RESEARCH ARTICLE

Identification of novel subtypes based on ssGSEA in immune-related prognostic signature for tongue squamous cell carcinoma

Yi Jin[1,2,3,#] | Zhanwang Wang[1,#] | Dong He[4] | Yuxing Zhu[1] | Xingyu Chen[1] | Ke Cao[1]

[1]Department of Oncology, Third Xiangya Hospital of Central South University, Changsha, China

[2]Department of Radiation Oncology, Hunan Cancer Hospital, The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, China

[3]Key Laboratory of Translational Radiation Oncology, Department of Radiation Oncology, Hunan Cancer Hospital and The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, China

[4]Department of Respiratory, The Second People's Hospital of Hunan Province, Changsha, China

Correspondence
Ke Cao, Department of Oncology, Third Xiangya Hospital of Central South University, Changsha, China.
Email: csucaoke@163.com

Dong He, Department of Respiratory, The Second People's Hospital of Hunan Province, Changsha, China.
Email: hedong_4420@sina.com

Abstract

Background: Tongue squamous cell carcinoma (TSCC) is characterized by aggressive invasion and poor prognosis. Currently, immune checkpoint inhibitors may prolong overall survival compared with conventional treatments. However, PD1/PDL1 remain inapplicable in predicting the prognosis of TSCC; thus, it is urgent to explore the genetic characteristics of TSCC.

Materials and methods: We utilized single-sample gene set enrichment analysis (ssGSEA) to classify TSCC patients from the TCGA database into clusters with different immune cell infiltrations. ESTIMATE (immune-related scores) and CIBERSORT (immune cell distribution) analyses were used to evaluate the immune landscape among clusters. GO, KEGG, and GSEA analyses were performed to analyze the different underlying molecular mechanisms in the clusters. Based on the immune characteristics, we applied the LASSO Cox regression to select hub genes and construct a prognostic risk model. Finally, we established an interactive network among these hub genes by using Cytoscape, and a pan-cancer analysis to further verify and decipher the innate function of these genes.

Results: Using ssGSEA, we constructed three functional clusters with different overall survival and immune-cell infiltration. ESTIMATE and CIBERSORT analyses revealed the different distributions of immune cells (T cells, B cells, and macrophages) with diverse immune-related scores (ESTIMATE, immune, stromal, and tumor purity scores). Moreover, pathways including those of the

interferon-gamma response, hypoxia, and glycolysis of the different subtypes were investigated to elucidate their involvement in mediating the heterogeneous immune characteristics. Subsequently, after LASSO Cox regression, a signature of 15 immune-related genes was established that is more prognostically effective than the TNM stage. Furthermore, three hub genes—*PGK1*, *GPI*, and *RPE*—were selected using Cytoscape evaluation and verified by immunohistochemistry. *PGK1*, the foremost regulator, was a comprehensively profiled pan-cancer, and a *PGK1*-based interactive network was established.

Conclusion: Our results suggest that immune-related genes and clusters in TSCC have the potential to guide individualized treatments.

KEYWORDS

CIBERSORT, ESTIMATE, LASSO, PGK1, ssGSEA, tongue squamous cell carcinoma

1 | INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common malignancy globally, and its incidence has been significantly increasing worldwide.[1] HNSCC arises from the squamous epithelium of the oral cavity, oropharynx, larynx, and hypopharynx and represents a heterogeneous group of tumors. Tongue squamous cell carcinoma (TSCC) is one of the most frequent types of cancer, particularly of the oral cavity, and is characterized by aggressive local invasion and early lymph-node metastasis with a relatively poor prognosis.[2,3] Environmental carcinogens such as tobacco, alcohol, or HPV infection may induce the development of TSCC.[4] Conventional treatments for TSCC, including surgical techniques, chemoradiotherapy, and some other target therapies, result in inevitable short-term and long-term morbidity; it has been difficult to improve the overall survival (OS) of patients with TSCC, which has been approximately 50% over recent decades.[5] The tumor, lymph node, and metastasis (TNM) classification system has been widely used and is considered a vital tool in predicting the survival outcomes of TSCC patients. However, this system has been criticized because it cannot be used to decipher the heterogeneous outcomes of individual cancers with the same TNM stage. Thus, it is crucial to reveal the biologic molecular mechanism underlying TSCC development and to discover novel biomarkers for the prediction of clinical outcomes.

