



OPEN

## Identifying prognostic biomarkers in oral squamous cell carcinoma: an integrated single-cell and bulk RNA sequencing study on mitophagy-related genes

Minsi Li<sup>1,2</sup>, Yi Wei<sup>1,2</sup>, Wenhua Huang<sup>4</sup>, Cen Wang<sup>1</sup>, Shixi He<sup>1,3</sup>, Shuwen Bi<sup>5</sup>, Shuangyu Hu<sup>1,3</sup>, Ling You<sup>2,3</sup> & Xuanping Huang<sup>1,2</sup>✉

Oral squamous cell carcinoma (OSCC) has an extremely poor prognosis. Recent studies have suggested that mitophagy-related genes (MRGs) are closely correlated with the development and occurrence of cancer, but the role they play in oral cancer has not yet been explained. We conducted a comprehensive analysis of integrated single-cell and bulk RNA sequencing (RNA-seq) data retrieved from Gene Expression Omnibus (GEO) datasets and The Cancer Genome Atlas (TCGA) database. Multiple methods were combined to provide a comprehensive understanding of the genetic expression patterns and biology of OSCC, such as analysis of pseudotime series, CellChat cell communication, immune infiltration, Gene Ontology (GO), LASSO Cox regression, gene set variation analysis (GSVA), Kyoto Encyclopedia of Genes and Genomes (KEGG), gene set enrichment analysis (GSEA), Tumor Mutation Burden (TMB) and drug sensitivity assessments. The findings of this study demonstrated significantly greater activity of MRGs in NK cells than in other cells in OSCC. A reliable prognostic model was developed using 12 candidate genes strongly associated with mitochondrial autophagy. T stage, N stage and risk score were revealed as independent prognostic factors. Distinctively enriched pathways and immune cells were observed in different risk groups. Notably, low-risk patients were more responsive to chemotherapy. In addition, a nomogram model with excellent predictive ability was established by combining the risk scores and clinical features. The activity of MRGs suggest the potential for the development of new targeted therapies. The construction of a robust prognostic model also provides reference value for individualized prediction and clinical decision-making in patients with OSCC.

**Keywords** Mitochondrial autophagy, Single-cell RNA sequencing, Prognostic biomarkers, Immune, Oral squamous cell carcinoma

OSCC is a common type of tumor that has received worldwide attention, with 177,757 deaths and 377,713 new cases as of 2020<sup>1</sup>. The Global Cancer Observatory estimates that the number of patients with OSCC will continue to increase until 2030<sup>1</sup>.

More than 90% of oral malignant tumors are squamous cell carcinomas<sup>2</sup>, which can be categorized based on their anatomical location: the tongue, mouth floor, buccal mucosa, hard palate, gingiva, lips, posterior molar triangle and alveolar ridge<sup>3</sup>. Despite significant advancements in therapy and diagnosis, OSCC still has an extremely poor prognosis. The five-year survival rate is less than 60%, and the postoperative quality of life can

<sup>1</sup>Department of Oral and Maxillofacial Surgery, College & Hospital of Stomatology, Guangxi Medical University, NO.10 Shuangyong Road, Nanning 530021, China. <sup>2</sup>Guangxi Key Laboratory of Oral and Maxillofacial Rehabilitation and Reconstruction, Nanning 530021, China. <sup>3</sup>Guangxi Clinical Research Center for Craniofacial Deformity, Guangxi Medical University, Nanning 530021, China. <sup>4</sup>Department of Radiation Oncology, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, China. <sup>5</sup>Department of Pathology, Beihai People's Hospital (Ninth Affiliated Hospital of Guangxi Medical University), Beihai 536000, China. ✉email: hxp120@126.com

be unsatisfactory. Oral cancer also constitutes a significant portion of the global cancer burden<sup>4</sup>. Therefore, identifying potential molecular biomarkers and therapeutic targets is crucial for early diagnosis of OSCC and improving clinical treatment outcomes.

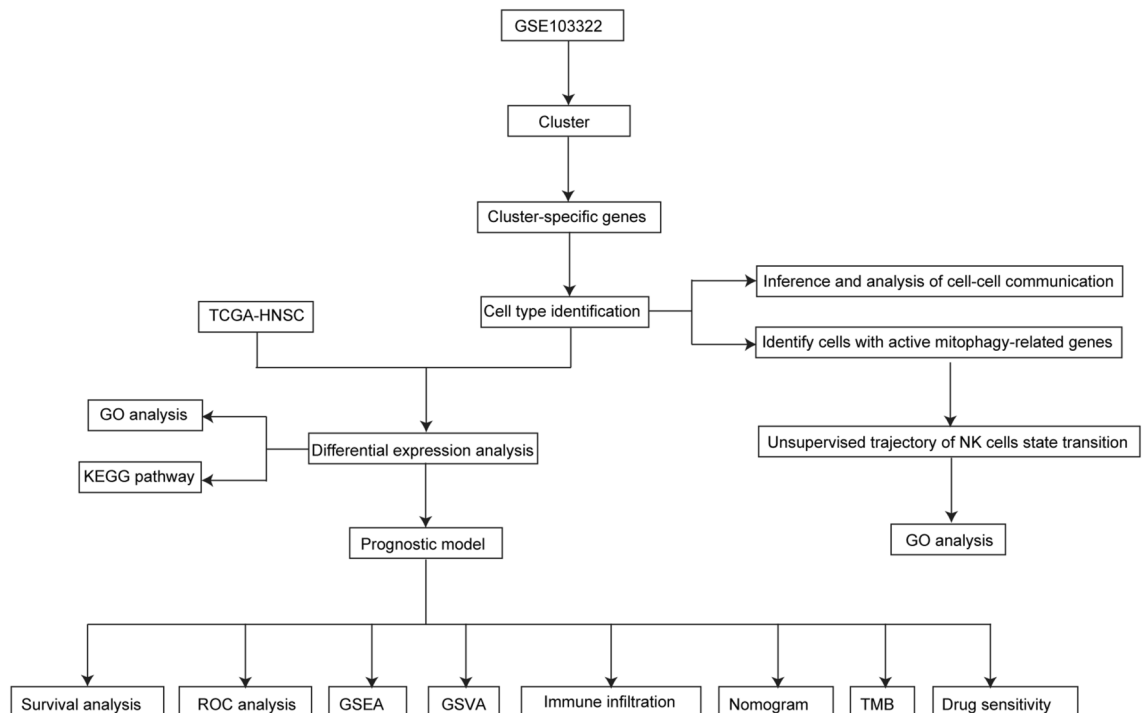
Mitochondria are the main energy producers within cells, and their normal functioning is crucial for processes such as cell proliferation, metabolism, migration and death. Any interference or damage during this period causes mitochondrial dysfunction, leading to the occurrence of tumors<sup>5</sup>. Mitochondrial autophagy is a type of selective autophagy that aims to maintain cellular homeostasis by selectively clearing excess or functionally impaired mitochondria to maintain a balance between the number and quality of mitochondria. Autophagosomes specifically encapsulate and fuse with lysosomes containing excess or damaged mitochondria, resulting in mitochondrial degradation<sup>6</sup>. The maintenance of mitochondrial homeostasis and cell survival rely on the dynamic balance of mitochondrial autophagy, which also aids in cell survival during inflammatory conditions. Therefore, mitochondrial autophagy may be a potential anticancer target in tumor cells. Recent research has demonstrated that mitochondrial autophagy promotes cancer cell death. He et al. discovered a potential antitumor drug, tanshinone IIA, which inhibits the protein kinase pathway activated by adenosine monophosphate, inactivates parkin, and reduces mitochondrial autophagy activity, thereby promoting apoptosis in colorectal cancer cells<sup>7</sup>. Chang et al<sup>8</sup> reported that upregulated expression of the protooncogene ROS 1 (located in mitochondria) is associated with the invasiveness of OSCC. Upregulation of ROS1 promotes fragmented changes in mitochondrial morphology, improves the respiratory levels of cancer cells, and reduces mitochondrial biogenesis, thereby enhancing OSCC invasiveness. There are many differences between the mitochondria in cancer cells and those in normal cells<sup>9</sup>, and these findings can guide the study of cancer cell apoptosis and lead to the exploration of novel targets for OSCC treatment.

scRNA-seq is a cutting-edge technique that allows for the analysis of transcriptomes in individual cells, providing comprehensive transcriptional information for the entire genome<sup>10</sup>. It is well suited for exploring the heterogeneity of immune cells<sup>11</sup>. Additionally, scRNA-seq allows for the concurrent detection of gene expression signatures in thousands of single cells<sup>12</sup>. We used scRNA-seq in this study to analyze distinct cell populations in human oral cancer, investigate their mitochondrial autophagy-related gene activity and expression patterns, and further investigate the function of these heterogeneous cells as well as their communication with other cell types within the tumor microenvironment. Drawing from a substantial dataset of tumor transcriptome information and clinical data from the TCGA, we identified genes linked to survival and characterized their prognostic features. The prognostic significance of these genes was further validated across various risk subgroups and patient cohorts.

