Functional & Integrative Genomics (2025) 25:25
https://doi.org/10.1007/s10142-024-01525-6

RESEARCH

Heterogeneity analysis and prognostic model construction of HPV negative oral squamous cell carcinoma T cells using ScRNA-seq and bulk-RNA analysis

Chunyan Li[1] · Zengbo Lv[1] · Chongxin Li[1] · Shixuan Yang[1] · Feineng Liu[1] · Tengfei Zhang[1] · Lin Wang[1] · Wen Zhang[1] · Ruoyu Deng[1] · Guoyu Xu[1] · Huan Luo[1] · Yinhong Zhao[1] · Jialing Lv[1] · Chao Zhang[1]

Received: 8 October 2024 / Revised: 8 December 2024 / Accepted: 31 December 2024

Abstract

Background T cells are involved in every stage of tumor development and significantly influence the tumor microenvironment (TME). Our objective was to assess T-cell marker gene expression profiles, develop a predictive risk model for human papilloma virus (HPV)-negative oral squamous cell carcinoma (OSCC) utilizing these genes, and examine the correlation between the risk score and the immunotherapy response.

Methods We acquired scRNA-seq data for HPV-negative OSCC from the GEO datasets. We performed cell−cell communication, trajectory, and pathway enrichment analyses of T-cell-associated genes. In addition, we constructed and validated a T-cell-associated gene prognostic model for HPV-negative OSCC patients using TCGA and GEO data and assessed the immune infiltration status of HPV-negative OSCC patients .qRT-PCR was used to detect the expression level of prognosis-related genes in different risk groups.

Results ScRNA-seq was conducted on 28,000 cells derived from 14 HPV-negative OSCC samples and 6 normal samples. We identified 4,635 T cells from these cells and identified 774 differentially expressed genes(DEGs) associated with T cells across five distinct T-cell subtypes. Through the integration of bulk-RNAseq data, we established a prognostic model based on DEGs related to T cells. By separating patients into high-risk and low-risk groups according to these prognostic related genes, we can accurately predict their survival rates and the immune infiltration status of the TME.qRT-PCR results showed that compared with the patients of low risk group, the expression of PMEPA1, SH2D2A, SMS and PRDX4 were significantly up-regulated in high risk group.

Conclusion This study provides a resource for understanding the heterogeneity of T cells in HPV-negative OSCC patients and associated prognostic risk models. It provides new insights for predicting survival and level of immune infiltration in patients with HPV-negative OSCC.

Keywords HPV-negative oral squamous cell carcinoma · ScRNA-seq · Bulk-RNA · T cells · Prognostic model

Introduction

Oral squamous cell carcinoma (OSCC) is the predominant form of head and neck cancer and accounts for more than 90% of all oral malignancies(Ng et al. 2016; Liu et al. 2023; Chai et al. 2020).OSCC arises from the alveolar ridge, buccal mucosa, floor of the mouth, palate, tongue, and other oral regions. In 2018, there were over 350,000 new cases and approximately 170,000 related fatalities. Globally, the prevalence of oral squamous cell carcinoma (OSCC) is highest in Asia(Sung et al. 2021).OSCC development is typically linked to lifestyle-related risk factors such as tobacco use, excessive alcohol intake, and human papilloma virus

✉ Jialing Lv
lvjialinglvjialing@126.com

✉ Chao Zhang
chesanjin@163.com

[1] Department of Oncology, the First People's Hospital of Qujing City/the Qujing Affiliated Hospital of Kunming Medical University, 1 Yuanlin Road, Qujing, Yunnan, China

Published online: 23 January 2025

(HPV) infection(Johnson et al. 2020). The use of betel nuts is frequently linked to diseases in regions endemic to OSCC, particularly in South Asia and Southeast Asia(Guha et al. 2014).While HPV is a common risk factor for cancer in other head and neck regions, such as the oropharynx, HPV is responsible for only 2–5% of OSCCs (Melo et al. 2021), and the implications of HPV infection in OSCC remain unclear. HPV-negative OSCC is caused predominantly by tobacco and alcohol consumption and constitutes the majority of OSCC cases. Despite advancements in the understanding of OSCC biology during recent decades, the prognosis has largely remained unchanged; the 5-year survival rate for HPV-negative OSCC patients is less than 50%(Sagheer et al. 2021). The therapeutic approaches for OSCC primarily include surgery, chemotherapy, radiotherapy, and targeted therapy; nevertheless, their efficacy is limited(Caudell et al. 2022).Consequently, investigating novel therapeutic targets is imperative for improving OSCC prognosis.

New prognostic biomarkers are needed to help stratify patients and to predict treatment response. Many bulk RNA-seq studies have been performed to identify OSCC prognostic markers and improved our understanding of carcinogenesis and progression in recent decades. Chen et al. identified prognosis-related immune genes of OSCC via bioinformatics analysis(Chen et al. 2022a, b). Wang et al. developed an RNA binding protein model to predict OSCC patient overall survival (OS)(Wang et al. 2023). The model in the above study had good predictive power, but bulk RNA-seq reveals the "average" expression levels for all cells in a sample, this approach cannot identify molecular-level differences between cells and cannot be used to assess the relationships among cell interactions, immune cell types, and cell-specific markers. Previous research has suggested that gene expression profiles based on immune cell molecular features derived from single-cell RNA sequencing (scRNA-seq) data may be used to predict cancer patient survival outcomes and immunotherapy response(Shi et al. 2022a, b; Kim et al. 2024).

Single-cell technology (SCT) is a revolutionary sequencing method for studying cell types, heterogeneity and tumour cell function in the tumour microenvironment (TME) (Liu and Wu 2023; Kamperman et al. 2018).Single-cell RNA sequencing (scRNA-seq) can be used to measure gene expression at the single-cell level, revealing heterogeneity of gene expression in single cells or homologous cell types and helping us to understand the role of T-cells, which are critical for tumour immunotherapy. To identify T-cell marker genes and assess tumour infiltration of T cells, we first analysed the TME of HPV-negative OSCC samples by scRNA-seq. Next, we used bulk-RNAseq data to create a T-cell marker gene profile for predicting the prognosis of HPV-negative OSCC. The predictive value of the T-cell marker gene profile was confirmed in two independent cohorts using the Gene Expression Omnibus (GEO) database. The study methodology is shown in Fig. 1.

Materials and methods

Data source and preparation

Single cell RNA-sequencing data of HPV-negative OSCC from GSE181919(https://www.ncbi.nlm.nih.gov/geo/) and were used to determine the T cell marker genes. In this study, we collected 14 primary HPV-negative OSCC tumor samples and 6 normal samples. The selection criteria were as follows: (1) every sample must have at least 500 cells; (2) each cell must express more than 250 genes; (3)each gene must be expressed in at least 3 cells; (4)each cell must have less than 20% mitochondrial RNA (Butler et al. 2018); We filtered and collected 28,000 cells for subsequent analysis using Seurat package of R.

The TCGA public database provides bulk-RNA seq data of HPV-negative OSCC and clinical data and were downloaded from the UCSC Xena (https://xenabrowser.net/) for identifying survival-related genes and constructing prognostic model.The dataset has 95 samples, which include 2 normal and 93 tumor samples. For DEGs identification, the limma package was used to determine the filtering parameters. DEGs were selected according to a threshold of $|\log 2\text{FoldChange}| > 1.5$ and adjusted P value < 0.05 for further analysis(Chen et al. 2022a, b).GEO provided the GSE41613 and GSE85446 datasets and clinical data. GSE85446 has 66 HPV-positive OSCC tumor samples, whereas GSE41613 has 97 tumor samples(Yang et al. 2024; Chi et al. 2022).

Integrating data and reducing dimensionality

We used Seurat's "NormalizeData" function to standardize the expression values after data filtration. Following normalization of the expression of every gene by the overall expression of genes in each cell, multiplying by 10,000, and adding 1, natural logarithm transformation was applied to circumvent zero values. A total of 2,000 highly variable genes were identified via "FindVariableFeatures". The "ScaleData" tool centered the genes. Data integration was completed with 2,000 "RunHarmony"anchor points. This "anchor point" integration method easily translates identical cell types across datasets into a few anchor points, decreasing batch effects and simplifying data integration.

Data clustering with reduced dimensions

High-dimensional data visualization is complicated by each gene in the samples. Thus, dimensionality reduction algorithms are needed to accurately describe data structures with

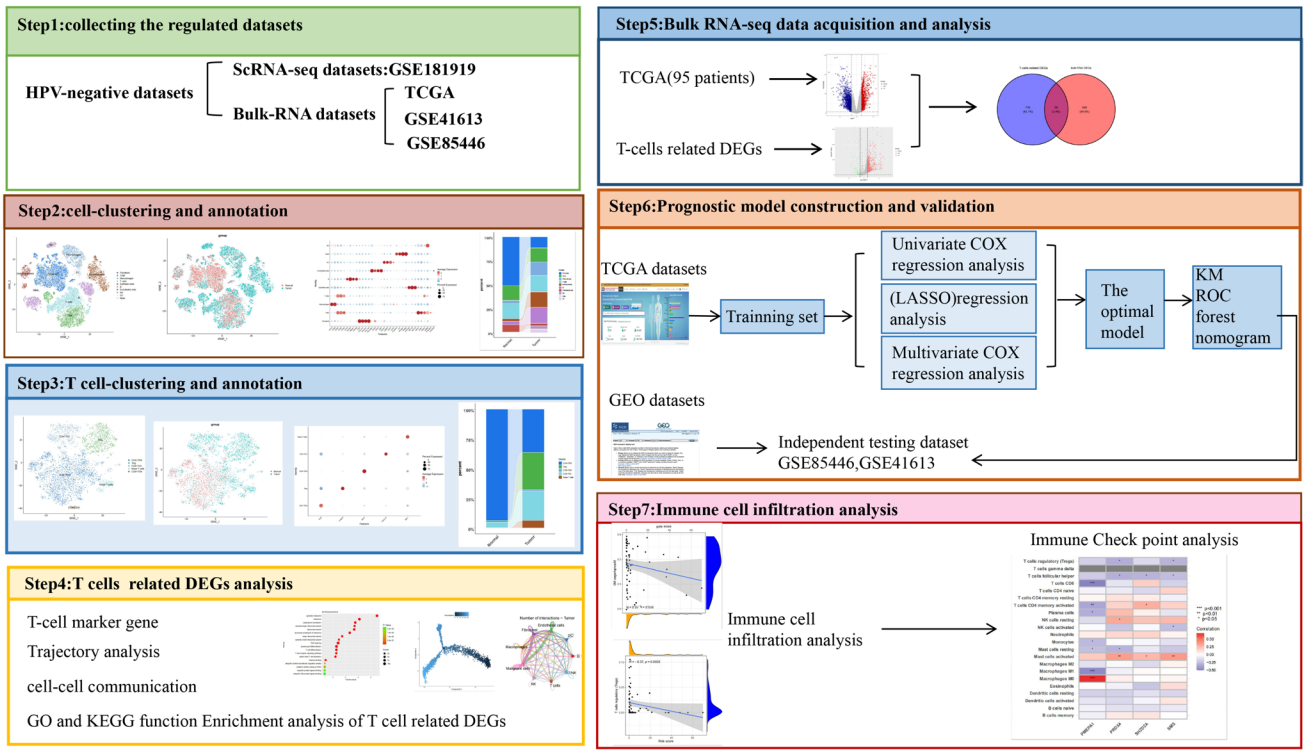


Fig. 1 The overall design process of this study

fewer dimensions(Stuart 2019). The "RunPCA" function of the Seurat package was used to reduces the dimensionality of the detected highly variable genes. We subsequently clustered the dimensionally reduced data via Seurat's "FindNeighbors" and "FindClusters" algorithms. The t-distributed stochastic neighbor embedding (t-SNE) method helps present clustering results by combining the t-distribution and random nearest neighbors(Zhou 2020). Clustering visualization was improved by using the Seurat package's "RunTSNE" function with dims = 1:20 and resolution = 0.7.

Cell-clustering and annotation

We annotated our data for cell type classification using the Cell Cytotaxonomy website. The differentially expressed genes (DEGs) of each cluster were calculated using the Wilcoxon Mann Whitney test using the "findallmarkers" function in the "Seurat" package. To identify marker genes for each cluster, we used a cutoff threshold with an adjusted p-value < 0.05 and |log2 (fold change)| >0.25, For cluster annotation.

Identification of marker genes and exploration of functional enrichment

The "FindMarkers" function of Seurat R package was used to find T cell marker genes in single-cell data using a |log2FoldChange| > 1. Later, the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) were performed in R used the "clusterProfiler" package to find the biological processes and pathways that were most strongly linked to DEGs related to T cells(Wu et al. 2021). GO annotations were based on genome-wide annotation packages (org.Hs.eg.db) released by the Bioconductor project .The data were plotted with ggplot2.

Pseudotime cellular trajectory analysis

Cells are constantly undergoing dynamic changes, transforming from one cell type to another.which alters gene expression and function(Haghverdi et al. 2016). Pseudotime analysis arranges cells along a biological trajectory and divides them into differentiation states on the basis of gene expression profiles. We performed pseudotime analysis on T-cell subtypes via monocle2(Ionkina et al. 2021). The"plot_cell_trajectory"function sorted cells in pseudotime to study T-cell subtype differentiation trajectories and related genes. The "plot_pseudotime_heatmap" function displayed the results.

