for investigating the causes and underlying mechanisms of HNSCC. These DEGs may show clinical relevance in the early diagnosis and prevention of HNSCC and may provide effective targets for treatment.

**Author Contributions**

Pin Dong and Xinwei Chen designed the study and provided technology and software. GLW, Xinjiang Ying and Bin Shen interpreted the data. Lixiao Chen drafted the initial manuscript and performed the experiments. All authors read and approved the final manuscript.

**Acknowledgments**

The authors thank the contributors of TCGA (https://cancergenome.nih.gov/) and GEO (https://www.ncbi.nlm.nih.gov/geo/) for sharing the HNSCC data on open access and thank for financial support. In addition, we would like to acknowledge the helpful comments and suggestions on this article received from our reviewers and editors.

**Data Statement**

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

**References**

1. Seiwert TY, Burtness B, Mehra R, et al., Safety and clinical activity of pembrolizumab for treatment of recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-012): an open-label, multicentre, phase 1b trial. vol 17; 2016.
Integrated Data Analysis Implicates PLAU, SERPINE1, SPP1, and MMP1 as Prognostic Factors in HNSCC

2. Licitra L, Felip E. Squamous cell carcinoma of the head and neck: ESMO clinical recommendations for diagnosis, treatment and follow-up. Ann Oncol 2009;20(Suppl 4):121–122. DOI: 10.1093/annonc/mdp149.

3. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136(5):E359–E386. DOI: 10.1002/ijc.29210.

4. Byeon HK, Ku M, Yang J. Beyond EGFR inhibition: multilateral combat strategies to stop the progression of head and neck cancer. Exp Mol Med 2019;51(1):8. DOI: 10.1038/s12276-018-0202-2.

5. Baxi SS, Pinheiro LC, Patil SM, et al. Causes of death in long-term survivors of head and neck cancer. Cancer-Am Cancer Soc 2014;120(10):1507–1513. DOI: 10.1002/cncr.28588.

6. Burtness B, Haddad R, Dinis J, et al. Afatinib vs placebo as adjuvant therapy after chemoradiotherapy in squamous cell carcinoma of the head and neck. Jama Oncol 2019;5(8):1170. DOI: 10.1001/jamaoncol.2019.1146.

7. Ho T, Wei Q, Sturgis EM. Epidemiology of carcinogen metabolism genes and risk of squamous cell carcinoma of the head and neck. Head Neck 2007;29(7):682–699. DOI: 10.1002/hed.20570.

8. Hughes RT, Beuerlein WJ, O’Neill SS, et al. Human papillomavirus-associated squamous cell carcinoma of the larynx or hypopharynx: clinical outcomes and implications for laryngeal preservation. Oral Oncol 2019;98:20–27. DOI: 10.1016/j.oraloncology.2019.09.008.

9. Dahm V, Haitel A, Kaider A, et al. Cancer stage and pack-years, but not p16 or HPV, are relevant for survival in hypopharyngeal and laryngeal squamous cell carcinomas. Eur Arch Otorhinolaryngol 2018;275(7):1837–1843. DOI: 10.1007/s00405-018-4997-1.

10. Gupta S, Kong W, Peng Y, et al. Temporal trends in the incidence and survival of cancers of the upper aerodigestive tract in Ontario and the United States. Int J Cancer 2009;125(9):2159–2165. DOI: 10.1002/ijc.24533.

11. Li S, Yang R, Sun X, et al. Identification of SPP1 as a promising biomarker to predict clinical outcome of lung adenocarcinoma individuals. Gene 2018;679:398–404. DOI: 10.1016/j.gene.2018.09.030.

12. Clough E, Barrett T. The gene expression omnibus database. Methods Mol Biol 2016;1418:93–110.

13. Lee JS. Exploring cancer genomic data from the cancer genome atlas project. BMRep 2016;49(11):607–611. DOI: 10.5483/BMRep.2016.49.11.145.

14. Cai J, Yu Y, Xu Y, et al. Exploring the role of Mir204/211 in HNSCC by the combination of bioinformatic analysis of ceRNA and transcription factor regulation. Oral Oncol 2019;96:153–160. DOI: 10.1016/j.oraloncology.2019.07.024.

15. Kuriakose MA, Chen WT, He ZM, et al. Selection and validation of differentially expressed genes in head and neck cancer. Cell Mol Life Sci 2004;61(11):1372–1383. DOI: 10.1007/s00018-004-4069-0.

16. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16(5):284–287. DOI: 10.1089/omi.2011.0118.

17. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med 2004;10(8):789–799. DOI: 10.1038/mm.2004.355.

18. Cramer JD, Burtness B, Le QT, et al. The changing therapeutic landscape of head and neck cancer. Nat Rev Clin Oncol 2019;16(11):669–683. DOI: 10.1038/s41571-019-0227-z.

19. Blanchard P, Baujat B, Holostenc V, et al. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): a comprehensive analysis by tumour site. Radiother Oncol 2011;100(1):33–40. DOI: 10.1016/j.radonc.2011.05.036.