## Methods

### Download and organization of conventional transcriptome data

Figure 1 outlines the workflow of the research. By utilizing the TCGA database and Bioconductor package “TCGAbiolinks” (version 2.25.0)<sup>13</sup>, we downloaded OSCC whole-genome expression profile data and clinical information in the TPM format (<https://portal.gdc.cancer.gov/>). The “VarScan2 Variant Aggregation and



**Figure 1.** Research flowchart.

Masking” tool was used to predict single nucleotide mutation (SNV) data. The TCGA-HNSC dataset (n = 566) included 522 tumor samples, 44 normal control samples. By selecting “Primarist” and “Osser Andil-Defined-sitesInlip, Olarcal Cavetiende Fernks”, we obtained 71 oral cancer samples and 13 oral control samples. This study respects the data access strategies of each database.

### Download and processing of scRNA-seq data

We obtained the scRNA-seq dataset GSE103322 for oral cancer from the GEO website (<https://www.ncbi.nlm.nih.gov/geo/>). The dataset included 18 patients’ oral cancer specimens. The dataset was read using “Seurat” (version 4.2.0)<sup>14</sup>, and the data were normalized by the “normalizeddata” function. By balancing average expression with dispersion, genes with high variation were identified in single cells. Then, principal component analysis (PCA) was conducted using “FindClusters” based on the shared nearest neighbor (SNN) modular optimization clustering algorithm to generate 19 clusters on 30 significant principal components (PCs) with a resolution of 0.4. The “Runtstne” algorithm performs t-distributed stochastic neighbor embedding (t-SNE). t-SNE-1 and t-SNE-2 demonstrate cell aggregation. To identify differentially expressed genes (DEGs), we used “FindAllMarkers” with default parameters set by Seurat to the normalized gene expression data. Subsequently, through cell type-specific biomarkers we identified cell clusters, calculated and evaluated the proportion of cell types.

### Score of mitophagy-related genes

Using the R package AUCCell<sup>15</sup>, each cell was scored based on GSEA with 73 selected mitochondrial autophagy-related genes (Table S1)<sup>16</sup> and their corresponding area under the curve (AUC) values. This generated gene expression rankings for each cell, enabling estimation of the proportion of highly expressed genes. Cells that expressed a greater number of genes exhibited a higher AUC score. Subsequently, “ggplot2” (version 3.3.5) was utilized to map the area under the curve (AUC) of each cell to the t-SNE embedding, visualizing activated cell clusters.

### Constructing cell trajectories via pseudotime analysis

Pseudotime analysis generates a pseudotime graph through analysis of reverse graph embedding for genes defined by the user by Monocle 2<sup>17</sup>, a leading tool for scRNA-seq analysis and time inference. This approach provides insights into branching and linear differentiation processes. A pseudotime trajectory was constructed using genes with high dispersion and expression (estimated value of dispersion  $\geq 1$  and average expression  $\geq 0.1$ )<sup>18</sup>. The parameters of the DDRtree algorithm were selected using default values. To examine the branching events in greater depth, we utilized branch expression analysis modeling, which was integrated into Monocle 2. This approach facilitates the identification of all genes with notable branch-specific expression and visualizes them as a heatmap, enabling quantitative analysis of gene expression changes during cell fate decisions<sup>17</sup>.

### Cellular communication analysis

Cell communication analysis was employed to determine the incoming and outgoing communication pathways and ligand–receptor pairs<sup>19,20</sup>. In this study, we used CellChat (version 1.1.3) to calculate and analyze intercellular communication for each cell type in OSCC samples, maintaining the software’s default parameters with a significance threshold of  $P \leq 0.05$ . The Benjamini–Hochberg (BH) method was used to correct the adjusted  $P$  value and conduct an in-depth analysis of NK cells, revealing the extent of communication among all signaling pathways and visualizing them.

### GO and KEGG pathway enrichment

GO analysis can achieve the goal of annotating and classifying genes according to their different functions<sup>21</sup>. KEGG is a biological informatics resource utilized to explore rich and significantly altered metabolic pathways in gene lists<sup>22</sup>. GO and KEGG analyses of DEGs related to mitochondrial autophagy in OSCC were also conducted using “clusterProfiler (version 4.2.2)” (adjusted  $p$  value  $< 0.05$ )<sup>23</sup>.

### Development and validation of prognostic models

To evaluate the prognostic significance of DEGs linked to mitochondrial autophagy in OSCC and identify key genes associated with patient survival, we used univariate Cox regression analysis to assess the association between each gene and overall survival (OS) in tumor samples. Genes ( $P \leq 0.05$ ) were considered strongly associated with OS for further analysis. A total of 442 patients were randomly separated into training or validation cohorts at a 7:3 ratio. Then, least absolute shrinkage and selection operator (LASSO)-Cox regression analysis<sup>24</sup> was applied to these prognostic candidates in training cohorts. Finally, by choosing the optimal penalty parameter  $\lambda$  correlated with the minimum tenfold crossvalidation, we established a twelve-gene optimal prognostic model. The following formula was used to calculate the risk score:

$$\text{riskScore} = \sum_{i=1}^n \text{Coef}(\text{gene}_i) * \text{Expression}(\text{gene}_i)$$

(Coef (gene<sub>i</sub>): risk coefficient, Expression (gene<sub>i</sub>): gene expression level).

A median split approach was used to assign patients to different risk groups. The Kaplan–Meier (KM) method is a nonparametric method for assessing intergroup survival differences. We utilized the receiver operating characteristic (ROC) curve to validate the effectiveness of this prognostic model. The area under the curve (AUC) varies from 0.5 to 1, with a closer value to 1 indicating superior performance.

### GSEA and GSVA

GSEA is a computational method that examines whether a predefined set of genes displays any notable distinctions in terms of their expression patterns across two distinct biological conditions<sup>25</sup>. The algorithm arranges all genes in order based on log<sub>2</sub>FC through “clusterProfiler (version 4.2.2)”, with 1000 gene set permutations performed for each analysis (adjusted *p* value < 0.05). This study employed “c2. cp. kegg. v7.5.1. symbols” as the guide gene set and was stored in the Molecular Signatures Database<sup>25,26</sup>. To compare the biological functional disparities between high and low risk groups, GSVA (version 1.42.0) was performed and visualized through “pheatmap (version 1.0.12)” (adjusted *p* value < 0.05).

### Nomogram model

We sourced clinical data on patient tumor stage, sex and age from the TCGA. We performed univariate and multivariate Cox regression analyses using risk scores derived from clinical data and prognostic models. To predict OS at one, three, and five years, we developed a nomogram model using the “RMS” package and evaluated its performance through calibration curves and time-dependent receiver operating characteristic (ROC) curves, emphasizing its clinical and prognostic significance.

### Immune infiltration

Single-sample gene set enrichment analysis (ssGSEA) allows the calculation of individual enrichment scores for every pair of samples and specific gene sets<sup>27</sup>. These scores represent the extent to which genes within a specific gene set are coordinated in a given sample. We sourced 28 distinct immune cell types from the Tumor and Immune System Interactions Database (TISIDB)<sup>28</sup> (Table S2) and calculated the relative enrichment score. We employed “ggplot2”<sup>29</sup> to visualize the differences in immune infiltration among the different risk groups, providing a comprehensive analysis of immune cell responses within tumors.

### Drug sensitivity assessment and TMB

We downloaded the half-maximal inhibitory concentration (IC<sub>50</sub>) and corresponding gene expression data from the Genomics of Disease Study (GDSC) database (<https://www.cancerrxgene.org/>)<sup>30</sup>. To predict the potential therapeutic drug sensitivity of patients with high- and low-risk OSCC, we employed “OncoPredict (version 0.2)”<sup>31</sup>. “Maftools” was employed to visualize the landscape of somatic mutations across distinct clusters, encompassing single nucleotide polymorphisms, insertions, deletions, tumor mutation burden, and mutation frequency<sup>32</sup>. Genes that were frequently mutated and ranked within the top 20 in mutation frequency were deemed the principal driver genes for malignant tumors<sup>33</sup>.

### Statistical analysis

Survival analysis was conducted using the Kaplan–Meier (KM) curve and log-rank test to compare the survival rates between the two groups. All survival curves were generated using the ‘survminer’ package. Either one-way analysis or double-tailed *t* tests of variance were used to establish significant quantitative disparities in variables with a normal distribution. The Kruskal–Wallis test or Wilcoxon test was applied to assess significant differences in data with abnormal distributions. R version 4.3.0 was utilized for diverse statistical analyses, with the significance threshold set to *P* < 0.05.