Cell-cell communication analysis

Intercellular communication is critical for tumor cell biological activity. We used 'CellChat'to construct a cellular

communication network to study the interaction relationships between different cell types. 'CellChat'is an analytical instrument employed to assess the types and intensities of receptor ligands expressed among cell subpopulations in single-cell data(Fang et al. 2022; Jin et al. 2021).

Bulk RNA-seq data acquisition and analysis

TCGA provided HPV-negative OSCC bulk RNA-seq data and clinicopathological characteristics. A total of 95 HPV-negative OSCC bulk RNA-seq samples and clinical data were collected(Foy et al. 2017). Gene expression data were normalized using the variance stable transformation method of the "limma" R package. Gene expression annotation information was downloaded from Ensembl website.We used the "limma" R software package to analyze the genetic differences between HPV negative OSCC tumor tissues and normal tissues to identify DEGs. In the HPV negative OSCC dataset of TCGA, DEGs between tumor tissues and normal tissues were identified, and genes intersecting with T cell related DEGs were analyzed subsequently.GSE41613 and GSE85446 as independent external cohorts were download from GEO database.GSE85446 data were obtained via the GPL6480 Agilent-014850 Whole Human Genome Microarray 4×44 K G4112F, which includes 66 HPV-negative OSCC samples. GSE41613 data were obtained via the GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array.

Prognostic model construction and validation

The first step was univariate Cox regression analysis to identify prognostic DEGs(Dijk et al. 2008). The regression analysis used the least absolute shrinkage and selection operator (LASSO) to choose parameters with a P value < 0.05(Zhang et al. 2023). "glmnet" package of R was used to detect highly correlated genes, reduce gene count, and reduce model overfitting in the final risk model using LASSO regression analysis. The following formula was used to build the prognostic model from LASSO regression genes selected for multivariate Cox regression analysis, risk scores were calculated as follows:

Risk score = (0.7968×PMEPA1 expression) +(1.54998×PRDX4 expression)+(0.8579×SH2D2A expression)+(1.125×SMS expression). All patients were separated into high-risk and low-risk groups according to the median riskscore. The "survminer" and "ggrisk"R packages were used to construct survival curves and risk plots to represent patient survival status. We also used the "timeROC" R package to evaluate the ability of the riskscore to predict the survival outcomes of HPV-negative OSCC at 1, 3, and 5 years. Most plots in were drawn using ggplot2.The GSE41613 and GSE85446 datasets were independent external datasets used to verify the prognostic model.

Risk score and clinical correlation analysis

We subsequently assessed the correlation between the risk score and TCGA patients' clinicopathological features. To determine whether the riskscore is an Independent prognostic factors for HPV-positive OSCC patients, we used the "survival" R package for Cox regression. The "forestplot" R package was used to generate univariate and multivariate Cox regression forest plots. Nomograms were constructed for TCGA HPVnegative OSCC patients. By utilizing the prognosis-related gene signature riskscore and additional clinical factors, both univariate and multivariate Cox regression analyses, factors affecting OS were identified. Hazard ratios (HRs) and 95% confidence intervals (CIs) were determined. Nomograms were generated according to the multivariate Cox regression results.

Immune cell infiltration analysis

The traditional linear support vector regression method CIBERSORT deconvolutes immune cell expression matrices to measure marker gene expression in invading immune cells.By integrating marker gene expression data from 22 immune cell types and transcriptomic data from HPV-negative TCGA-OSCC patients, we quantitatively assessed tumour-infiltrating lymphocytes using CIBERSORT. This study also assessed immune cell infiltration in different risk groups using CIBERSORT.

12 qRT-PCR

In order to verify the expression levels of these four prognostic-related genes in different risk groups of HPV-negative OSCC, we collected tumor tissues from patients with HPV negative OSCC who underwent OSCC resection at our hospital from January 2022 to November 2024,and divided the samples into two groups (III-IV vs. I-II) according to pathological stages, and 10 samples in each group were examined for the relative expression levels of the target genes by qRT-PCR, This study was approved by the Internal Review Committee of the first people's hospital of Qujing. Total RNA was extracted from OSCC tissues using the RNAQuick Purification kit (EScience, Shanghai, China). HiScript R II Q RT SuperMix (Vazyme) was used to synthesize complementary DNA (cDNA). The ChamQTM SYBR R qPCR Master Mix kit (Vazyme) was used to detect mRNA levels of targeted genes. The human gene primers used as follows (5'−3'):

GAPDH F-primer: CTGGGCTACACTGAGCACC.
GAPDH R-primer: AAGTGGTCGTTGAGGGCAATG.
PMEPA1 F-primer: TGTCAGGCAACGGAATCCC.
PMEPA1 R-primer: CAGGTACGGATAGGTGGGC.

SH2D2A F-primer: GACTTTCCCTGAGGACCGAAG.
SH2D2A R-primer: GCTTGCCCCTGTTTGATGATTG.
SMS F-primer: TAGTGGGGATGTTAATTTGGCAG.
SMS R-primer: CCACACGTTTTTCGCATGTATTT.
PRDX4 F-primer: GCAAAGCGAAGATTTCCAAGC.
PRDX4 R-primer: CGCCAAAAGCGATAATTTCAGTT.

GAPDH was used as an internal reference gene to calibrate the relative expression level of the target gene. The primers were synthesised by Sangon Biotech (Shanghai, China).We used $2^{-\Delta\Delta CT}$ to assess the relative expression levels of the target genes.

Statistical analysis

R (4.3.2) was used to analyze the data. Statistical significance was determined via Student's t test. Continuous variable relationships were tested via the Wilcoxon rank-sum test. Survival differences between the high-risk and low-risk groups were assessed via Kaplan−Meier analysis. Differences were considered statistically significant at $P < 0.05$.

Results

HPV-negative OSCC ScRNA-seq analysis

This study included a total of 20 single-cell samples of OSCC. Following the elimination of substandard cells, normalization, integration, and principal component analysis (PCA), 28,000 cells were categorized into 24 clusters (Fig. 2A). By annotating these subclusters for analysis, we identified ten cell types: fibroblasts (clusters 0, 8, 10, 15, 16, 19, and 21), T/NK cells (clusters 1 and 12), macrophages (clusters 2 and 23), T cells (clusters 3, 5, and 18), epithelial cells (clusters 4, 11, and 20), B cells (clusters 6 and 7), endothelial cells (clusters 9 and 17), NK cells (cluster 13), dendritic cells (cluster 14), and mast cells (cluster 22) (Fig. 2B). Figure 2C shows the t-SNE clustering diagram representing tumor and normal tissues. Figure 2D-F shows the significant marker genes associated with each cell type (Supplementary Table S1). Figure 2G illustrates the distribution of cell types in normal and tumor samples within the scRNA-seq dataset.

Analysis of single-cell atlas about T cell subsets in HPV-negative OSCC

T-cell mediated immune responses are significant in the TME of OSCC(Chao et al. 2021). The HPV-negative OSCC scRNA-seq atlas revealed a significant abundance of T cells across all samples. Consequently, our study isolated 4635 T cells from the all cells for further clustering and annotation. Figure 3A illustrates the T cells categorized into seven groups following an additional clustering process. Based on the marker genes documented in the literature, the resulting T cell subpopulations were classed and identified. The particulars were as follows: (1) effector memory CD4 + T cells(CD4 + TEM) exhibit high expression of IL7R, ANXA1, FOS, and JUN (clusters 0,3,5); (2) regulatory T cells (Treg) demonstrate elevated expression of FOXP3, IL2RA, and BATF (clusters 1); (3) exhausted CD4 + T cell(CD4 + TEX) cells show significant expression of PDCD1, CXCL13, and CD200 (clusters 2); (4) naive T cells are characterized by high expression of CRR7, LEF1, SELL, and TCF7 (clusters 4); (5) effector memory CD8 + T cells(CD8 + TEM) express high levels of FOXP3, IL2RA, and BATF (clusters 6). Figure 3B illustrates the annotation of T cells, comprising 2,652 CD4 + TEM cells, 942 regulatory T cells, 37 CD8 + TEM cells, 814 CD4 + TEX cells, and 190 naive T cells. Figure 3C illustrates the t-SNE clustering diagram of T lymphocytes present in tumor and normal tissues. Figure 3D-F displays the marker genes for each T cell subtype, including FOS, FOXP3, PRF1, CXCL13, and SELL, whereas Fig. 3D presents the top four differentially expressed genes (DEGs) in each T cell subtype (Supplementary Table S2). Figure 3G illustrated the distribution of T cell subtypes in normal and tumor samples within the scRNA-seq dataset.

Enrichment analysis of T cell-related DEGs pathways

In order to gain deeper understanding into the biological roles of marker genes in T cells, we carried out enrichment analyses using the GO and KEGG pathways(Supplementary Table S3-4). The results of these analyses showed that the majority of the marker genes are enriched in the T cell activation pathway. Figure 4A displays the volcano plot of the overall cell subpopulation intercellular differences analysis (red dots indicate genes specific to cell clusters relative to other cell clusters).Encompassing the T cell receptor signaling pathway, Th1 and Th2 cell differentiation, Th17 cell differentiation, PDL1 expression, and the PD-1 checkpoint pathway in cancer, as well as the NF-kappa B signaling pathway (Fig. 4B-C); GO enrichment analysis of the differentially expressed genes (DEGs) associated with T cells indicated a predominant enrichment in biological processes(BP) such as lymphocyte differentiation, T cell receptor signaling pathway, T cell differentiation, and alpha-beta T cell activation.The focus was mostly on the cytosolic ribosome, the cytosolic large ribosomal subunit, and the cytosolic small ribosomal subunit within cellular components (CC). Regarding molecular functions (MF), they were chiefly engaged in structural constituents of ribosomes, histone binding, ubiquitin-protein transferase regulatory activity, and binding to ubiquitin and ubiquitin-like protein ligases (Fig. 4D and

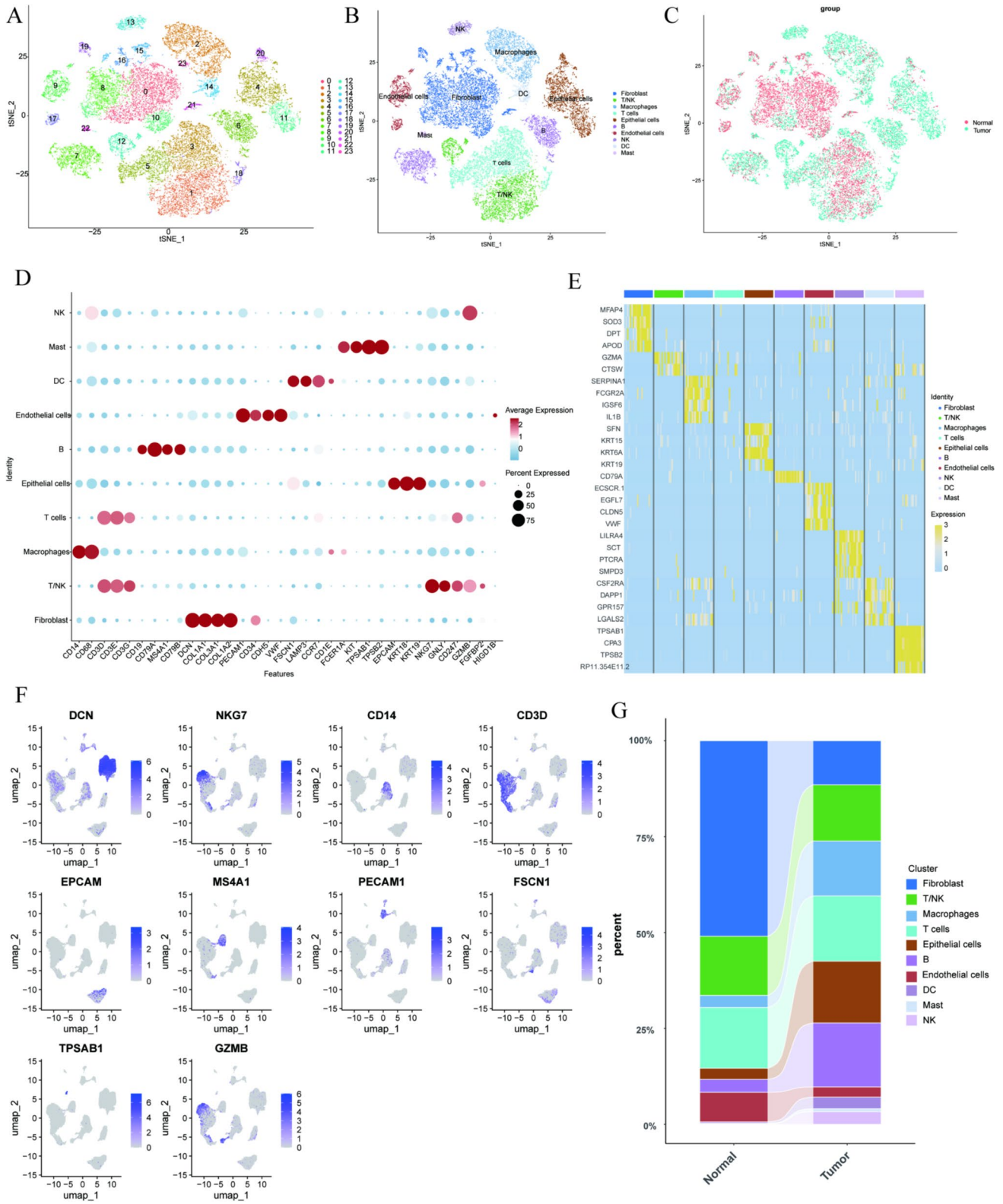


Fig. 2 HPV-negative OSCC ScRNA-seq analysis.**A** TSNE diagram depicts the clustering of single-cell samples into 24 clusters; **B** Identification of 10 cell types based on cell type marker genes; **C** Comparing TSNE clustering maps of normal and tumor tissues; **D** Bubble chart displays the expression of 10 different cell type related marker genes; **E** Heatmap displays the expression of differentially expressed genes among 10 cell types; **F** Umap diagram highlights the expression patterns of 10 cell type marker genes; **G** Cell type distribution of normal and tumor samples in ScRNA seq dataset;