20. Carpén T, Sorsa T, Jouhi L, et al. High levels of tissue inhibitor of metalloproteinase-1 (TIMP-1) in the serum are associated with poor prognosis in HPV-negative squamous cell oropharyngeal cancer. Cancer Immunol Immunother 2019;68(8):1263–1272. DOI: 10.1007/s00262-019-02362-4.

21. Nair RP, Timiri SP, Sunavala-Dossabhoy G. Discontinuous transduction of MMP-sensitized tousled in head and neck cancer. Mol Ther Oncolytics 2019;14:57–65. DOI: 10.1016/j.omto.2019.02.003.

22. Zinzindohoue F, Blons H, Hans S, et al. Single nucleotide polymorphisms in MMP1 and MMP3 gene promoters as risk factor in head and neck squamous cell carcinoma. Anticancer Res 2004;24(3b):2021–2026.

Fig. 7: Protein–protein interaction analysis of differentially expressed genes, comparing head and neck squamous cell carcinoma tissue and normal tissue. Circles represent the gene products and the lines represent the interactions between gene products.

Fig. 8: Top 20 differentially expressed genes, comparing head and neck squamous cell carcinoma tissue and normal tissue, visualized using the cytoscape platform.
Integrated Data Analysis Implicates PLAU, SERPINE1, SPP1, and MMP1 as Prognostic Factors in HNSCC

Fig. 9: Survival analysis of the top 20 differentially expressed genes, comparing head and neck squamous cell carcinoma tissue and normal tissue. In the GEPIA survival analysis, PLAU, SERPINE1, SPP1 and MMP1 were identified as being significantly associated with survival. These genes may be linked to the prognosis of patients with HNSCC.
Integrated Data Analysis Implicates PLAU, SERPINE1, SPP1, and MMP1 as Prognostic Factors in HNSCC

23. Kalfert D, Ludvikova M, Topolcan O, et al. Analysis of preoperative serum levels of MMP1, -2, and -9 in patients with site-specific head and neck squamous cell cancer. Anticancer Res 2014;34(12):7431–7441.

24. Yang K, Zhang S, Zhang D, et al. Identification of SERPINE1, PLAU and ACTA1 as biomarkers of head and neck squamous cell carcinoma based on integrated bioinformatics analysis. Int J Clin Oncol 2019;24(9):1030–1041. DOI: 10.1007/s10147-019-01435-9.

25. Xu X, Li M, Hu J, et al. Expression profile analysis identifies a two-gene signature for prediction of head and neck squamous cell carcinoma patient survival. J Cancer Res Ther 2018;14(7):1525–1534. DOI: 10.4103/jcrt.JCRT_557_18.

26. Hui L, Yang N, Yang H, et al. Identification of biomarkers with a tumor stage-dependent expression and exploration of the mechanism involved in laryngeal squamous cell carcinoma. Oncol Rep 2015;34(5):2627–2635. DOI: 10.3892/or.2015.4230.

27. Khan SS, Shah SJ, Klyachko E, et al. A null mutation in SERPINE1 protects against biological aging in humans. Sci Adv 2017;3(11):e1617. DOI: 10.1126/sciadv.aao1617.

28. Pang X, Xie R, Zhang Z, et al. Identification of SPP1 as an extracellular matrix signature for metastatic castration-resistant prostate cancer. Front Oncol 2019;9:924. DOI: 10.3389/fonc.2019.00924.

Figs 10A and B: SERPINE1 mRNA (A) and protein (B) expression levels in FaDu and DOK cell lines. SERPINE1 mRNA and protein expression was higher in the FaDu tumor cell line than in the DOK normal cell line.
Supplementary Table S1: The DGEs of HNSCC in those datasets

| DGEs | Gene names |
|------|------------|
| Upregulated | CRISP3, KRT4, MAL, KRT13, TGM3, ENDOU, PPP1R3C, SPINKS, EMP1, CYP3A5, SCEL, HOPX, APOD, COBL, GPD1L, SLURP1, SELERBP1, CRYAB, CEACAMS, CFD, ALOX12, GPX3, RRAGD, PPL, CLU, BARX2, MYH11, SASH1, NUCB2, CEACAM6, METTL7A, ACPR, MGLL, ECM1, ADH7, ALDH3A1, SCN1A, SERPINB1, BLNK, DUSP5, SULT2B1, AQP3, PTN, SPARC1L, CD24, ABLM1, CLDN7, IGBP5, PGO |
| Downregulated | MMP13, MMP1, MMP9, MMP3, MMP12, LAMC2, POSTN, PTHLH, FNI, MMP10, SPP1, COL5A2, LOXL2, COL1A1, SERPINE1, FAP, COL3A1, PLAU, COL4A1, COL6A3, NEFL, TNC, FEZ1, RBPI, IGFBP3, COL4A2, POPN, LUM, LAMB3, LPCAT1, CXCL8, MYO1B, KRT17, SERPINE2 |

**Supplementary Fig. S1:** The pathway of ECM-receptor interaction. It shows the pathway of ECM-receptor interaction associated with HNSCC.