## Results

### Single-cell dimensionality reduction, clustering and annotation

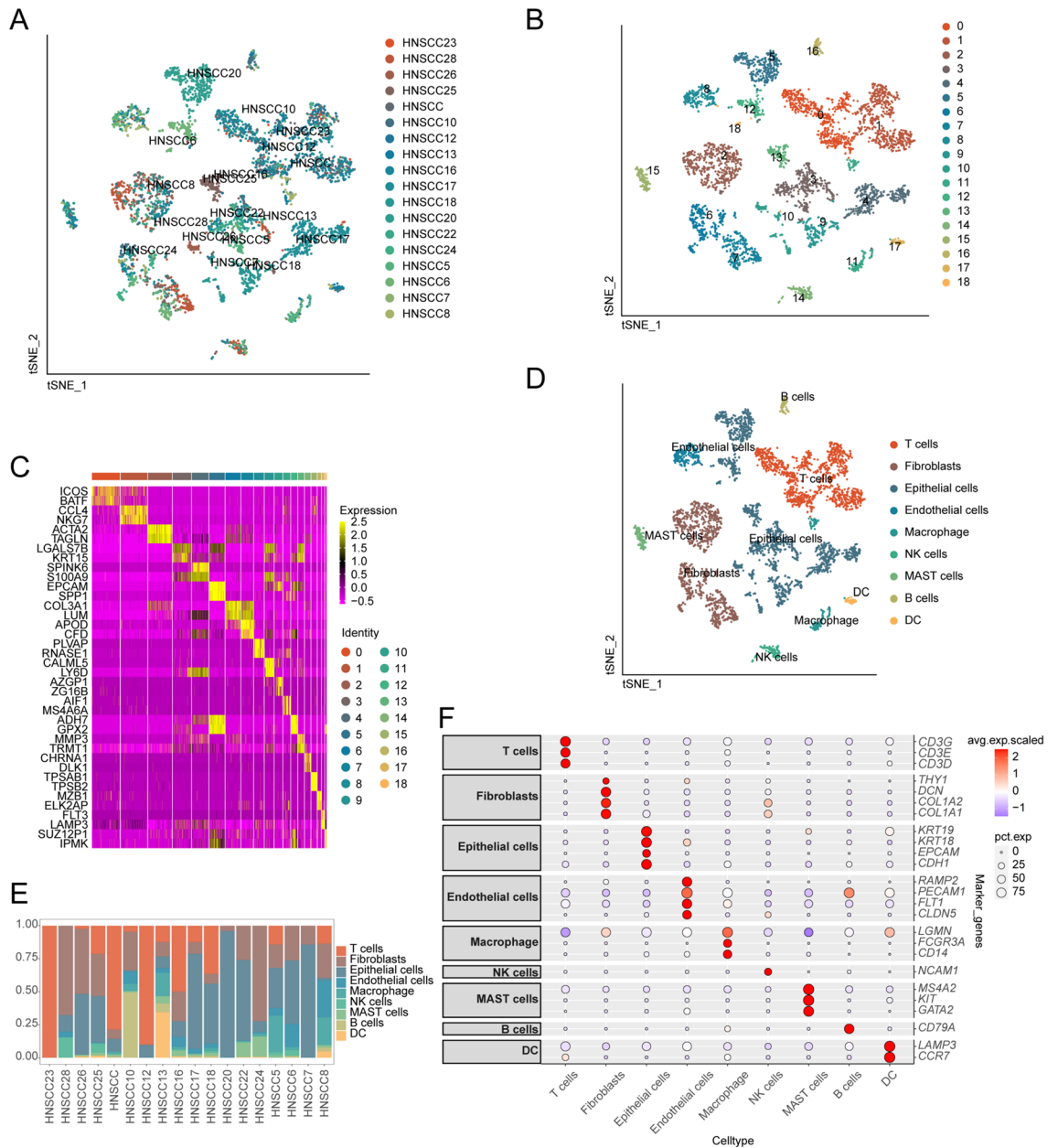
To explore the sources of the highly expressed genes, we analyzed the OSCC cell population using the scRNA-seq dataset GSE103322, which included 4,541 single-cell transcriptomes from a sample of 18 patients. These cells were subsequently grouped into 19 clusters (Fig. 2B). The two most significantly differentially expressed genes in each cluster were displayed using a heatmap (Fig. 2C). The distinct cell types were annotated according to cell-specific biomarkers and expression patterns (Table S2). As shown in Fig. 2D, nine cell types were identified: epithelial cells, macrophages, and T cells. The proportions of the various cell types in each patient sample are presented in Fig. 2E. Specific genes of each cell type were visualized using point plots (Fig. 2F).

### Determination of mitochondrial autophagy activity and pseudotime analysis

The identification of active cell subpopulations at single-cell resolution relies on the expression patterns of MRGs. After determining the optimal threshold for active cells, 1,354 active mitochondrial autophagic cells were identified (Fig. 3A). The activity of MRGs in NK cells was markedly greater than that in other cell types (*p* < 0.0001; Fig. 3B). Using the identified NK cells, we established a pseudotime cell trajectory to identify the important gene expression programs that determine OSCC progression. The transcriptional states in the trajectory revealed different processes (Fig. 3C–F). To elucidate the molecular basis of this transformation, we explored the genes that determine the branching of OSCC cells. The highly expressed genes in Cell fate 1 were enriched mainly in GO CC pathways such as “CatSper complex”, “condensed chromosome” and “male germ cell nucleus” (Fig. 3G).

### Cellular communication analysis

To further elucidate the comprehensive role of these immune cells, intercellular communication analysis was conducted. The interactions between these cell types are shown in Fig. 4A–B. Further research was conducted on potential efferent and afferent signals and specific ligand–receptor pairs in these nine cell types. As shown in Fig. 4C–D, NK cells have the least number of signaling receptors; however, they are the main signaling providers, and the potential signaling pathways in NK cells include FGF, LEP, CSF3, EDN and ANGPTL3. Subsequently,