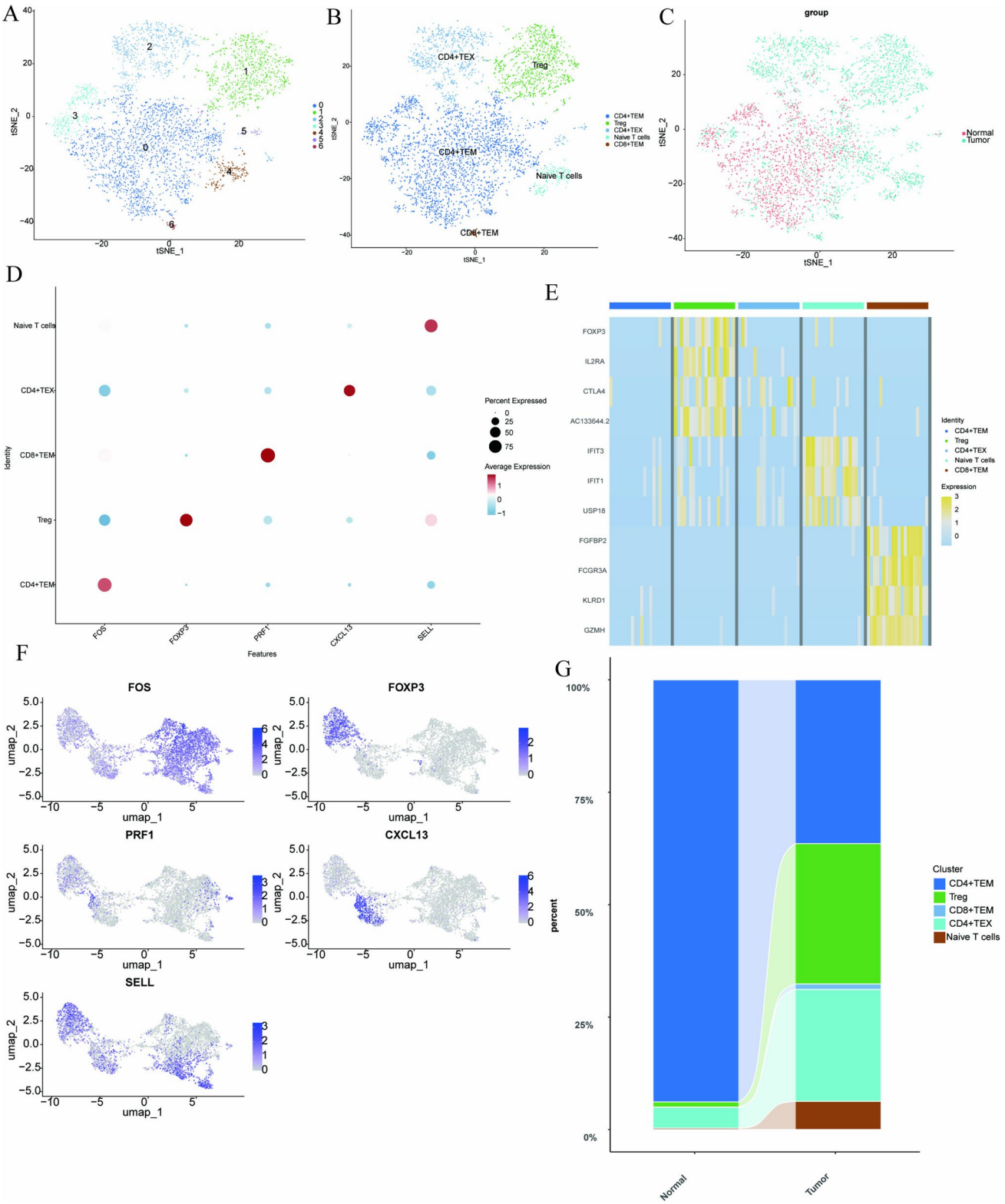
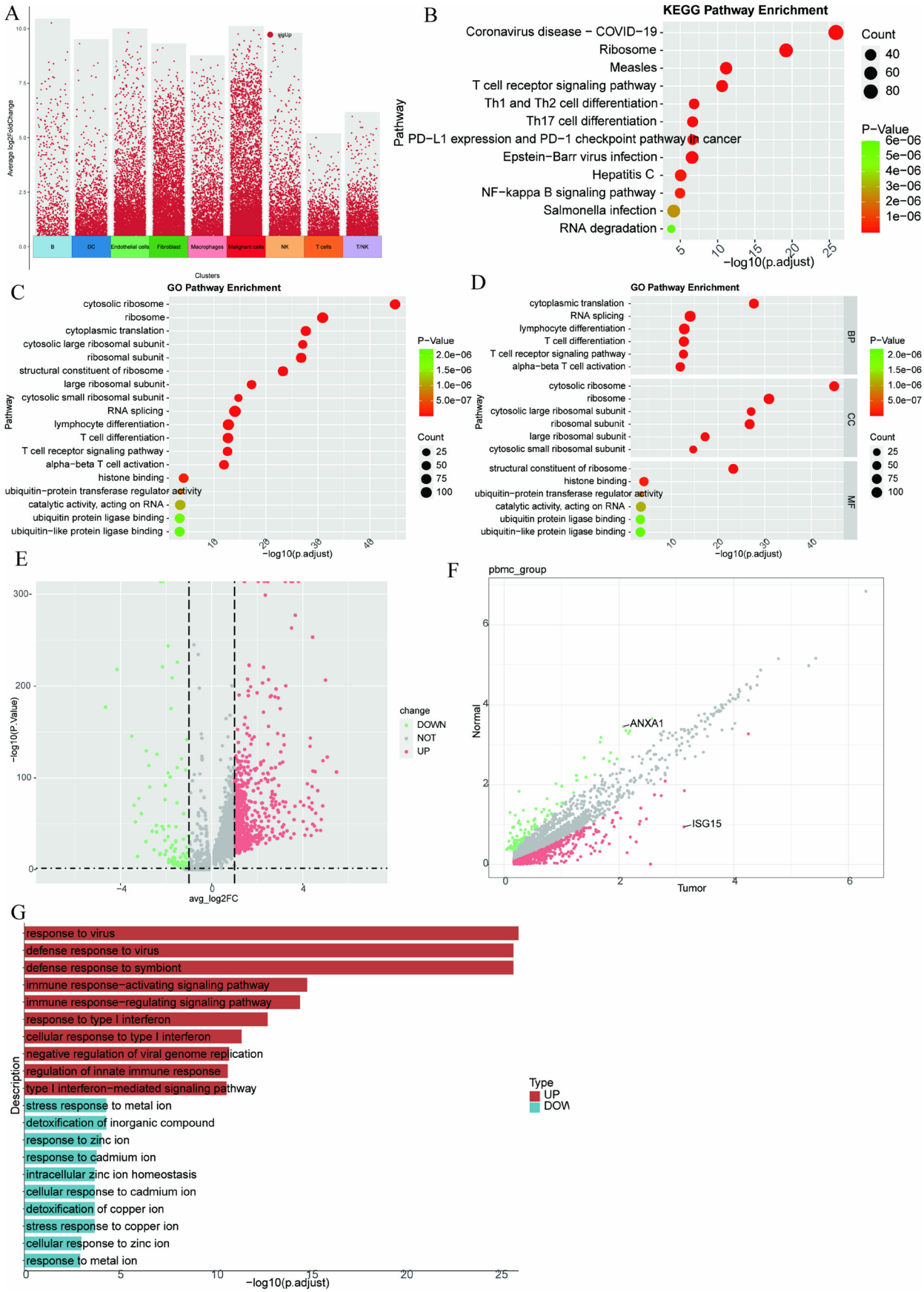


Fig. 3 Analysis of single-cell atlas about T cell subsets in HPV-negative OSCC. **A** T-SNE map of T cells in ScRNA seq, clustered into 5 clusters; **B** Identification of four T Cell Subsets based on marker gene expression status; **C** t-SNE clustering diagram comparing T cells in normal and tumor tissues; **D** Bubble plot displaying four T cell subtype marker genes; **E** Heatmap displays differential gene expression among four subtypes of T cells; **F** The Umap plot the distinct expression patterns exhibited by the marker genes of four distinct T cell subtypes; **G** Distribution of T cell subsets in tumor tissue and normal tissue in ScRNA seq dataset;


A
Average log2FoldChange
Clusters
B
Cell
DC
Endothelial cells
Fibroblast
Macrophages
Malignant cells
NK
T cells
T/NK
KEGG Pathway Enrichment
Coronavirus disease - COVID-19
Ribosome
Measles
T cell receptor signaling pathway
Th1 and Th2 cell differentiation
Th17 cell differentiation
PD-L1 expression and PD-1 checkpoint pathway in cancer
Epstein-Barr virus infection
Hepatitis C
NF-kappa B signaling pathway
Salmonella infection
RNA degradation
Pathway
-log10(p.adjust)
Count
P-Value
C
GO Pathway Enrichment
cytosolic ribosome
ribosome
cytoplasmic translation
cytosolic large ribosomal subunit
ribosomal subunit
structural constituent of ribosome
large ribosomal subunit
cytosolic small ribosomal subunit
RNA splicing
lymphocyte differentiation
T cell differentiation
T cell receptor signaling pathway
alpha-beta T cell activation
histone binding
ubiquitin-protein transferase regulator activity
catalytic activity, acting on RNA
ubiquitin protein ligase binding
ubiquitin-like protein ligase binding
D
BP
CC
MF
E
-log10(P.Value)
avg_log2FC
change
DOWN
NOT
UP
F
pbmc_group
Normal
Tumor
ANXA1
ISG15
G
response to virus
defense response to virus
defense response to symbiont
immune response-activating signaling pathway
immune response-regulating signaling pathway
response to type I interferon
cellular response to type I interferon
negative regulation of viral genome replication
regulation of innate immune response
type I interferon-mediated signaling pathway
stress response to metal ion
detoxification of inorganic compound
response to zinc ion
response to cadmium ion
intracellular zinc ion homeostasis
cellular response to cadmium ion
detoxification of copper ion
stress response to copper ion
cellular response to zinc ion
response to metal ion
Description
Type
UP
DOWN

◂**Fig. 4** Enrichment analysis of T cell-related DEGs pathways.**A** Volcanic diagram displays the analysis of inter cell differences in overall cell subpopulations, with red dots representing genes specific to cell clusters relative to other cell clusters; **B** Analysis of KEGG functional enrichment of T cell; **C** Functional enrichment analysis of GO with intercellular differences in T cells; **D** Functional enrichment analysis of DEGs among T cells, with a faceted presentation of the enrichment patterns of biological processes (BP), molecular functions (MF), and cellular components (CC) within GO entries; **E** Scatter plot analyzing inter-group differences in T cells; **F** Volcano plot of inter-group differential analysis of T cells; **G** Bar chart depicting gene enrichment disparities among various T cell groups

Supplementary Table S4).The examination of inter-group variations in T cells indicated that, relative to normal tissues, 672 T cell-associated marker genes were markedly elevated and 102 were dramatically downregulated in tumor tissues (Fig. 4E-F, Supplementary Table S5).The enrichment analysis of differentially expressed genes between groups indicated that the immune response activating signaling pathway, immune response regulating signaling pathway, and response to type I interferon were upregulated in the tumor group, while the stress response to metal ions and detoxification of inorganic compounds were downregulated (Fig. 4G).