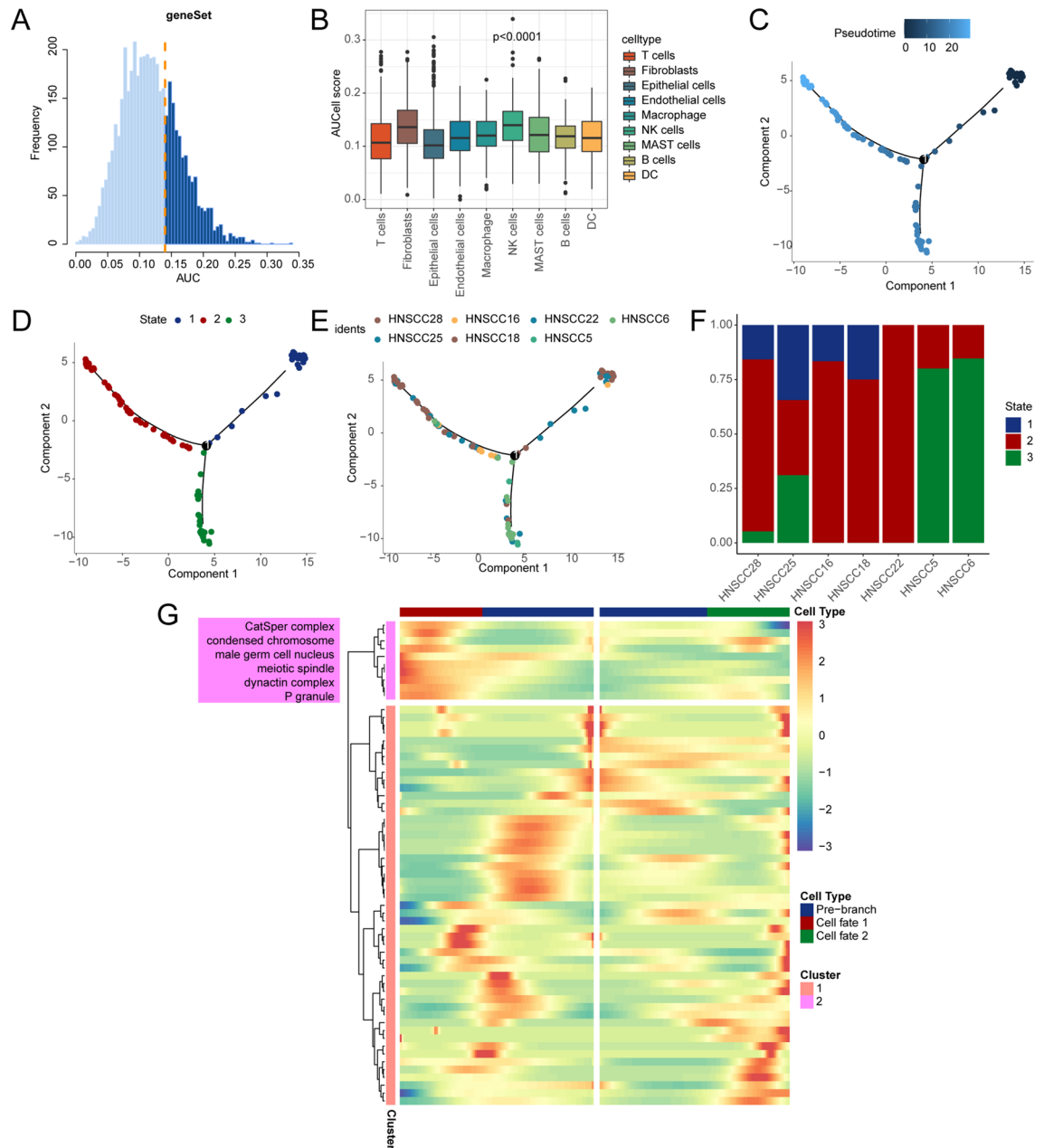


**Figure 2.** Shows the cell subpopulations identified from the scRNA-seq data. (A) t-SNE map of the distribution of 18 OSCC samples. (B) t-SNE map shows the distribution of 19 cell subsets in OSCC. (C) Heatmap displaying genes specifically expressed in the cell cluster. (D) Annotated t-SNE map of oral cancer cell subpopulations. (E) The histogram displays the distributions of different cell types in 18 OSCC patients. (F) Expression of key genes across different cell types.

specific signaling pairs between NK cells and other types of cells were studied. The findings revealed that the strongest interaction between NK cells and other cell types occurred through the MIF (CD74 + CXCR4) ligand–receptor pair from NK cells to T cells, B cells, and DCs (Fig. S1). The ligand MIF was expressed in various cells, as was the receptor CD74. The receptor CXCR4 is predominantly expressed in T cells, B cells, macrophages, DCs and NK cells, whereas CD44 is expressed in all cell types except endothelial cells (Fig. S1). These preliminary results elucidate the potential interactions between NK cells and other cell types and will help us further explore the comprehensive function of NK cells in OSCC.

### Enrichment analysis of DEGs in OSCC

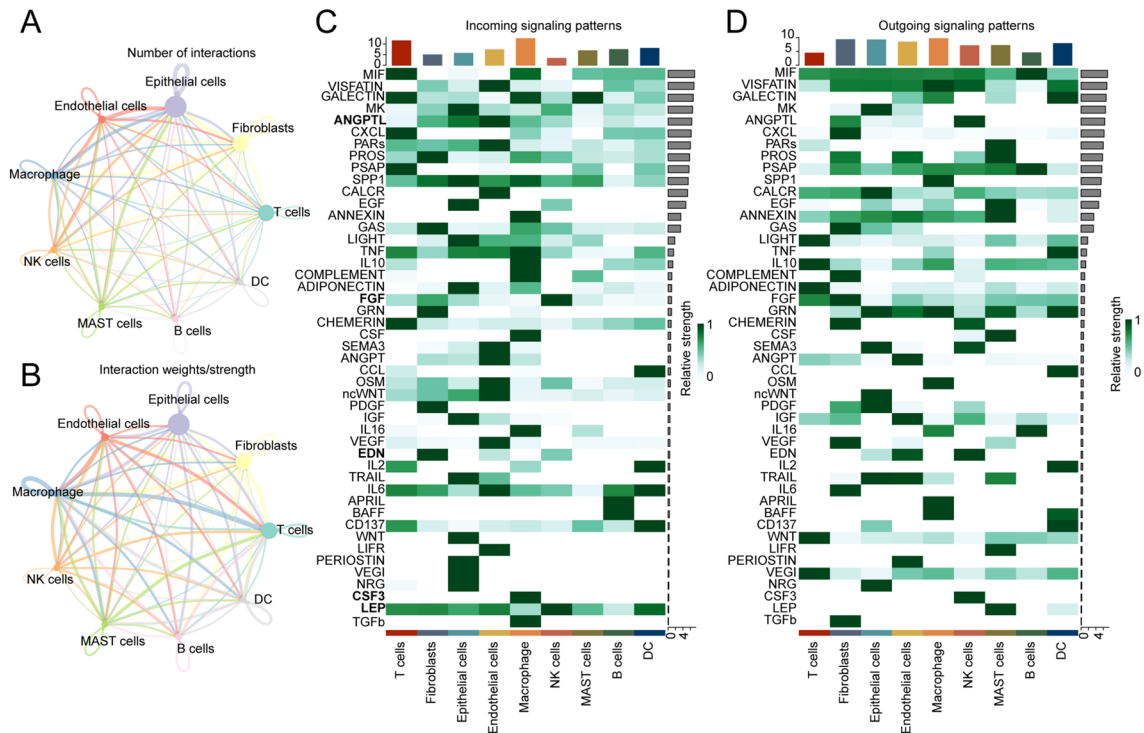
There were 1,192 DEGs between NK cells and other cell types ( $|\text{Log}_2\text{-fold change}| > 0.25$ , adjusted  $p$  value  $< 0.05$ ; Table S3). The top 10 upregulated genes (CHRNA1, DLK1, MYF5, MUSK, CHRDL2, DES, SGCA.2, APOC1, SERPINA3, and MEG3) in NK cells and the top 10 downregulated genes (ARHGDI1, SRGN, PTPRC, LAPTM5, LCP1, IL2RG, CD53, SAMS1, RGS1, and CD2) are shown in heatmaps (Fig. 5A).



**Figure 3.** The identification of mitochondrial autophagy activity subgroups and the transcriptional trajectory analysis of the cells in Fig. 3 reveal the transcriptional patterns of the NK cells. (A) The area under the curve (AUC) for mitochondrial autophagy-related genes was 0.14. (B) The activity of MRGs appears to differ significantly between NK cells and other cell types. (C) Pseudotime color gradient transitions from dark blue to light blue. (D) Pseudotime trajectory of OSCC cell distribution displayed based on sample patient sources. (E) The pseudotime trajectory is divided into three different states using Monocle2. (F) Stacked bar charts displaying the distribution of patients in different states. (G) The heatmap shows DEGs in different branches (cell fate). The GO pathways significantly enriched in different gene clusters are shown on the left of the heatmap.

By comparing OSCC samples with normal samples, 2214 genes were found to be differentially expressed. The differences in the expression of these genes between the two samples were statistically significant ( $|\text{Log}_2\text{-fold change}| > 1, \text{adjusted } p\text{-value} < 0.05$ ; Table S4). The top 10 upregulated genes (PTHLH, MMP9, LAMC2, INHBA, MMP12, MMP1, CA9, COL1A1, MMP13, and MMP11) and the top 10 downregulated genes (CRISP3, AQP5, STATH, PLA2G2A, LTF, PIP, ZG16B, SMR3B, MUC7, and BPIFB2) in the OSCC samples are displayed in heatmaps (Fig. 5B). Intersection of the two sets of DEGs resulted in 227 hub genes (Fig. 5C, Table S5).

To explore the biological functions associated with the marked genes, our group conducted enrichment analysis of the GO entries (Table S6) and KEGG pathways (Table S7). GO results revealed that the genes were concentrated in pathways related to skin development (GO:0,043,588), epidermal development (GO:0,008,544),



**Figure 4.** Communication mode of OSCC cells. **(A)** The network graph illustrates the frequency of interactions between different OSCC cell types. **(B)** The network graph demonstrates the strength of interactions among OSCC cell types. **(C)** The heatmap depicts the signaling pathways associated with OSCC cell types. **(D)** The heatmap depicts the signaling pathways associated with OSCC cell types.

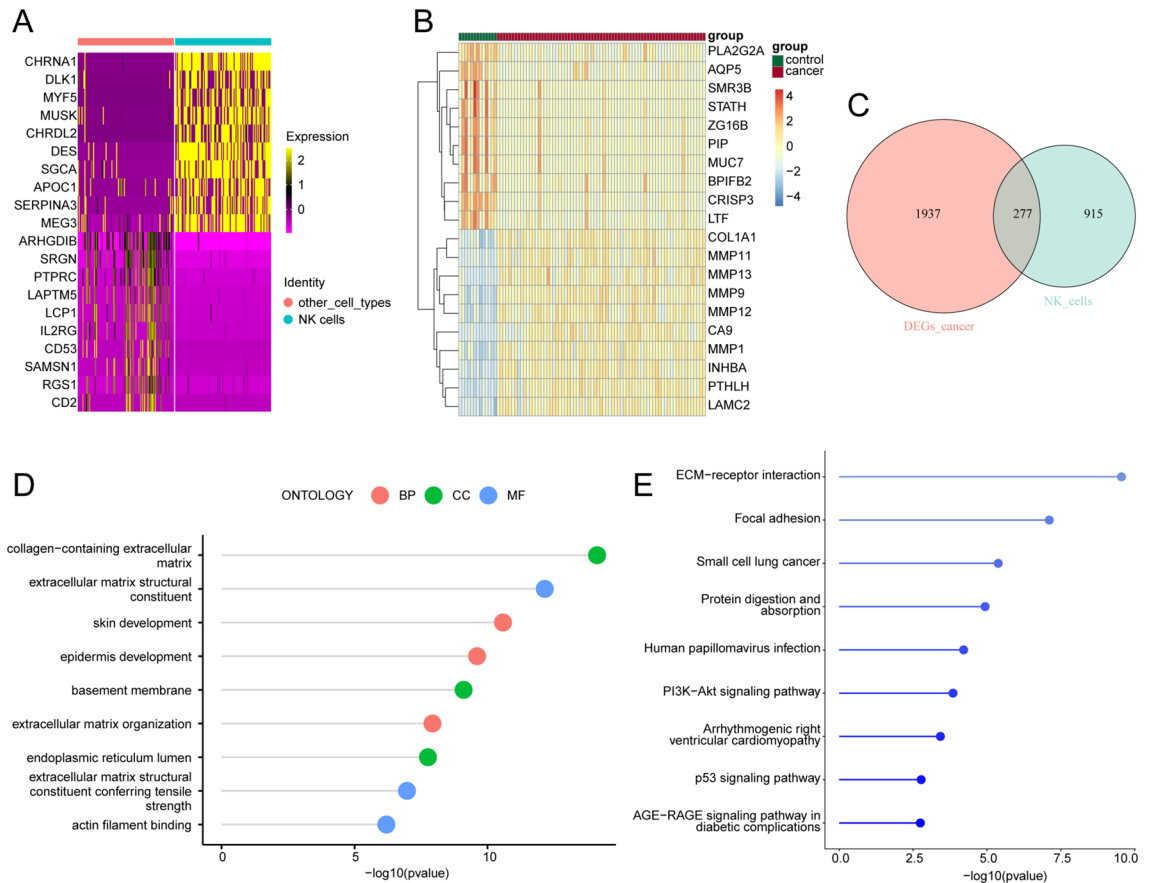
and extracellular matrix organization (GO:0,030,198) (BP) (Fig. 5D). The enriched KEGG pathways included ECM-receiver interaction (hsa04512), focal adhesion (hsa04510), and small-cell lung cancer (hsa05222) (Fig. 5E).

### Development and evaluation of prognostic models

Through single-factor Cox analysis, we identified key genes related to prognosis, ultimately identifying 28 genes linked to OSCC prognosis ( $p < 0.05$ ; Table S8). Next, The OSCC samples ( $n = 442$ ) were classified randomly into two groups: a training group ( $n = 301$ ) and a validation group ( $n = 141$ ). LASSO regression analysis was also conducted on the training dataset to eliminate redundant genes and establish a random number seed of four. Twelve genes associated with the prognosis of patients with OSCC were identified (Table S9). Based on the median risk value, the samples were divided into high-risk and low-risk groups to assess the stability of the 12-gene model. Figure 6A displays the expression levels of the 12 genes, survival time, and risk score distribution in the training dataset. The results revealed a more dense distribution of deaths and higher expression levels of the COBL and AREG genes in the high-risk group. This indicates that patients in the high-risk group may have more severe disease than those in the low-risk group. Figure 6B and C display the KM survival curves for different patient populations in both the training and validation datasets. The results showed that patients in the high-risk group had a significantly worse prognosis than did those in the low-risk group, further validating the effectiveness of our 12-gene model in predicting patient prognosis. To further evaluate the predictive performance of our model, receiver operating characteristic (ROC) curves were generated to assess its ability to predict patient prognosis. Figure 6D and E display the ROC curves for both the training and validation datasets. In the training cohort, the area under the curve (AUC) values at 1, 3, and 5 years were 0.674, 0.669 and 0.665, respectively (Fig. 6D). In the validation dataset, the area under the curve (AUC) values at 1, 3, and 5 years were 0.725, 0.715, and 0.733, respectively (Fig. 6E). These data demonstrated that our model has good predictive performance and can be used to distinguish between high-risk and low-risk patients effectively.

### GSEA and GSVA

We used pathway information from the MsigDB database as a reference for GSEA of DEGs. We selected the most significant pathway based on the normalized enrichment score and used the FDR-adjusted  $P$  value to control for errors. The results showed that three pathways were highly enriched in the high-risk group: HYPERTROPHIC\_CARDIOMYOPATHY\_HCM (NES = 2.1171, adjusted  $P = 0.0163$ , FDR = 0.0177; Fig. 7A), FOCAL\_ADHESION (NES = 2.0956, adjusted  $P = 0.0163$ , FDR = 0.0117; Fig. 7B), and DILATED\_CARDIOMYOPATHY (NES = 2.0596, adjusted  $P = 0.0163$ , FDR = 0.0117; Fig. 7C). In contrast, three pathways were significantly enriched in the low-risk group: DRUG\_METABOLISM\_CYTOCHROME\_P450 (NES = -1.8336, adjusted  $P = 0.0181$ , FDR = 0.013; Fig. 7D), GLUTATHIONE\_METABOLISM (NES = -1.8679, adjusted  $P = 0.0181$ , FDR = 0.013; Fig. 7E), and METABOLISM\_OF\_XENOBIOTICS\_BY\_CYTOCHROME\_P450 (NES = -1.9868, adjusted  $P = 0.0181$ ,



**Figure 5.** Enrichment analysis of DEGs in OSCC. **(A)** The heatmap shows the genes with significant differential expression in NK cells in OSCC. **(B)** The heatmap shows that 20 DEGs were significantly differentially expressed in OSCC. **(C)** Venn plot displaying key genes. **(D)** GO analysis of the key genes. **(E)** KEGG analysis of the key genes.

FDR = 0.013; Fig. 7F). Additionally, pathway activity heatmaps were generated using the 5 pathways associated with the genes exhibiting the greatest difference between the high-risk subgroup and low-risk subgroup (Fig. 7G, Supplementary Table S11).

### Nomogram

Univariate (Fig. 8A) and multivariate (Fig. 8B) Cox regression analyses were conducted to evaluate the clinical characteristics of patients. Both analytical approaches consistently indicate that the risk score serves as an independent prognostic risk factor for patients. Furthermore, a multivariate Cox regression analysis was used to devise a Nomogram, and the findings reveal that the risk score possesses substantial predictive value in forecasting clinical outcomes (Fig. 8C). The AUC values (1-, 3- and 5 years) are 0.730, 0.706, and 0.670, respectively (Fig. 8D), indicating its reliability and effectiveness in prognostic prediction.