Pseudotime analysis of T lymphocytes in single-cell RNA sequencing

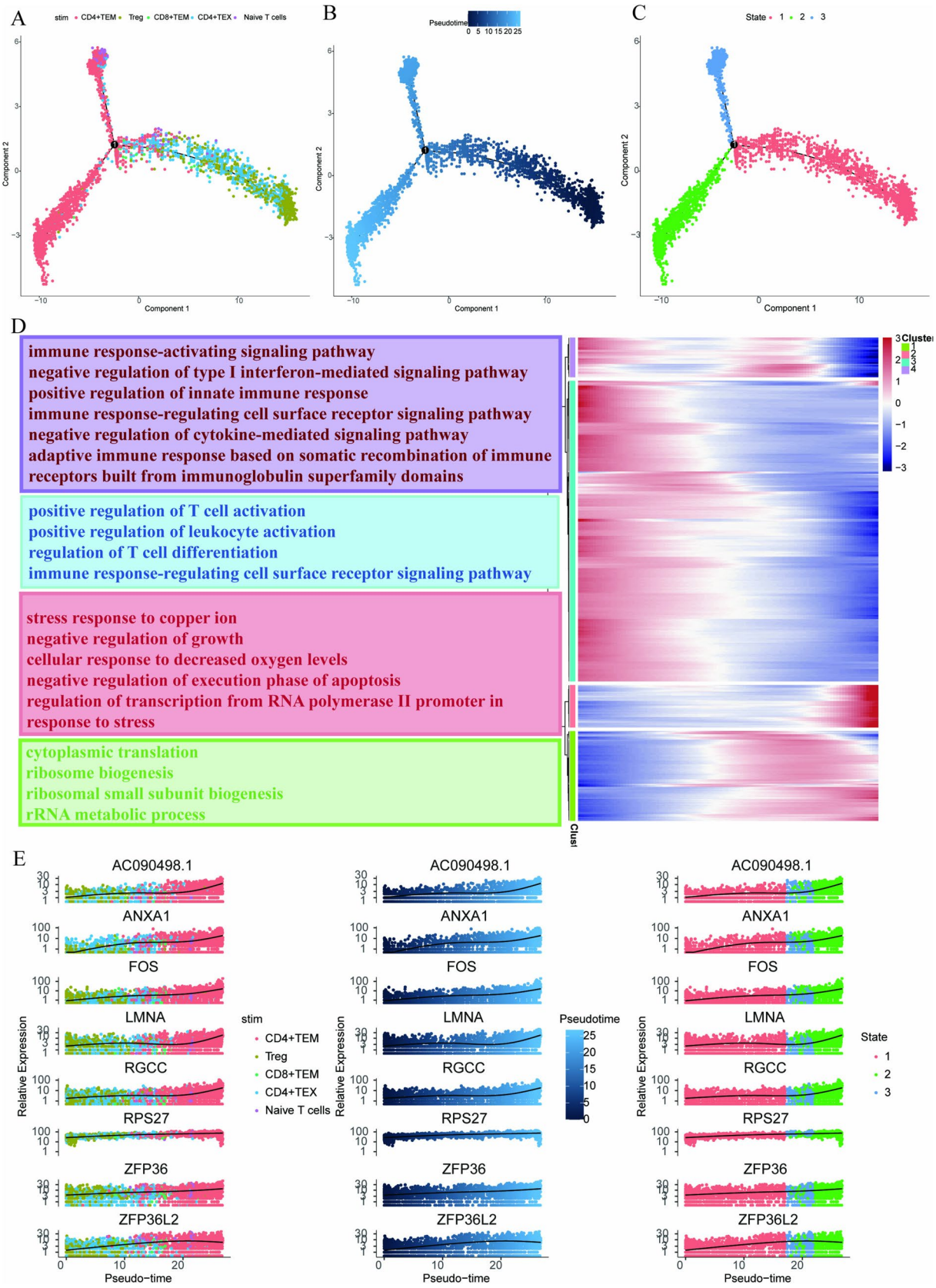
Pseudotime analysis examines the developmental trajectories of various cells by meticulously evaluating the expression patterns of temporal genes in single-cell data. This work involved the extraction of T cells to clarify their developmental pathway, revealing three branches of T cell subtypes (Fig. 5A). Figure 5C illustrates that T cells progress through three phases of differentiation during development, whereas Fig. 5B represents the sequence of cell subtype differentiation, with dark cells evolving into light cells. This signifies that CD4 + TEM cells are in the early phase of cellular formation, whereas memory CD8 + TEM cells denote the final differentiation stage. We subsequently employed heatmaps to evaluate differentially expressed genes (DEGs) across time, categorized them into four clusters, and performed Gene Ontology (GO) pathway enrichment analysis. The findings suggest that cluster 1 is primarily associated with cytoplasmic translation(Cao and Qin 2016), ribosome biosynthesis, ribosomal subunit biosynthesis, and rRNA metabolic activities. Cluster 2 primarily encompasses the stress response to copper ions, the negative regulation of development, the cellular reaction to reduced oxygen levels, the negative regulation of the execution stage of apoptosis, and the transcriptional regulation of the stress response by RNA polymerase II promoter. Cluster 3 mostly pertains to the positive regulation of T cell activation. The cell surface receptor signaling pathway that facilitates leukocyte activation and modulates T cell differentiation to govern the immunological response. Cluster 4 is primarily linked to immune response activation signaling pathways, negative regulation of type I interferon-mediated signaling pathways, immune response regulation of cell surface receptor signaling pathways, and adaptive immune responses resulting from somatic recombination of immune receptors composed of immunoglobulin superfamily domains (Fig. 5D, Supplementary Table S6). The expression levels of AC090498.1, ANXA1, FOS, LMNA, and RGCC progressively elevate throughout the development of tumor-associated T cells. Nonetheless, ZFP36L2 continues to exhibit a declining tendency (Fig. 5E).

Different cell-type communication analysis

This study simulates ligand receptor interactions between cell types using CellChat to create an intercellular communication network. Researchers found that tumor samples have more intercellular communication than normal samples (Fig. 6A-C). Upon examining signal pathway analysis, we observed significant differences in the frequency and intensity of cellular connections between tumor and normal tissues (Fig. 6D). This investigation revealed the significant activation of pathways including FN1, LAMININ, and GALECTIN in cancers, suggesting their possible critical roles in tumor evolution. Continuing along this trajectory, our following analysis, as illustrated in Fig. 6E, revealed substantial changes in ligand-receptor pairs within tumor tissues. These particular molecular interactions indicate essential regulatory functions in intercellular communication, thereby revealing possible novel molecular targets for therapeutic intervention. An in-depth examination of the GALECTIN pathways' functions in cellular communication uncovered unique dynamics (Fig. 6F). The GALECTIN signaling pathway is prevalent in ligand-receptor interactions across diverse cell types, with the LGALS9-CD44 ligand-receptor pair making the most substantial contribution to this route. The comparative study of essential molecules in the GALECTIN pathway revealed significant expression differences in LGALS9 between tumor and normal samples (Fig. 6G–J).

Forecasting and verification of T cell-associated prognostic gene models

The TCGA-HNSC dataset was analyzed for differential genes (Supplementary Table S7). We analyzed 95 HPV-negative OSCC samples from the TCGA-HNSCC database (93 oral cancer and 2 normal). The 986 DEGs intersected with the 774 T cell-related DEGs from scRNA-seq, yielding 58 overlapping genes (Fig. 7A and Supplementary Table S8). After univariate Cox regression analysis, prognostic genes were discovered. Their risk scores were calculated using LASSO regression, 15 genes were included in the above analysis (Supplementary


A
stim CD4+TEM Treg CD8+TEM CD4+TEX Naive T cells
Component 1
Component 2
B
Pseudotime
0 5 10 15 20 25
C
State 1 2 3
D
immune response-activating signaling pathway
negative regulation of type I interferon-mediated signaling pathway
positive regulation of innate immune response
immune response-regulating cell surface receptor signaling pathway
negative regulation of cytokine-mediated signaling pathway
adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains
positive regulation of T cell activation
positive regulation of leukocyte activation
regulation of T cell differentiation
immune response-regulating cell surface receptor signaling pathway
stress response to copper ion
negative regulation of growth
cellular response to decreased oxygen levels
negative regulation of execution phase of apoptosis
regulation of transcription from RNA polymerase II promoter in response to stress
cytoplasmic translation
ribosome biogenesis
ribosomal small subunit biogenesis
rRNA metabolic process
3Cluster
Clust
E
AC090498.1
ANXA1
FOS
LMNA
RGCC
RPS27
ZFP36
ZFP36L2
Relative Expression
Pseudo−time
stim
CD4+TEM
Treg
CD8+TEM
CD4+TEX
Naive T cells
Pseudotime
State

◂**Fig. 5** Pseudotime analysis of T lymphocytes in single-cell RNA sequencing.**A** The positioning of distinct T-cell subpopulations within the pseudotime; **B** Sequential differentiation patterns of T cell subtypes; **C** The positions of cells at various pseudotime; **D** A heatmap depicting the distribution and enrichment status of genes exhibiting pseudotime differences is proposed. This heatmap will showcase the genes that undergo changes over time, organized into four distinct clusters. Additionally, the text within the left-side bar chart will indicate the enriched pathways associated with these genes; **E** Scatterplot of pseudotime differences of genes changes

Table S9). Finally, the results of multivariate COX regression showed that four genes (PMEPA1, PRDX4, SH2D2A and SMS) could serve as independent prognostic genes in HPV-negative OSCC patients (Supplementary Table S10). A prognostic model was created based on the above four genes using the HPV-negative TCGA-OSCC dataset as a training set, GSE41613 and GSE85446 as the testing sets.We categorised patients into high and low risk groups based on median risk scores (Fig. 7B-F). In the training set, overall survival was shorter in the higher risk group compared to the lower risk group (Fig. 7G). The time-dependent ROC curve analysis showed that the prognostic model accurately predicted HPV-OSCC patient survival within the training sets, with an AUC of 0.75 for 1-, 3-, and 5-year survival (Fig. 7H). We next performed univariate and multivariate Cox analyses to assess if the risk score is an independent prognostic factor for HPV-negative OSCC patients compared to other clinicopathological factors. Figure 7I–J shows that PMEPA1, PRDX4, SH2D2A, SMS, N, stage, and risk scores correlated positively with overall survival (OS) in univariate Cox regression analysis. A multivariate analysis revealed that PRDX4 (HR: 2.5, 95% CI: 1.12–5.5, $P<0.05$) and risk Scores (HR: 1, 95% CI: 1–1, $P<0.05$) significantly impact overall survival (OS) in HPV-negative OSCC patients, suggesting they may be independent prognostic markers, Risk score's hazard ratio is about 1, making it unreliable as a predictor. This may be due to our small sample.To corroborate our findings, we will collect a larger sample. Following the technique, the nomogram had significant multivariate logistic analysis variables. Age, stage, T, N, M, and risk score were included in the nomogram (Fig. 7K). We tested our T cell-related differential gene prognosis model with GSE41613 and GSE85446 to confirm its accuracy. Our model predicts HPV-negative OSCC prognoses with great accuracy (Fig. 8).

Correlation of the HPV- OSCC with immune cell infiltration and immune checkpoint

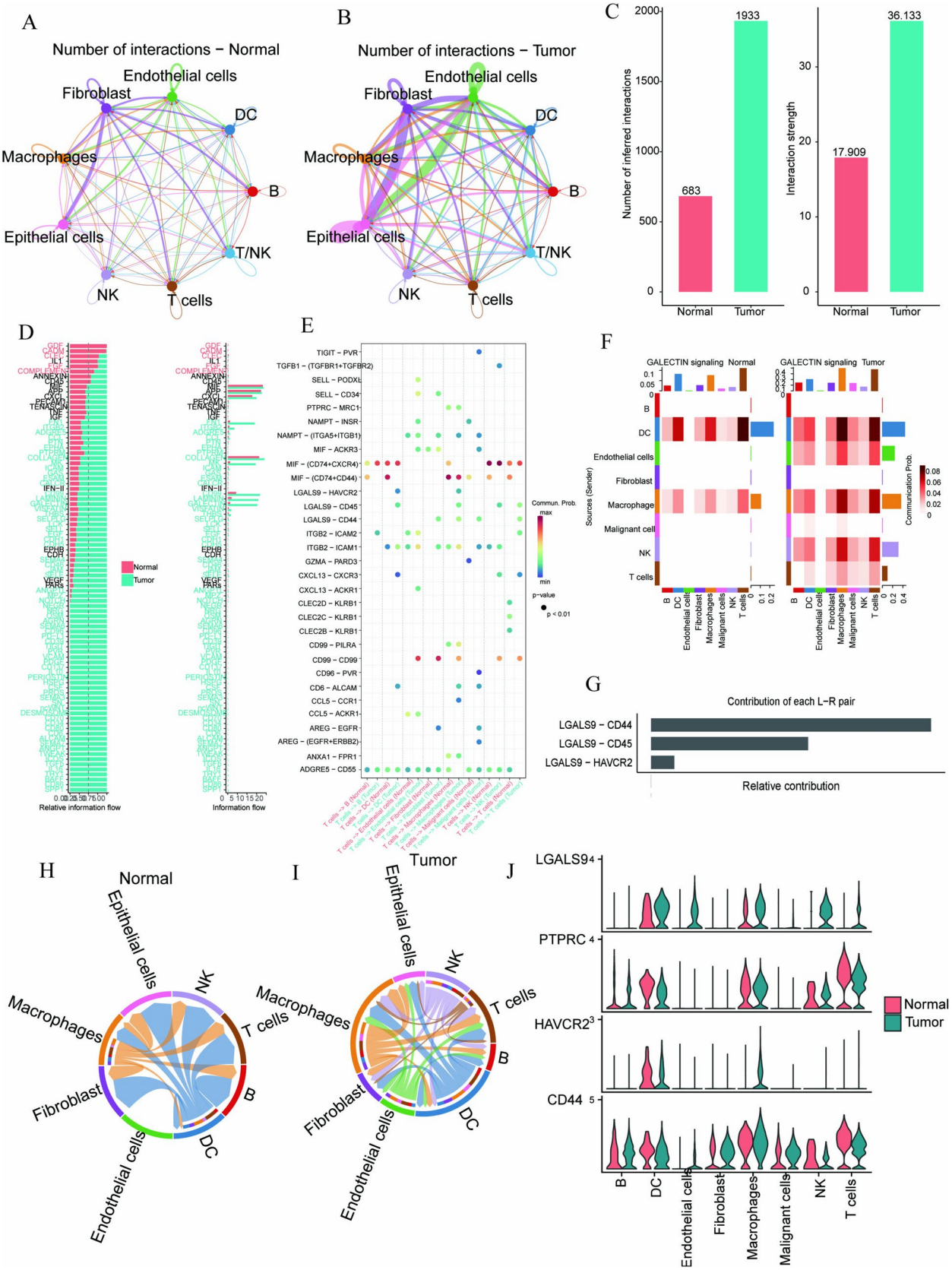
To examine the disparities in tumor microenvironment between high-risk and low-risk cohorts, tumor microenvironment scores were calculated utilizing the ESTIMATE R package. The findings indicated that activated Macrophages M0 and Mast cells were significantly expressed in the high-risk group ($P<0.05$) (Fig. 9A-B), whereas resting mast cells, T follicular helper cells, and regulatory T cells were mostly expressed in the low-risk group (Fig. 9C-E). Furthermore, we examined the relationship between these four prognostic characteristics and twenty-two immune cell types, uncovering significant associations between PMEPA1, PRDX4, SH2D2A, and SMS with various T cell subtypes (Fig. 9F).

qRT-PCR

To detect the expression of PMEPA1, SH2D2A, SMS, and PRDX4 in different stages of HPV-negative OSCC tissues, we performed qRT-PCR experiments.The results showed that PMEPA1 (**$P<0.01$), SH2D2A(*$P<0.05$), SMS (*$P<0.05$), PRDX4(*$P<0.01$) expression was significantly up-regulated in stageIII-IV tissues vs. stageI-II ($P<0.05$, Fig. 10A–D and Supplementary Table S11).