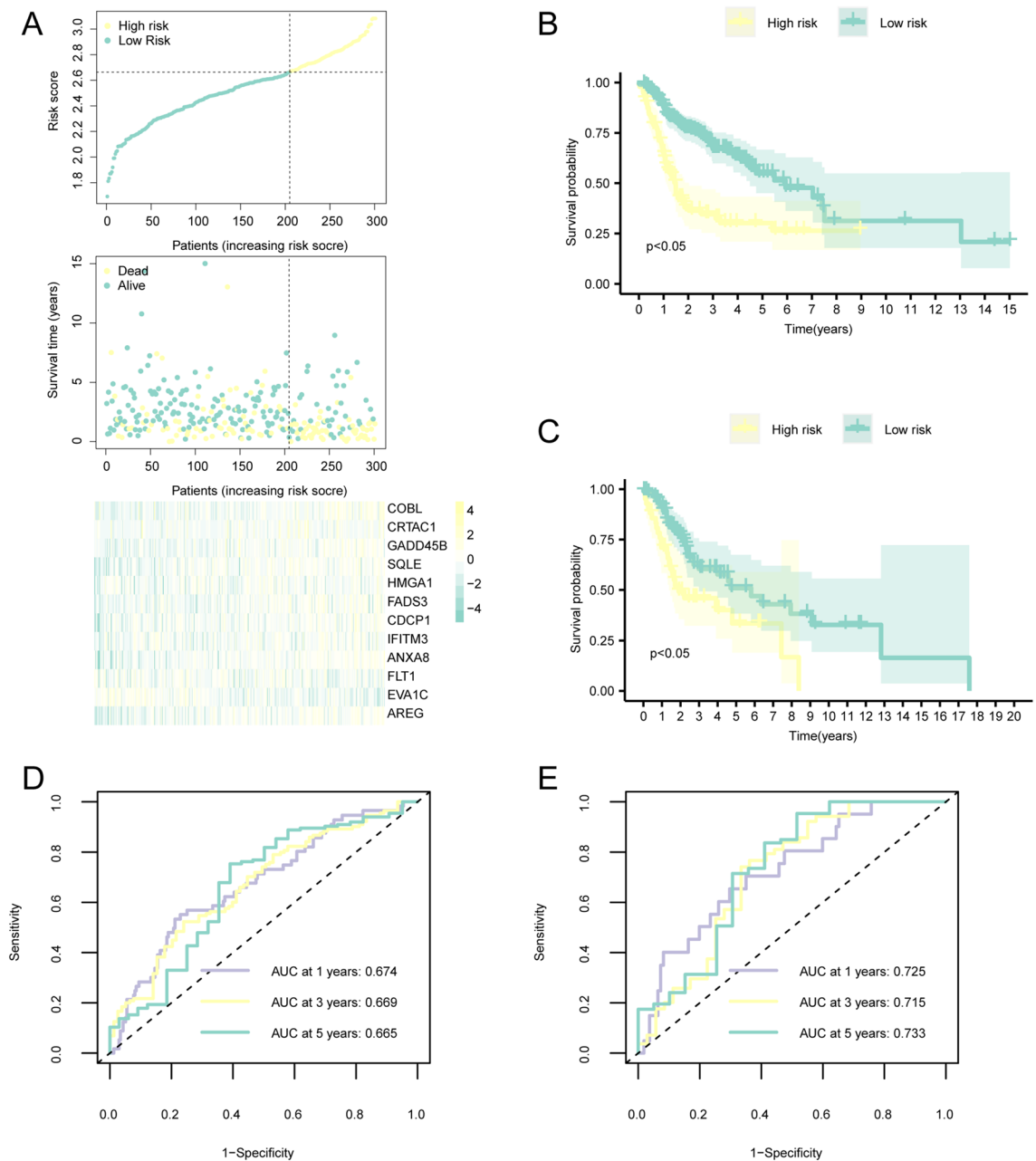
### Immune infiltration analysis

Table S12 shows the infiltration degree of 28 immune cell subsets. In addition, we also displayed the proportions of 28 immune cell subsets through a histogram, revealing the heterogeneity of immune cells in oral cancer. Figure S2 shows that most immune cells exhibited a positive correlation, while CD56bright NK cells and myeloid-derived suppressor cells exhibited a negative correlation. We also found significant differences in immune cells, including activated B cells, immature B cells, central memory CD8 + T cells, activated CD8 + T cells, gamma delta T cells, T follicular helper cells, eosinophils, and neutrophils ( $P < 0.05$ ), between the high-risk and low-risk groups. To further explore the relationships between specific genes and their corresponding immune cells, we analyzed the significant correlation between each central gene and its corresponding immune cell. AREG gene is negatively correlated with activated B cells ( $R = -0.3615$ ,  $P < 0.001$ ), while FLT1 gene is positively correlated with central memory CD8 + T cells ( $R = 0.4756$ ,  $P < 0.001$ ) (Fig. S2).

### TMB and drug sensitivity analysis

TMB result shows the top 20 genes with the highest mutation frequency in both groups are the same. The gene with the highest mutation frequency is TP53, followed by TTN (Fig. S3). Patients with low-risk scores may be



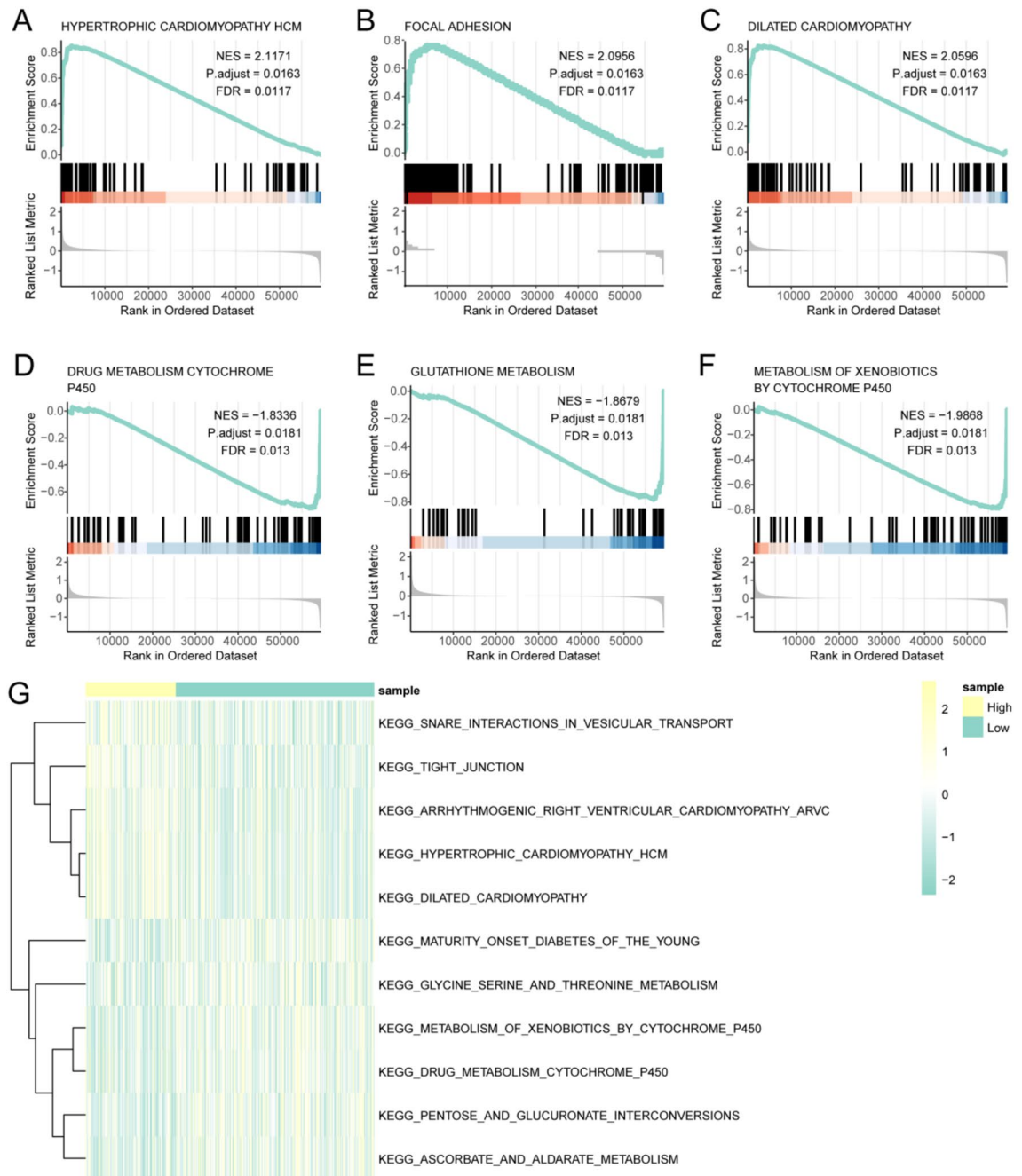


**Figure 6.** A prognostic model constructed for the OSCC dataset. **(A)** The distribution of risk scores, heatmap of key genes and survival profiles in the OSCC dataset. **(B, C)** Survival curves of the high- and low-risk patients from the training dataset **(B)** and validation dataset **(C)** Yellow indicates the high-risk group, while green indicates the low-risk group. **(D, E)** Time-dependent ROC curves for 1-, 3-, and 5-year survival in the training cohort **(D)** and validation cohort **(E)**.

more sensitive to the reactions of Dihydronotenone\_1827, GNE-317\_1926, and Sabutoclax\_1849, indicating that chemotherapy is a promising option for the low-risk scoring group (Fig. S3, Table S13).