Discussion

Oral squamous cell carcinoma (OSCC) is one of the most aggressive forms of head and neck squamous cell carcinoma (HNSCC)(Badwelan et al. 2023).The prognosis is poor, and the 5-year survival rate of patients with oral squamous cell carcinoma (OSCC) is approximately 50%; these poor outcome are associated with high rates of lymph node metastases and locoregional recurrence(Zhu et al. 2022). Smoking and alcohol consumption are well-established risk factors, whereas human papillomavirus (HPV) infection is becoming increasingly recognized as a significant risk factor for OSCC(Hübbers and Akgül 2015). The HPV status has been incorporated for the stratification of oral cancer patients regarding outcomes and therapeutic responses, with HPV-negative OSCC patients exhibiting markedly inferior survival rates and diminished therapeutic responses compared to those with HPV-positive tumors(Zhang et al. 2020; Lohavanichbutr et al. 2013). At present, scRNA seq is widely used to describe the basic characteristics of tumor infiltrating immune cells. As an important component of the immune system, T cells play a crucial role in exerting tumor killing effects and regulating the overall immune response of the body(O'Donnell et al. 2018).This study utilized ScRNA seq to explore the DEGs associated with T cells in HPV-negative OSCC, and integrated bulk RNA data from TCGA and GEO to preliminarily explore their potential value in prognosis assessment and molecular targeted therapy.TME is a complex ecosystem composed of tumor cells, fibroblasts, and various immune cells. Numerous research results have shown that TME plays an important role in promoting tumor invasion and metastasis, resistance to radiotherapy and


A
Number of interactions − Normal
Endothelial cells
Fibroblast
DC
Macrophages
B
Epithelial cells
T/NK
NK
T cells
B
Number of interactions − Tumor
Endothelial cells
Fibroblast
DC
Macrophages
B
Epithelial cells
T/NK
NK
T cells
C
Number of inferred interactions
2000
1500
1000
500
0
1933
683
Normal
Tumor
Interaction strength
30
20
10
0
36.133
17.909
Normal
Tumor
D
Normal
Tumor
Relative information flow
Information flow
E
TIGIT − PVR
TGFB1 − (TGFBR1+TGFBR2)
SELL − PODXL
SELL − CD34
PTPRC − MRC1
NAMPT − INSR
NAMPT − (ITGA5+ITGB1)
MIF − ACKR3
MIF − (CD74+CXCR4)
MIF − (CD74+CD44)
LGALS9 − HAVCR2
LGALS9 − CD45
LGALS9 − CD44
ITGB2 − ICAM2
ITGB2 − ICAM1
GZMA − PARD3
CXCL13 − CXCR3
CXCL13 − ACKR1
CLEC2D − KLRB1
CLEC2C − KLRB1
CLEC2B − KLRB1
CD99 − PILRA
CD99 − CD99
CD96 − PVR
CD6 − ALCAM
CCL5 − CCR1
CCL5 − ACKR1
AREG − EGFR
AREG − (EGFR+ERBB2)
ANXA1 − FPR1
ADGRE5 − CD55
Commun. Prob.
max
min
p-value
p < 0.01
F
GALECTIN signaling Normal
GALECTIN signaling Tumor
Sources (Sender)
B
DC
Endothelial cells
Fibroblast
Macrophage
Malignant cell
NK
T cells
Communication Prob.
G
Contribution of each L−R pair
LGALS9 − CD44
LGALS9 − CD45
LGALS9 − HAVCR2
Relative contribution
H
Normal
Epithelial cells
NK
Macrophages
T cells
Fibroblast
B
Endothelial cells
DC
I
Tumor
Epithelial cells
NK
Macrophages
T cells
B
Fibroblast
DC
Endothelial cells
J
LGALS9
PTPRC
HAVCR2
CD44
Normal
Tumor
B
DC
Endothelial cells
Fibroblast
Macrophages
Malignant cells
NK
T cells

◂**Fig. 6** Different Cell-type communication analysis. **A**-**B** Number of cell interactions between normal and tumor groups; **C** Comparison of the quantity and intensity of cell communication between tumor tissue and normal tissue; **D** Differences in the quantity and intensity of cell communication related to signaling pathways between tumors and normal tissues; **E** Identifying signal transduction by comparing the probability of ligand receptor pairs mediated communication between T cells and other cell types in control and tumor samples; **F** Heatmap depicting the ligand-receptor interaction patterns in the GALECTIN signaling pathway across various cell types in both tumor and normal groups; **G** Relative contribution of tumor receptor ligand pairs in the GALECTIN pathway; **H-I** Cell to cell communication mediated by LGALS9-CD44 single receptor ligand pairs in tumor and normal groups; **J** Differential expression of GALECTIN pathway molecules between tumor and normal samples

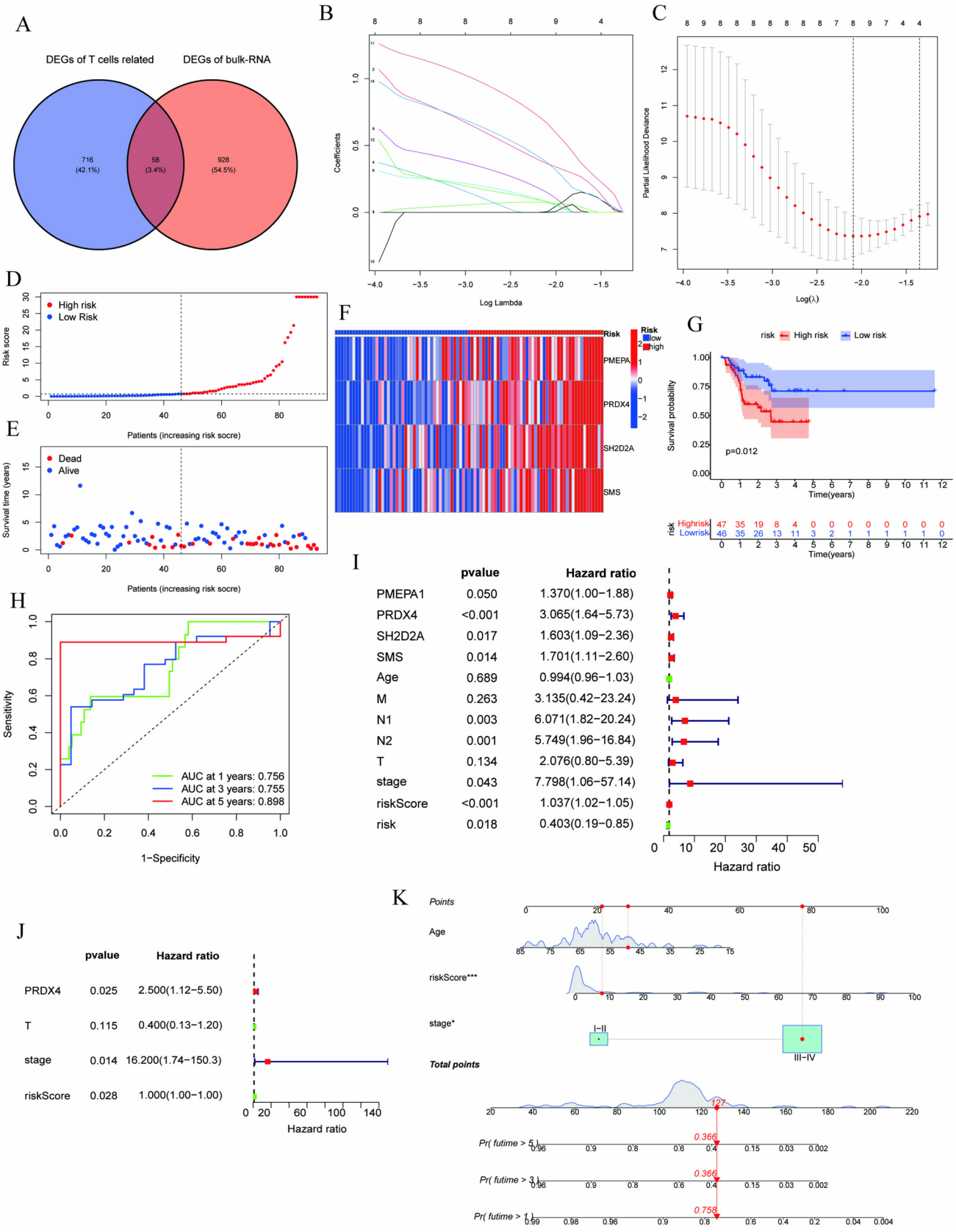
chemotherapy, and low response to immunotherapy(Visser and Joyce 2023; Quail and Joyce 2013). Therefore, it is crucial to systematically elucidate the TME of HPV negative OSCC patients and construct T cell related prognostic models.

This study conducted ScRNA seq analysis on 14 human HPV negative OSCCs and 6 normal control tissues in the GSE181919 dataset to elucidate the tumor microenvironment of HPV negative OSCCs. After filtering, a total of 28,000 cells were included for subsequent analysis. Based on T cell marker genes, 4635 T cells were identified, including 942 Treg cells, 2652 CD4 + TEM cells, 814 CD4 + TEX cells, 37 CD8 + TEM cells, and 190 Naive T cells, Treg cells play an important role in tumor immune suppression and immune escape processes, and appropriately reducing Treg cells may promote anti-tumor immune responses(Shan et al. 2022). TEM cells induce DC cell activation by connecting members of the tumor necrosis factor receptor superfamily, exhibiting significant anti-tumor effects(Gattinoni et al. 2011). Naive T cells recognize MHC antigen peptides presented by DCs, monocytes, or B cells and differentiate into different T cell subpopulations under the induction of microenvironmental cytokines. Therefore, specific regulation of the differentiation microenvironment may contribute to the differentiation of effector T cells and enhance anti-tumor effects(Das et al. 2019). Currently, a large amount of research is focused on CD8 + TEX, with limited understanding of CD4 + TEX. Some studies have found that CD4 + TEX can promote the progress of bladder cancer and immune therapy resistance by regulating the epithelial mesenchymal transformation of bladder cancer(Wu et al. 2023). In summary, our study indicates a significant increase in the proportion of CD4 + TEX, Naive T cells, CD8 + TEM, and Treg in the TME of HPV negative OSCCs, which is consistent with previous research.

The results of the enrichment analysis indicated that T cell marker genes were predominantly linked to the T cell receptor signaling pathway, Th1 and Th2 cell differentiation, PD-L1 expression and PD-1 checkpoint pathway and Th17 cell differentiation pathways. There are four primary subsets of CD4 + T helper (Th) cells, each characterized by a unique cytokine profile that regulates the immune response through a complex network. These include Th1, Th2, Th17, and T regulatory (Tregs) cells. The balance between Th1 and Th2 cells is essential for tumor-specific immune responses and pro-tumor immune control(Shang et al. 2024; Xiao et al. 2023). Prior research has demonstrated that immunological dysregulation in OSCC patients is characterized by a predominance of Th2 /Th1 cytokines(Gaur et al. 2014; Fraga et al. 2021). The polarization towards Th2 cytokines escalates with illness severity, correlating with a significantly elevated risk, whereas Th1 cytokines (IL2, IFNc) have been identified as protective factors(Tano et al. 2013). The association among various cytokines indicated a direct relationship between IL17A and Th1 cytokines, while TGFb1 exhibited a positive correlation with Th2 cytokines (IL4, IL10). Substantial evidence has implicated Th17 in protective immunity against cancer by promoting the production of Th1 chemokines, such as CXCL10 and CXCL9, and by attracting effector cells to neoplastic tissues(Tano et al. 2013; Leivonen and Kähäri 2007). Furthermore, Th17 cells can stimulate tumor-specific CD8 + T cells and enhance dendritic cell accumulation, hence exerting an anticancer effect(Wang et al. 2010; Ankathatti Munegowda et al. 2011). Consequently, a thorough investigation of the functions and processes of several T cell types in HPV-negative OSCC is crucial for enhancing the prognosis of oral cancer.

The analysis of intercellular communication revealed that T cells engaged in direct and robust interactions with different cell subtypes, facilitated by LAMININ, and GALECTIN signaling pathways. GALECTIN pathway have a broad influence on tumour progression via glycosylation-dependent or independent mechanisms that modulate proliferation, evasion of growth suppressors and immune responses, resistance to cell death, induction of angiogenesis, inflammation and metastasis(Mariño et al. 2023). Laminin can bind to integrins in the extracellular matrix, promote tumor invasion and metastasis, evade apoptosis and stimulate angiogenesis(Stewart and O'Connor 2015). Therefore, targeted therapy against LAMININand GALECTIN may become a new therapeutic method, which may improve the prognosis of tumor patients.

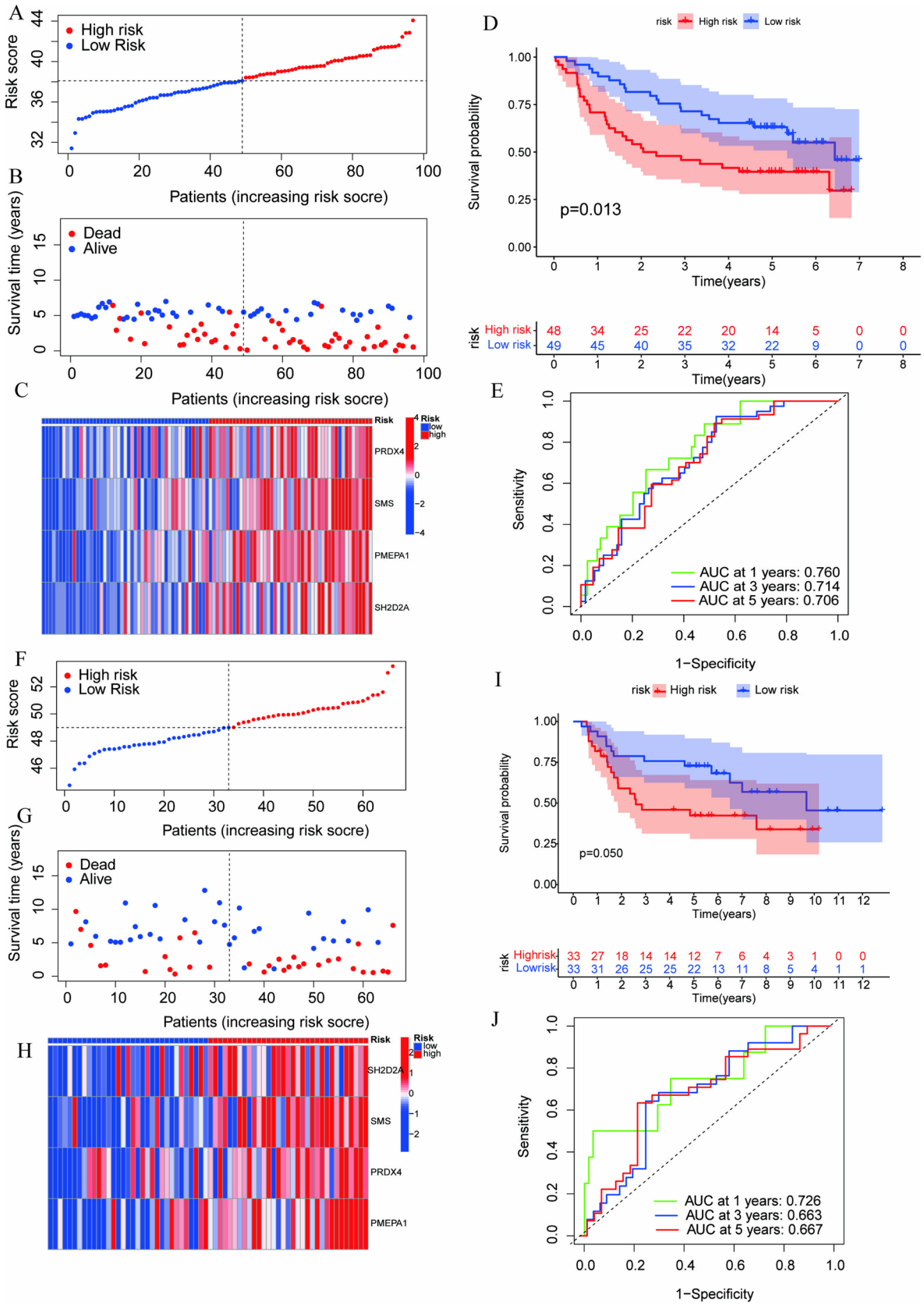
To comprehensively elucidate the correlation between T cells in various tissue types and clinical outcomes in patients with HPV-negative OSCC, a predictive risk model was developed in the TCGA cohort utilizing DEGs of T cell and subsequently confirmed with the GEO cohort. PMEPA1, PRDX4, SH2D2A, and SMS were recognized as independent prognostic biomarkers. PMEPA1, as well as transmembrane prostate androgen-inducible protein 1, is sometimes designated as transmembrane prostate androgen-induced protein (TMEPAI) or solid tumor-associated gene 1 (STAG1). It was first


A
DEGs of T cells related
DEGs of bulk-RNA
716
(42.1%)
58
(3.4%)
928
(54.5%)
B
Coefficients
Log Lambda
C
Partial Likelihood Deviance
Log(λ)
D
High risk
Low Risk
Risk score
Patients (increasing risk socre)
E
Dead
Alive
Survival time (years)
Patients (increasing risk socre)
F
Risk
low
high
PMEPA1
PRDX4
SH2D2A
SMS
G
risk
High risk
Low risk
Survival probability
p=0.012
Time(years)
Highrisk 47 35 19 8 4 0 0 0 0 0 0 0 0
Lowrisk 46 35 26 13 11 3 2 1 1 1 1 1 0
Time(years)
H
Sensitivity
1−Specificity
AUC at 1 years: 0.756
AUC at 3 years: 0.755
AUC at 5 years: 0.898
I
pvalue
Hazard ratio
PMEPA1 0.050 1.370(1.00−1.88)
PRDX4 <0.001 3.065(1.64−5.73)
SH2D2A 0.017 1.603(1.09−2.36)
SMS 0.014 1.701(1.11−2.60)
Age 0.689 0.994(0.96−1.03)
M 0.263 3.135(0.42−23.24)
N1 0.003 6.071(1.82−20.24)
N2 0.001 5.749(1.96−16.84)
T 0.134 2.076(0.80−5.39)
stage 0.043 7.798(1.06−57.14)
riskScore <0.001 1.037(1.02−1.05)
risk 0.018 0.403(0.19−0.85)
Hazard ratio
J
pvalue
Hazard ratio
PRDX4 0.025 2.500(1.12−5.50)
T 0.115 0.400(0.13−1.20)
stage 0.014 16.200(1.74−150.3)
riskScore 0.028 1.000(1.00−1.00)
Hazard ratio
K
Points
Age
riskScore***
stage*
I–II
III–IV
Total points
Pr(futime > 5)
Pr(futime > 3)
Pr(futime > 1)
0.366
0.366
0.758

◂**Fig. 7** Forecasting and verification of T cell-associated prognostic gene models.**A** The Venn diagram illustrates the intersection genes of DEGs associated with T cells in both single-cell samples and TCGA-OSCC samples; **B** Constructing LASSO Regression Model with TCGA-OSCC Samples as Training Set; **C** Model parameter optimization cross validation (the left line represents the optimal λ value); **D-F** Distribution of risk score and survival status in the training set; **G** Kaplan-Meier survival curve constructed based on patient risk scores and survival statuses within the training set; **H** ROC curve for predicting the probability of patient mortality in the training set at 1, 3, and 5 years; **I** Univariate Cox regression analysis was conducted on age, tumor stage, risk score, and the genes selected through LASSO regression screening.**J** Multivariate Cox regression analysis of age, tumor stage, risk score, and genes screened by LASSO regression analysis; **K** The nomogram illustrates the predicted survival rates of patients over a period of 3, 4, and 5 years

identified as an androgen-inducible gene in hormone-responsive LNCaP prostate cancer cells(Sharad et al. 2020). A significant amount of data suggests that PMEPA1 is associated with the advancement of many tumors. Xu et al.(Xu et al. 2003) found that PMEPA1 adversely affects the proliferation of both androgen-responsive and refractory androgen-treated LNCaP prostate cancer cells, likely via its interaction with the NEDD4 protein, which is involved in the ubiquitin-proteasome pathway. Moreover, Wen et al.(Wen et al. 2023) found that elevated expression of PMEPA1 is associated with poor prognosis and reduced anti-tumor immunity in gastric cancer patients. PMEPA1 promotes carcinogenic effects on cell proliferation, colony formation, invasion, and migration in vitro via activating the Wnt/β-catenin signaling pathway, and substantially facilitates the progression of gastric cancer in vivo. In addition, PMEPA1 negatively influenced PTEN protein levels in TNBC cell lines, hence activating the PI3K signaling pathway. This subsequently activated EMT-associated Snai1 by influencing the Snai1/Smad3/Smad4 complex, leading to reduced production of E-cadherin. In triple-negative breast cancer, increased Smad3 and reduced Smad2 levels enable TGF-β to shift from a growth inhibitor to a promoter of growth and metastasis through TMEPAI/PMEPA1(Singha et al. 2014, 2019). We hypothesize that the PMEPA1 gene is a vital regulatory element in tumor proliferation and various biological processes. The biological functions, regulatory processes, and implications of PMEPA1 in HPV-negative OSCC have yet to be investigated. This study shown that HPV-negative OSCC patients with increased PMEPA1 mRNA expression experience worse survival rates, with the underlying mechanism remaining unclear and requiring further exploration. Peroxiredoxin IV (PRDX4) is a multifunctional protein implicated in cellular protection against oxidative damage, regulation of cell proliferation, modulation of intracellular signaling, and tumor pathogenesis(Jain et al. 2021). Research indicates its potential utility as a biomarker for various cancers, including pancreatic, prostate, and gastric cancer(Jia et al. 2019).Zheng et al.(Zheng et al. 2020). demonstrated that PRDX4 overexpression influenced the tumor microenvironment and facilitated tumor progression in a mouse model of urethane-induced lung cancer. A separate study indicated that PRDX4 serves as a negative prognostic indicator for OSCC, with reduced cell migration and invasiveness shown in OEC-M1 cells following in vitro PRDX4 knockdown(Chang et al. 2011). Consequently, it is essential to elucidate the potential mechanistic relationship of PRDX4 in facilitating OSCC progression.SH2D2A, or SH2 domain containing 2 A, also referred to as T cell specific adaptor protein (TSAd), is integral to T cell biology(Kassiotis et al. 2012). It is essential for the interaction between vascular endothelial growth factor receptor 2 (VEGFR2) and c-Src, facilitating c-Src activation and enhancing vascular permeability in tumor vasculature. SH2D2A may impede the progression of bladder cancer(Lin et al. 2023). Conversely, SH2D2A is an essential driver gene that facilitates the spread of breast cancer(Shi et al. 2022a, b). Currently, there is no research investigating the involvement of SH2D2A in the progression of HPV-negative OSCC and the potential mechanisms involved. This work indicates that SH2D2A correlates with unfavorable outcomes in HPV-negative OSCC; nevertheless, the specifics of its mechanism whether it pertains to the modulation of T-cell immune surveillance or T-cell activation necessitate additional investigation. The spermine synthase (SMS) gene, also referred to as MRSR, SPMSY, SRS, and SpS, belongs to the spermidine/spermine synthase family, a group of polyamine biosynthetic enzymes. Research indicates that elevated SMS mRNA expression correlates with diminished overall survival (Hanash et al. 2020).Guo et al.(Guo et al. 2020). discovered that SMS is overexpressed in colorectal cancer (CRC). Targeted disruption of SMS in CRC cells leads to the accumulation of spermidine, which inhibits the acetylation of FOXO3a, facilitating its translocation to the nucleus to stimulate the expression of the pro-apoptotic protein Bim. Qiu et al.(Qiu et al. 2024) discovered that SMS was markedly overexpressed in M2 macrophages of non-small cell lung cancer and exhibited elevated metabolic activity in these macrophages, potentially linked to the polarization of M2 macrophages. Pan et al. (Pan et al. 2022)discovered that SMS is markedly overexpressed in head and neck squamous cell carcinoma and correlates with tumor grade, N stage, and prognosis. Patients exhibiting elevated SMS expression demonstrate significantly reduced overall survival and progression-free survival, aligning with our findings.Consequently, SMS could serve as a prospective target for anti-tumor interventions. In conclusion, PMEAP1, PRDX4, SH2D2A, and SMS warrant consideration as prognostic indicators for HPV-negative OSCC.