## Discussion

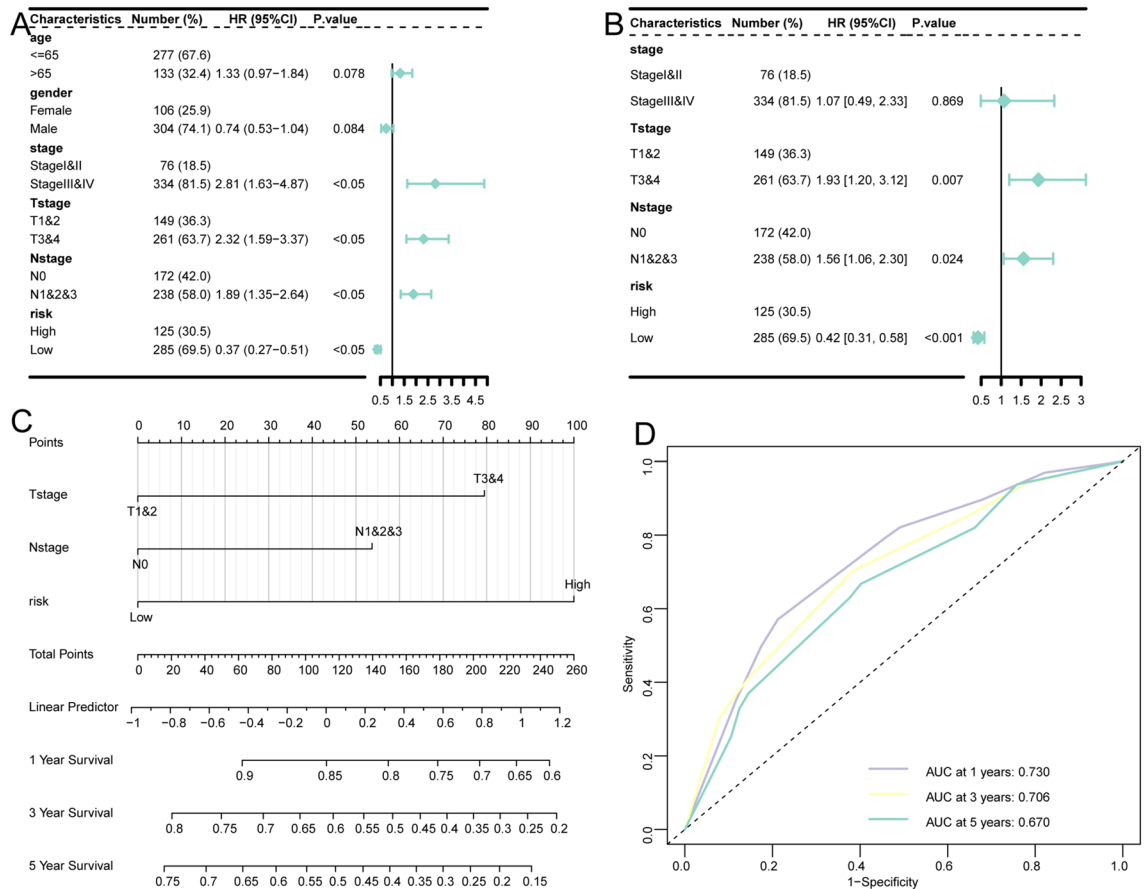
In this study, a total of 4541 cells were obtained from the single-cell transcriptome, and all the cells were classified into 19 clusters. According to the characteristics of each cluster, nine cell types were identified using cell-specific biomarker annotations. The expression patterns of mitochondrial autophagy-related genes were used to identify active cell subpopulations at the single-cell scale. We found that the activity of MRGs in different cell types in oral cancer is heterogeneous, with NK cells showing significantly greater activity than other cells. NK cells are a crucial component of the human innate immune system. They are cytotoxic and secrete various immunoregulatory cytokines and chemokines<sup>34</sup>. Letai et al. reported that the mitochondrial apoptosis mechanism in tumor cells is important for efficient NK cell killing and that targeting mitochondrial apoptosis can enhance the tumor



**Figure 7.** GSEA and GSVA revealed significantly enriched pathways. GSEA revealed that HYPERTROPHIC\_CARDIOMYOPATHY\_HCM (A), FOCAL\_ADHESION (B), DILATED\_CARDIOMYOPATHY (C), DRUG\_METABOLISM\_CYTOCHROME\_P450 (D), GLUTATHIONE\_METABOLISM (E), and METABOLISM\_OF\_XENOBIOTICS\_BY\_CYTOCHROME\_P450 (F) were significantly enriched. (G) A pathway heatmap demonstrating GSVA enrichment in different risk groups.

inhibitory function of NK cells<sup>35</sup>. Research has also suggested that NK cells infiltrating liver cancer cells undergo mitochondrial rupture, resulting in loss of their own antitumor function<sup>36</sup>. Combining the conclusions of the above studies with those of our study, it can be concluded that mitochondrial function is correlated with the dysfunction of tumor-invading NK cells, leading to the loss of tumor immune monitoring mechanisms.

Using GO enrichment analysis, we characterized the biological roles of key genes implicated in mitochondrial autophagy, providing novel insights into the regulatory mechanisms of these genes. GO analysis revealed that key genes might play potential roles in skin development, epidermal development, extracellular matrix tissue, and other processes. The KEGG pathways included “ECM-receiver interaction”, “focus adhesion”, and “small-cell lung cancer”. Squamous cell carcinoma affects the skin and mucosa mainly. OSCC indicates that the site of cancer transformation is the oral cavity. Combining clinical and pathological features, OSCC usually does not metastasize in its early stages. As the diameter of the tumor increases, the tumor gradually enters the dermis



**Figure 8.** Risk score is an independent prognostic factor. The forest plot shows the results of univariate (A) and multivariate (B) Cox regression analysis on clinical features. (C) The column chart of the prediction model. The line segment represents the contribution of clinical factors to the outcome event, the total score represents the sum of corresponding individual scores for all variable values, and the bottom three lines represent the prognosis for 1-year, 3-year, and 5-year survival corresponding to each value point. (D) The time dependent ROC curves of the column chart model for 1 year, 3 years, and 5 years.

and extends to a greater area and depth, invading the lymph nodes. If the patient's condition progresses to an advanced stage and cancer cells start undergoing distant metastasis, other organs can be significantly affected, which can lead to failure and a very low probability of cure. The development and progression of OSCC from the surface to the inside are closely related to the enrichment results of the GOKEGG pathway, but the specific mechanisms of each pathway in OSCC still require further research.

We also analyzed the infiltration levels of 28 types of immune cells in both the high- and low-risk groups and found that the proportion of immune cells varied among individuals. Most immune cells were positively correlated with each other, whereas CD56bright natural killer cells and myeloid-derived suppressor cells were negatively correlated. We observed significant differences in immune cell expression between the high- and low-risk groups ( $P < 0.05$ ). Furthermore, significant correlations were observed between each hub gene and the corresponding immune cells. The AREG gene was strongly negatively associated with activated B cells ( $R = -0.3615$ ,  $p < 0.001$ ). The FLT1 gene was significantly associated with central memory CD8 T cells ( $R = 0.4756$ ,  $p < 0.001$ ).

AREG is a protein in the epithelial growth factor family that binds and activates the epithelial growth factor receptor EGFR, promotes the formation of homodimers or heterodimers with EGFR and triggers the phosphorylation of downstream signals within cells, thereby regulating cell proliferation and differentiation<sup>37</sup>. AREG was first identified in the MCF-7 human breast cancer cell line. It was primarily characterized as a bifunctional growth factor that can inhibit the proliferation of tumor cell lines but also promotes the growth of normal cells, such as fibroblasts and keratinocytes<sup>38</sup>. Multiple studies have demonstrated that AREG plays a crucial role in cell biology and tumor development. AREG is positively regulated in various tumors, including breast, liver, colon, lung, head and neck, prostate, stomach, bladder, and skin tumors<sup>39</sup>. Studies have shown that AREG can act as a tumor-promoting factor and play a significant role in the development of numerous tumors<sup>40</sup>. New research has also demonstrated that AREG expression is triggered by various mitochondrial stressors and that upregulation of AREG expression serves as a biomarker for mitochondrial damage in mouse models of liver injury.

FLT1, a member of the vascular endothelial growth factor receptor family, is a growth factor that stimulates endothelial cell proliferation and inhibits apoptosis, promoting angiogenesis<sup>41</sup>. Therefore, it actively regulates both normal and pathological angiogenesis. Its expression is observed in most tumor tissues and is significantly

associated with tumor growth, differentiation, recurrence, metastasis, and patient survival<sup>42,43</sup>. Aggarwal et al. reported an increase in FLT1 expression at the site of the tumor and in the peripheral circulation of patients with OSCC, indicating a correlation between FLT1 expression and the occurrence, development, and distant spread of oral cancer<sup>44</sup>. The risk model constructed in this research included 12 mitochondrial autophagy-related genes that potentially function and play roles in the tumor immune environment.

Mitochondria serve as the central hubs for metabolic signaling and are highly dynamic, undergoing continuous morphological changes through fusion and fission reactions, along with alterations in their intracellular distribution. The involvement of these proteins in diverse cellular processes and the growing evidence for the therapeutic significance of mitochondrial autophagy in various diseases have placed them at the forefront of recent research. In the current scenario of widespread resistance to chemotherapeutic agents in clinical tumors, mitochondrial autophagy assumes a crucial role in the antitumor response. A recent study reported a positive correlation between increased mitochondrial autophagy in OSCC stem cells and cisplatin resistance, suggesting that mitochondrial autophagy aids tumor cells in evading targeted destruction by chemical agents<sup>45</sup>. These findings raise the prospect of utilizing mitochondrial autophagy as a potential treatment for OSCC. Our analysis revealed that patients with low risk scores exhibit high sensitivity to chemotherapy, suggesting that chemotherapy represents a promising option for this subset of patients. The present challenge lies in our inadequate understanding of the molecular mechanisms underlying mitochondrial autophagy, which hinders our ability to distinguish its benefits from potential drawbacks in cancer therapy. The aforementioned conclusions are based solely on bioinformatics analysis and may offer novel insights into OSCC research; however, it is worth noting that there are some selection biases in the population datasets obtained from the GEO and TCGA cohorts, necessitating further validation through large-scale prospective studies.