TME plays an important role in tumor progression, and comprehending the immune cells within the TME is crucial for advancing immunotherapy. This research examined the disparities in immune cell infiltration between high-risk and low-risk HPV-negative oral squamous cell carcinoma patients.The results indicated that M0 macrophages and activated mast cells were significantly expressed in the high-risk


A
High risk
Low Risk
Risk score
Patients (increasing risk socre)
B
Dead
Alive
Survival time (years)
Patients (increasing risk socre)
C
Risk
low
high
PRDX4
SMS
PMEPA1
SH2D2A
D
risk
High risk
Low risk
Survival probability
p=0.013
Time(years)
High risk 48 34 25 22 20 14 5 0 0
Low risk 49 45 40 35 32 22 9 0 0
E
Sensitivity
1−Specificity
AUC at 1 years: 0.760
AUC at 3 years: 0.714
AUC at 5 years: 0.706
F
High risk
Low Risk
Risk score
Patients (increasing risk socre)
G
Dead
Alive
Survival time (years)
Patients (increasing risk socre)
H
Risk
low
high
SH2D2A
SMS
PRDX4
PMEPA1
I
risk
High risk
Low risk
Survival probability
p=0.050
Time(years)
Highrisk 33 27 18 14 14 12 7 6 4 3 1 0 0
Lowrisk 33 31 26 25 25 22 13 11 8 5 4 1 1
J
Sensitivity
1−Specificity
AUC at 1 years: 0.726
AUC at 3 years: 0.663
AUC at 5 years: 0.667

◂**Fig. 8** Validation of T cell prognostic risk model.**A**-**C** Distribution of Risk Scores and Survival Status in the GSE41613 validation set; **D** Kaplan-Meier survival curve constructed based on patient risk scores and survival statuses within the GSE41613 validation set; **E** ROC curve for predicting the probability of patient mortality in the GSE41613 validation set at 1, 3, and 5 years; **F**-**H** Distribution of Risk Scores and Survival Status in the GSE85446 validation set; **I** Kaplan-Meier survival curve constructed based on patient risk scores and survival statuses within the GSE85446 validation set; **J** ROC curve for predicting the probability of patient mortality in the GSE85446 validation set at 1, 3, and 5 years;

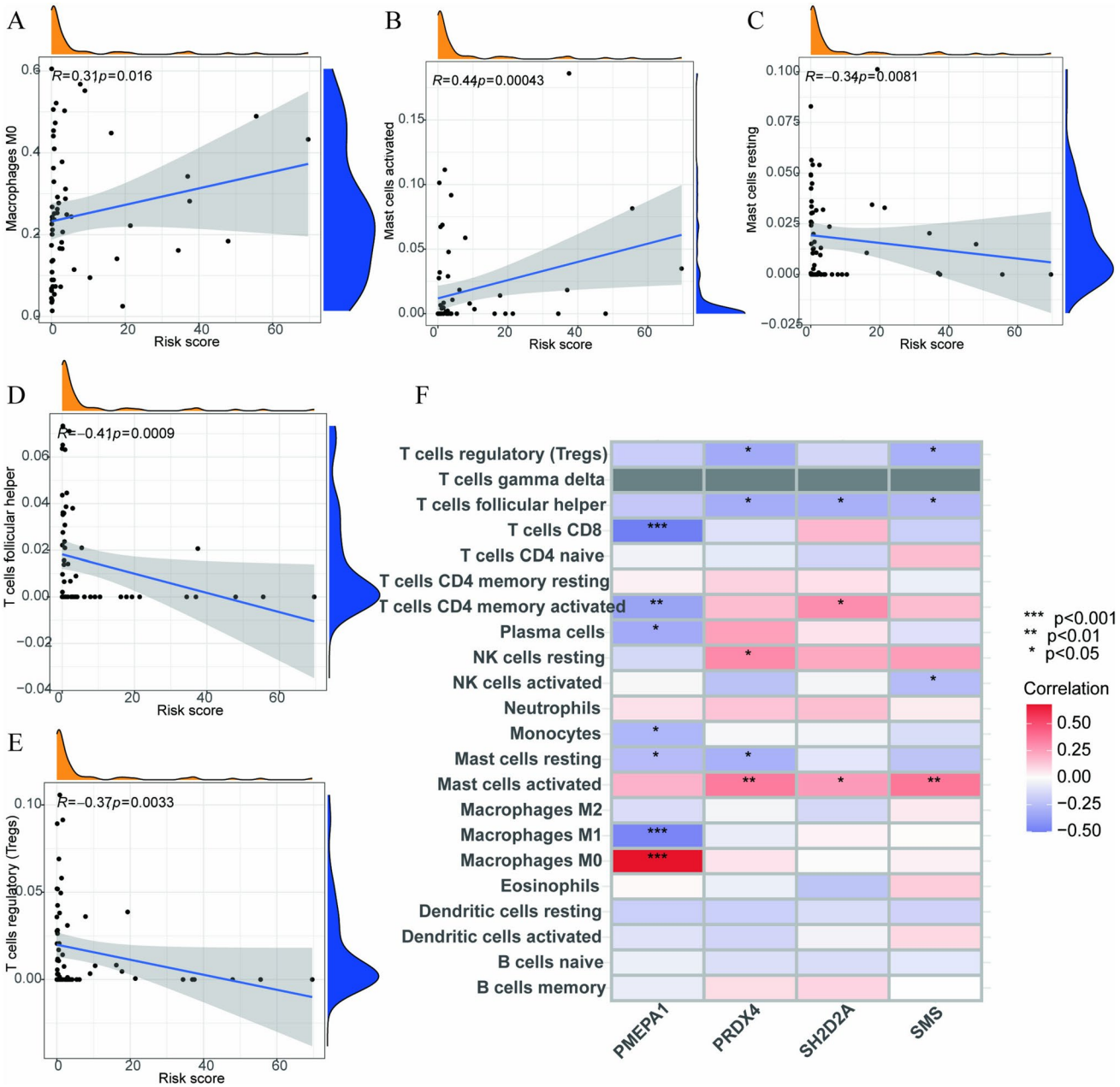


Fig. 9 Correlation of the HPV- OSCC with immune cell infiltration and immune checkpoint.**A**-**B** Macrophages M0 and Mast cells were significantly expressed in the high-risk group; **C**-**E** resting mast cells, group.In the low-risk group, mast cells resting, T cells follicular helper and Tregs were significantly expressed. Some studies indicate that Tregs impede anti-tumor immunity. Nonetheless, various subtypes of Treg exert distinct influences on the prognosis of HNSCC.Neurotrophin-1 (Nrp1) is essential for preserving the stability and functionality of regulatory T cells (Tregs) within malignancies. A high prevalence of NRP1 + Tregs in melanoma and head and neck squamous cell carcinoma correlates with unfavorable T follicular helper cells, and regulatory T cells were mostly expressed in the low-risk group; **F** The heatmap illustrates the correlation among prognostic features associated with immune cells and T cells

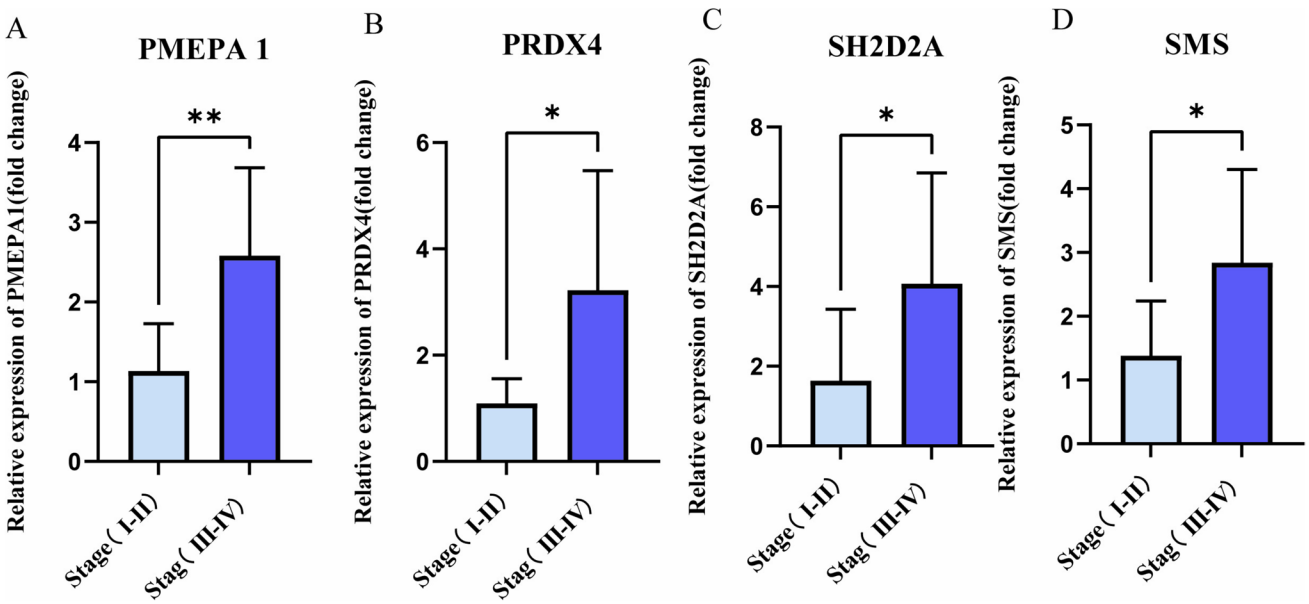


Fig. 10 Relative expression levels of the these four genes in patients with different stages of HPV-negative OSCC .(**A**-**D**) PMEPA1,PRDX4,SH2D2A and SMS are significantly up-regulated in patients with stage III-IV.*$P<0.05$;**$P<0.01$;***$P<0.001$;

prognosis.Conversely, Nrp1-/- Tregs generate interferon-γ (IFNγ), which augments anti-tumor immunity and facilitates tumor eradication(Overacre-Delgoffe et al. 2017).

Consequently, Treg functions as a double-edged sword. Investigating the genetic determinants of Treg subtype transition is crucial for boosting tumor prognosis and improving responses to tumor immunotherapy.T follicular helper cells can enhance immune responses facilitated by B cells and cytotoxic CD8 T cells, and their quantity is frequently correlated with improved prognostic results.This aligns with our research and further substantiates the great precision of the prognostic model we have developed.

This study has limitations.First, prognostic genes were found using public datasets, but the sample size of HPV-negative OSCC patients was modest, requiring validation with a large prospective clinical sample.Second, this study did not investigate the mechanism of HPV-OSCC prognostic gene using molecular biology.

Conclusion

We utilised scRNA-seq to create a single-cell transcriptome profile of HPV-negative oral squamous cell carcinoma and to investigate the role of T-cells in cell trajectory and intercellular communication. In addition, we integrated single-cell data with bulk RNA data from TCGA and GEO datasets to construct a prognostic model of T cell associated DEGs, which enabled us to predict the 1 -, 3 -, and 5-year survival rates of patients.Notably, we can also evaluate immune infiltration within the TME based on these prognostic genes.In the future, we will further confirm the specific mechanisms of these four genes promote tumourigenesis and development through basic experiments, and search for targeted inhibitors of the genes through network pharmacology or molecular drug library screening to provide new strategies for clinical therapy.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10142-024-01525-6.

Acknowledgements We express our profound gratitude to GEO and TCGA for supplying a substantial volume of data. Simultaneously, we express our gratitude to all staff engaged in this investigation.

Author contributions CZ and JL conceptualized this study and were accountable for its evaluation and editing. CL, TZ, FL, CL, SY, and YZ extracted information. WZ and RD engaged in the analysis and interpretation of the data. LW, HZ, HL, and GX conducted a literature review. CL and ZL authored this paper. All authors have reviewed and endorsed the final manuscript.

Funding This research was funded by the Yunnan Fundamental Research Projects (202401AY070001-137,202401AY070001-140), the Basic Research Program of the First People's Hospital of Qujing (2023YJKTB01),the Basic Research Program of Yunnan Provincial Department of Education (2023Y0699,2024J0321,2024Y917).

Data availability The datasets utilized in this investigation are accessible in public web databases.

Declarations

Ethics approval and consent to participate This research employed publicly accessible summary statistical data. All original studies have received approval from their individual ethics review committees, and

participants have submitted informed consent forms. Approval from a new ethical review committee is unnecessary.The clinical samples for this study were approved by the Internal Review Committee of the Qujing affiliated Hospital of Kunming Medical University. Participants all submitted written consent to participate in this study.

Competing interests The authors declare no competing interests.

References

Ankathatti Munegowda M, Deng Y, Mulligan SJ, Xiang J (2011) Th17 and Th17-stimulated CD8 + T cells play a distinct role in Th17-induced preventive and therapeutic antitumor immunity. Cancer Immunol Immunother 60(10):1473–1484

Badwelan M, Muaddi H, Ahmed A, Lee KT, Tran SD (2023) Oral squamous cell carcinoma and Concomitant Primary Tumors, what do we know? A review of the literature. Curr Oncol 30(4):3721–3734

Butler A, Hoffman P, Smibert P, Papalexi E, Satija R (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36(5):411–420

Cao X, Qin Y (2016) Mitochondrial translation factors reflect coordination between organelles and cytoplasmic translation via mTOR signaling: implication in disease. Free Radic Biol Med 100:231–237

Caudell JJ, Gillison ML, Maghami E, Spencer S, Pfister DG, Adkins D et al (2022) NCCN Guidelines® insights: Head and Neck cancers, Version 1.2022. J Natl Compr Canc Netw 20(3):224–234

Chai AWY, Lim KP, Cheong SC (2020) Translational genomics and recent advances in oral squamous cell carcinoma. Sem Cancer Biol 61:71–83

Chang K-P, Yu J-S, Chien K-Y, Lee C-W, Liang Y, Liao C-T et al (2011) Identification of PRDX4 and P4HA2 as Metastasis-Associated proteins in oral cavity squamous cell carcinoma by comparative tissue proteomics of microdissected specimens using iTRAQ Technology. J Proteome Res 10(11):4935–4947

Chao JL, Korzinkin M, Zhavoronkov A, Ozerov IV, Walker MT, Higgins K et al (2021) Effector T cell responses unleashed by regulatory T cell ablation exacerbate oral squamous cell carcinoma. Cell Rep Med. 2(9):100399

Chen Y, Feng Y, Yan F, Zhao Y, Zhao H, Guo Y (2022a) A Novel Immune-related gene signature to identify the Tumor Microenvironment and Prognose Disease among patients with oral squamous cell carcinoma patients using ssGSEA: a Bioinformatics and Biological Validation Study. Front Immunol 13:922195

Chen HLH, Wang J, Li J, Jiang Y (2022b) Identification of a pyroptosis-related prognostic signature in breast cancer. BMC Cancer 22(1):429

Chi HXX, Yan Y, Peng G, Strohmer DF, Lai G, Zhao S, Xia Z, Tian G (2022) Natural killer cell-related prognosis signature characterizes immune landscape and predicts prognosis of HNSCC. Front Immunol 13:101865

Das RK, Vernau L, Grupp SA, Barrett DM (2019) Naïve T-cell deficits at diagnosis and after Chemotherapy Impair Cell Therapy potential in Pediatric Cancers. Cancer Discov 9(4):492–499

de Visser KE, Joyce JA (2023) The evolving tumor microenvironment: from cancer initiation to metastatic outgrowth. Cancer Cell 41(3):374–403

Fang Z, Tian Y, Sui C, Guo Y, Hu X, Lai Y et al (2022) Single-cell transcriptomics of proliferative phase endometrium: systems Analysis of cell–cell Communication Network using CellChat. Front Cell Dev Biology 10:919731

Foy JPBC, Michallet MC, Deneuve S, Incitti R, Bendriss-Vermare N, Albaret MA, Ortiz-Cuaran S, Thomas E, Colombe A, Py C, Gadot N, Michot JP, Fayette J, Viari A, Van den Eynde B, Goudot P, Devouassoux-Shisheboran M, Puisieux A, Caux C, Zrounba P, Lantuejoul S, Saintigny P (2017) The immune microenvironment of HPV-negative oral squamous cell carcinoma from never-smokers and never-drinkers patients suggests higher clinical benefit of IDO1 and PD1/PD-L1 blockade. Ann Oncol 28(8):1934–1941

Fraga M, Yáñez M, Sherman M, Llerena F, Hernandez M, Nourdin G et al (2021) Immunomodulation of T Helper cells by Tumor Microenvironment in oral Cancer is Associated with CCR8 expression and Rapid Membrane Vitamin D Signaling Pathway. Front Immunol 12:643298

Gattinoni L, Lugli E, Ji Y, Pos Z, Paulos CM, Quigley MF et al (2011) A human memory T cell subset with stem cell–like properties. Nat Med 17(10):1290–1297

Gaur P, Singh AK, Shukla NK, Das SN (2014) Inter-relation of Th1, Th2, Th17 and Treg cytokines in oral cancer patients and their clinical significance. Hum Immunol 75(4):330–337

Guha N, Warnakulasuriya S, Vlaanderen J, Straif K (2014) Betel quid chewing and the risk of oral and oropharyngeal cancers: a meta-analysis with implications for cancer control. Int J Cancer 135(6):1433–1443

Guo Y, Ye Q, Deng P, Cao Y, He D, Zhou Z et al (2020) Spermine synthase and MYC cooperate to maintain colorectal cancer cell survival by repressing bim expression. Nat Commun 11(1):3243

Haghverdi L, Büttner M, Wolf FA, Buettner F, Theis FJ (2016) Diffusion pseudotime robustly reconstructs lineage branching. Nat Methods 13(10):845–848

Hanash S, Arun B, Disis ML, Katayama H, Peterson C, Irajizad E et al (2020) Association between Plasma Diacetylspermine and Tumor Spermine Synthase with Outcome in Triple-negative breast Cancer. JNCI: J Natl Cancer Inst 112(6):607–616

Hübbers CU, Akgül B (2015) HPV and cancer of the oral cavity. Virulence 6(3):244–248

Ionkina AA, Balderrama-Gutierrez G, Ibanez KJ, Phan SHD, Cortez AN, Mortazavi A et al (2021) Transcriptome analysis of heterogeneity in mouse model of metastatic breast cancer. Breast Cancer Res 23(1):93

Jain PD-GA, Mollen E, Malbeteau L, Xie M, Jessa F, Dhavarasa P, Chung S, Brown KR, Jang GH, Vora P, Notta F, Moffat J, Hedley D, Boutros PC, Wouters BG, Koritzinsky M (2021) NOX4 links metabolic regulation in pancreatic cancer to endoplasmic reticulum redox vulnerability and dependence on PRDX4. Sci Adv 7(19):eabf7114

Jia W, Chen P, Cheng Y (2019) PRDX4 and its roles in various cancers. Technol Cancer Res Treat 18:1533033819864313

Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H et al (2021) Inference and analysis of cell-cell communication using CellChat. Nat Commun 12(1):1088

Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR (2020) Head and neck squamous cell carcinoma. Nat Reviews Disease Primers 6(1):92

Kamperman T, Karperien M, Le Gac S, Leijten J (2018) Single-cell microgels: Technology, challenges, and applications. Trends Biotechnol 36(8):850–865

Kassiotis G, Berge T, Grønningsæter IHB, Lorvik KB, Abrahamsen G, Granum S et al (2012) SH2D2A modulates T cell mediated Protection to a B cell derived Tumor in Transgenic mice. PLoS ONE 7(10):e48239

Kim S, Leem G, Choi J, Koh Y, Lee S, Nam S-H et al (2024) Integrative analysis of spatial and single-cell transcriptome data from human pancreatic cancer reveals an intermediate cancer cell population associated with poor prognosis. Genome Med 16(1):20

Leivonen SK, Kähäri VM (2007) Transforming growth factor-β signaling in cancer invasion and metastasis. Int J Cancer 121(10):2119–2124

Lin F, Ke Z-B, Xue Y-T, Chen J-Y, Cai H, Lin Y-Z et al (2023) A novel CD8 + T cell-related gene signature for predicting the prognosis and immunotherapy efficacy in bladder cancer. Inflamm Res 72(8):1665–1687

Liu Y, Wu G (2023) The utilization of single-cell sequencing technology in investigating the immune microenvironment of ccRCC. Front Immunol 14:1276658

Liu B, Si W, Wei B, Zhang X, Chen P (2023) PTP4A1 promotes oral squamous cell carcinoma (OSCC) metastasis through altered mitochondrial metabolic reprogramming. Cell Death Discovery 9(1)

Lohavanichbutr P, Méndez E, Holsinger FC, Rue TC, Zhang Y, Houck J et al (2013) A 13-Gene signature Prognostic of HPV-Negative OSCC: Discovery and External Validation. Clin Cancer Res 19(5):1197–1203

Mariño KV, Cagnoni AJ, Croci DO, Rabinovich GA (2023) Targeting galectin-driven regulatory circuits in cancer and fibrosis. Nat Rev Drug Discovery 22(4):295–316

Melo BAC, Vilar LG, Oliveira NRd L, POd, Pinheiro MB, Domingueti CP et al (2021) Human papillomavirus infection and oral squamous cell carcinoma - a systematic review. Braz J Otorhinolaryngol 87(3):346–352

Ng JH, Iyer NG, Tan MH, Edgren G (2016) Changing epidemiology of oral squamous cell carcinoma of the tongue: a global study. Head Neck 39(2):297–304

O'Donnell JS, Teng MWL, Smyth MJ (2018) Cancer immunoediting and resistance to T cell-based immunotherapy. Nat Reviews Clin Oncol 16(3):151–167

Overacre-Delgoffe AE, Chikina M, Dadey RE, Yano H, Brunazzi EA, Shayan G et al (2017) Interferon-γ drives Treg Fragility to promote anti-tumor immunity. Cell 169(6):1130–41e11

Pan X, Xue L, Sun Y (2022) Spermine synthase (SMS) serves as a prognostic biomarker in head and neck squamous cell carcinoma: a bioinformatics analysis. Annals Translational Med 10(22):1213

Qiu J, Wang Z, Yu Y, Zheng Y, Li M, Lin C (2024) Prognostic and immunological implications of glutathione metabolism genes in lung adenocarcinoma: a focus on the core gene SMS and its impact on M2 macrophage polarization. Int Immunopharmacol 132:111940

Quail DF, Joyce JA (2013) Microenvironmental regulation of tumor progression and metastasis. Nat Med 19(11):1423–1437

Sagheer SH, Whitaker-Menezes D, Han JYS, Curry JM, Martinez-Outschoorn U, Philp NJ (2021) 4NQO induced carcinogenesis: a mouse model for oral squamous cell carcinoma. Carcinogen-driven mouse models of oncogenesis. Methods Cell Biol 163:p. 93–111

Shan F, Somasundaram A, Bruno TC, Workman CJ, Vignali DAA (2022) Therapeutic targeting of regulatory T cells in cancer. Trends Cancer 8(11):944–961

Shang Q, Yu X, Sun Q, Li H, Sun C, Liu L (2024) Polysaccharides regulate Th1/Th2 balance: a new strategy for tumor immunotherapy. Biomed Pharmacother 170:115976

Sharad S, Dobi A, Srivastava S, Srinivasan A, Li H (2020) PMEPA1 gene isoforms: a potential biomarker and therapeutic target in prostate Cancer. Biomolecules 10(9):1221

Shi X, Dong A, Jia X, Zheng G, Wang N, Wang Y et al (2022a) Integrated analysis of single-cell and bulk RNA-sequencing identifies a signature based on T-cell marker genes to predict prognosis and therapeutic response in lung squamous cell carcinoma. Front Immunol 13:992990

Shi Y, Zhang D, Chen J, Jiang Q, Song S, Mi Y et al (2022b) Interaction between BEND5 and RBPJ suppresses breast cancer growth and metastasis via inhibiting notch signaling. Int J Biol Sci 18(10):4233–4244

Singha PKPS, Geng H, Lan R, Venkatachalam MA, Saikumar P (2014) TGF-β induced TMEPAI/PMEPA1 inhibits canonical smad signaling through R-Smad sequestration and promotes non-canonical PI3K/Akt signaling by reducing PTEN in triple negative breast cancer. Genes Cancer 5(9–10):320–336

Singha PKPS, Geng H, Lan R, Venkatachalam MA, Dobi A, Srivastava S, Saikumar P (2019) Increased Smad3 and reduced Smad2 levels mediate the functional switch of TGF-β from growth suppressor to growth and metastasis promoter through TMEPAI/PMEPA1 in triple negative breast cancer. Genes Cancer 10(5–6):134–149

Stewart RL, O'Connor KL (2015) Clinical significance of the integrin α6β4 in human malignancies. Lab Invest 95(9):976–986

Stuart TSR (2019) Integrative single-cell analysis. Nat Rev Genet 20(5):257–272

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A et al (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. ;71(3):209 – 49

Tano T, Okamoto M, Kan S, Bando T, Goda H, Nakashiro K-i et al (2013) Immunochemoradiotherapy for patients with oral squamous cell carcinoma: augmentation of OK-432-Induced helper T cell 1 response by 5-FU and X-ray irradiation. Neoplasia 15(7):805–814

van Dijk PC, Jager KJ, Zwinderman AH, Zoccali C, Dekker FW (2008) The analysis of survival data in nephrology: basic concepts and methods of Cox regression. Kidney Int 74(6):705–709

Wang F-S, Agrawal S, Gupta S, Agrawal A (2010) Human dendritic cells activated via Dectin-1 are efficient at priming Th17, cytotoxic CD8 T and B cell responses. PLoS ONE 5(10):e13418

Wang S, Liu S, Zhu Y, Zhang B, Yang Y, Li L et al (2023) A novel and independent survival prognostic model for OSCC: the functions and prognostic values of RNA-binding proteins. Eur Arch Otorhinolaryngol 281(1):397–409

Wen F, Yang S, Cai W, Zhao M, Qin L, Jiao Z (2023) Exploring the role of PMEPA1 in gastric cancer. Mol Cell Probes 72:101931

Wu THE, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innov (Camb) 2(3):100141

Wu C, Duan L, Li H, Liu X, Cai T, Yang Y et al (2023) PD1hi CD200hi CD4 + exhausted T cell increase immunotherapy resistance and tumour progression by promoting epithelial–mesenchymal transition in bladder cancer. Clin Translational Med 13(6):e1303

Xiao Y, Huang Y, Jiang J, Chen Y, Wei C (2023) Identification of the prognostic value of Th1/Th2 ratio and a novel prognostic signature in basal-like breast cancer. Hereditas 160(1):2

Xu LLSY, Petrovics G, Sun C, Makarem M, Zhang W, Sesterhenn IA, McLeod DG, Sun L, Moul JW, Srivastava S (2003) PMEPA1, an androgen-regulated NEDD4-binding protein, exhibits cell growth inhibitory function and decreased expression during prostate cancer progression. Cancer Res 63(15):4299–4304

Yang FNX, Zhou M, Li W (2024) Development and validation of a novel disulfidptosis-related lncRNAs signature in patients with HPV-negative oral squamous cell carcinoma. Sci Rep 14(1):14436

Zhang L, Li Z, Skrzypczynska KM, Fang Q, Zhang W, O'Brien SA et al (2020) Single-cell analyses inform mechanisms of myeloid-targeted therapies in Colon cancer. Cell 181(2):442–59e29

Zhang J, Liu X, Huang Z, Wu C, Zhang F, Han A et al (2023) T cell-related prognostic risk model and tumor immune environment modulation in lung adenocarcinoma based on single-cell and bulk RNA sequencing. Comput Biol Med 152:106460

Zheng J, Guo X, Nakamura Y, Zhou X, Yamaguchi R, Zhang J et al (2020) Overexpression of PRDX4 modulates Tumor Microenvironment and promotes urethane-Induced Lung Tumorigenesis. Oxidative Med Cell Longev 2020:1–11

Zhou BJW (2020) Visualization of single cell RNA-Seq Data using t-SNE in R. Methods Mol Biol 2117:159–167

Zhu W, Wang J, Liu X, Xu Y, Zhai R, Zhang J et al (2022) lncRNA CYTOR promotes aberrant glycolysis and mitochondrial respiration via HNRNPC-mediated ZEB1 stabilization in oral squamous cell carcinoma. Cell Death Dis 13(8):703