## Conclusion

In summary, this study revealed that the activity of MRGs in different cell types in oral cancer is heterogeneous, and the activity of MRGs in NK cells is markedly greater than that in other immune cells, suggesting the potential for the development of new targeted therapies. The construction of a robust prognostic model also provides reference value for individualized prediction and clinical decision-making in patients with OSCC.

## Data availability

The datasets generated and/or analyzed during the current study are available in TCGA (<https://portal.gdc.cancer.gov/>) and GEO databases (<https://www.ncbi.nlm.nih.gov/geo/>). The single-cell dataset is GSE103322. This article and supplemental material included all the data generated during this study. For further inquiries, please contact the corresponding author.

Received: 30 April 2024; Accepted: 16 August 2024

Published online: 28 August 2024

## References

- Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
- Almangush, A., Pirinen, M., Youssef, O., Mäkitie, A. A. & Leivo, I. Risk stratification in oral squamous cell carcinoma using staging of the eighth American Joint Committee on Cancer: Systematic review and meta-analysis. *Head Neck* **42**, 3002–3017 (2020).
- Montero, P. H. & Patel, S. G. Cancer of the oral cavity. *Surg. Oncol. Clin. N. Am.* **24**, 491–508 (2015).
- Petersen, P. E. Oral cancer prevention and control – The approach of the World Health Organization. *Oral Oncol.* **45**, 454–460 (2009).
- Simula, L., Nazio, F. & Campello, S. The mitochondrial dynamics in cancer and immune-surveillance. *Semin. Cancer Biol.* **47**, 29–42 (2017).
- Qiu, Y.-H. et al. Mitochondria autophagy: a potential target for cancer therapy. *J. Drug Target.* **29**, 576–591 (2021).
- He, L. & Gu, K. Tanshinone IIA regulates colorectal cancer apoptosis via attenuation of Parkin-mediated mitophagy by suppressing AMPK/Skp2 pathways. *Mol. Med. Rep.* **18**(2), 1692–1703. <https://doi.org/10.3892/mmr.2018.9087> (2018).
- Chang, Y.-J., Chen, K.-W. & Chen, L. Mitochondrial ROS1 increases mitochondrial fission and respiration in oral squamous cancer carcinoma. *Cancers* **12**, 2845 (2020).
- Denisenko, T. V., Gorbunova, A. S. & Zhivotovsky, B. Mitochondrial Involvement in Migration, Invasion and Metastasis. *Front. Cell Dev. Biol.* **7**, 355 (2019).
- Xu, G. et al. Dissecting the human immune system with single cell RNA sequencing technology. *J. Leukocyte Biol.* **107**, 613–623 (2020).
- Liu, S. & Trapnell, C. Single-cell transcriptome sequencing: recent advances and remaining challenges. *F1000Research* **5**, 182 (2016).
- Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
- Colaprico, A. et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71–e71 (2016).
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
- Van De Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
- Onishi, M., Yamano, K., Sato, M., Matsuda, N. & Okamoto, K. Molecular mechanisms and physiological functions of mitophagy. *EMBO J.* **40**, e104705 (2021).
- Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
- Karmaus, P. W. F. et al. Metabolic heterogeneity underlies reciprocal fates of TH17 cell stemness and plasticity. *Nature* **565**, 101–105 (2019).
- Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 (2021).
- Fang, Z. et al. Single-cell transcriptomics of proliferative phase endometrium: Systems analysis of cell–cell communication network using cell chat. *Front. Cell Dev. Biol.* **10**, 919731 (2022).

21. The Gene Ontology Consortium. Gene Ontology Consortium: Going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).
22. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1), 27–30. <https://doi.org/10.1093/nar/28.1.27>. PMID:10592173;PMCID:PMC102409 (2000).
23. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: An R Package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).
24. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1 (2010).
25. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
26. Liberzon, A. *et al.* The molecular signatures database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
27. Wu, S. *et al.* Integrated machine learning and single-sample gene set enrichment analysis identifies a TGF-beta signaling pathway derived score in headneck squamous cell carcinoma. *J. Oncol.* **2022**, 1–12 (2022).
28. Ru, B. *et al.* TISIDB: An integrated repository portal for tumor-immune system interactions. *Bioinformatics* **35**, 4200–4202 (2019).
29. Ito, K. & Murphy, D. Application of ggplot2 to pharmacometric graphics. *CPT: Pharmacometrics Syst. Pharmacol.* **2**, 79 (2013).
30. Yang, W. *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–D961 (2012).
31. Maeser, D., Gruener, R. F. & Huang, R. S. oncoPredict: An R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief. Bioinform.* **22**, bbab260 (2021).
32. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).
33. Liu, Z. *et al.* TTN/OBSCN ‘Double-Hit’ predicts favourable prognosis, ‘immune-hot’ subtype and potentially better immunotherapeutic efficacy in colorectal cancer. *J. Cell. Mol. Med.* **25**, 3239–3251 (2021).
34. Freud, A. G., Mundy-Bosse, B. L., Yu, J. & Caligiuri, M. A. The broad spectrum of human natural killer cell diversity. *Immunity* **47**, 820–833 (2017).
35. Pan, R., Ryan, J., Pan, D., Wucherpennig, K. W. & Letai, A. Augmenting NK cell-based immunotherapy by targeting mitochondrial apoptosis. *Cell* **185**, 1521–1538.e18 (2022).
36. Zheng, X. *et al.* Mitochondrial fragmentation limits NK cell-based tumor immunosurveillance. *Nat. Immunol.* **20**, 1656–1667 (2019).
37. Berasain C, Avila M A. Gene Section. <http://AtlasGeneticsOncology.org>, 2010; 1.
38. Shoyab, M. *et al.* Amphiregulin: A bifunctional growth-modulating glycoprotein produced by the phorbol 12-myristate 13-acetate-treated human breast adenocarcinoma cell line MCF-7. *Proc. Natl. Acad. Sci.* **85**(17), 6528–6532 (1988).
39. Busser, B. *et al.* The multiple roles of amphiregulin in human cancer. *Biochim. et Biophys. Acta –Rev. Cancer* **1816**(2), 119–131 (2011).
40. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**(5), 646–674 (2011).
41. Melincovici, C. S. *et al.* Vascular endothelial growth factor (VEGF) - key factor in normal and pathological angiogenesis. *Rom. J. Morphol. Embryol.* **59**(2), 455–467 (2018).
42. Qian, B. Z. *et al.* FLT1 signaling in metastasis-associated macrophages activates an inflammatory signature that promotes breast cancer metastasis. *J. Exp. Med.* **212**(9), 1433–1448. <https://doi.org/10.1084/jem.20141555> (2015).
43. Masood, R. *et al.* Vascular endothelial growth factor (VEGF) is an autocrine growth factor for VEGF receptor-positive human tumors. *Blood* **98**(6), 1904–1913 (2001).
44. Aggarwal, S., Devaraja, K., Sharma, S. C. & Das, S. N. Expression of vascular endothelial growth factor (VEGF) in patients with oral squamous cell carcinoma and its clinical significance. *Clin. Chim. Acta.* **25**(436), 35–40. <https://doi.org/10.1016/j.cca.2014.04.027> (2014).
45. Naik, P. P. *et al.* Autophagy regulates cisplatin-induced stemness and chemoresistance via the upregulation of CD 44, ABCB 1 and ADAM 17 in oral squamous cell carcinoma. *Cell Prolif.* **51**, e12411 (2018).

### Author contributions

X.H. and M.L. designed the research. M.L. and Y.W. wrote the main manuscript text. W.H. and M.L. performed the data collection and bioinformatics analysis. S.H. and C.W. was responsible for statistical analysis and Drafting. S.B. and S.H. investigated and organize the literature. X.H. and L.Y. was responsible for the integrity of the entire study and manuscript review. All authors reviewed the manuscript.

### Funding

This work was supported by the National Natural Science Foundation of China (82360187), Guangxi Science and Technology Base and Talents Special Project (2021AC18031), Guangxi Medical and health suitable technology development and popularization applications project (S20210850) and Nanning Qingxiu District Science and Technology Plan (2021004).

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-70498-0>.

**Correspondence** and requests for materials should be addressed to X.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024