Recent research has led to the discovery of several novel biomarkers linked to the establishment and progression of TSCC, especially those involving the tumor genes *TP53*[6] and *EGFR*.[7] Moreover, cetuximab, a monoclonal antibody that targets EGFR, has been shown to have efficacy in locoregionally advanced, recurrent, or metastatic HNSCC in combination with chemotherapy or radiation.[8] However, the efficacy of such treatment with curative intent is limited to the best median OS of about 10 months, and this treatment is associated with substantial toxicity.[9] Furthermore, there are no approved treatment options after TSCC progression.[10] According to the KEYNOTE-048, a randomized, open-label, phase III study, the results suggest that pembrolizumab monotherapy, which targets the programmed cell death 1 (PD-1) receptor,[11] has significantly longer OS and a favorable safety profile in the population with PD-L1 combined positive score (CPS) ≥20 compared with cetuximab plus chemotherapy.[9] Nowadays, the field of immunology has attracted much attention for its irreplaceable clinical benefits. Innate immunologic mechanisms are often observed to correlate with the mediation of the tumor microenvironment, which is composed of immune cells and stromal cells, among others.[12] There is no denying that immune-related genes and immune infiltrating cells have indispensable roles in the tumor microenvironment; yet, little is known regarding the immune landscape of TSCC.[13] Therefore, it is crucial to comprehensively analyze immune-related cells or genes in TSCC to further decipher the mechanism underlying immunology resistance and response.

In this study, we used single-sample gene set enrichment analysis (ssGSEA) to assign TSCC patients from the Cancer Genome Atlas (TCGA) database into different immune-cell infiltration clusters; we used ESTIMATE, CIBERSORT, and K-M analyses to identify the prognostic significance of immune-related risk models or regulators and explored the mechanisms underlying the characterization of the tumor microenvironment.

2 | MATERIALS AND METHODS

2.1 | Data processing

The RNA-seq data (FPKM value) and clinical characteristics (including age, sex, smoking status, tumor stage, and TNM staging) of the TSCC cohort were obtained from TCGA Head and Neck Squamous Cell Carcinomas database. We included samples wherein the primary site was the tongue and obtained publicly available data.[14]

2.2 | ssGSEA and hierarchical clustering analysis

The ssGSEA algorithm was based on 29 immune gene sets, including genes related to different immune cell types, functions, pathways, and checkpoints. We employed the ssGSEA algorithm via R packages (GSVA, GSEABase, and limma) to comprehensively assess the immunologic characteristics of every sample included in the study.[15]

To study the correlation between immunity and clinical phenotype of TSCC, we used "Consensus Cluster Plus" (50 iterations, 80% resampling rate) according to the enrichment score of immune items in ssGSEA to cluster the TSCC samples into three different groups (Clusters A, B, and C). OS analysis between different clusters was conducted using the Kaplan–Meier method.

2.3 | ESTIMATE and CIBERSORT analyses

To verify differences in the immunity of the three clusters, we utilized the "estimate" package for R (https://sourceforge.net/projects/estimateproject) to calculate the tumor purity, stromal score, immune score, and ESTIMATE score of each TSCC tumor sample.[16] The Mann–Whitney U test was used to compare the scores of the three clusters. Then, we used the CIBERSORT package to assess the distribution of 22 immune cell types in each sample. Results with $p < 0.05$ were used in further analysis.[17]

2.4 | Gene Ontology (GO) analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) database analysis, and GSEA

The "edgeR" package in R was used to perform a differential analysis of the mRNAs in the low-expression and high-expression groups.[18] Differentially expressed genes (DEGs) were selected based on the following significance criteria: |logFC| >1 and FDR <0.05. We further performed GO functional annotation, including molecular function, cellular component, and biologic process analysis. The KEGG database analyzes the metabolic pathways and signal transduction pathways in which DEGs are significantly enriched. GSEA was then performed to identify the signaling pathways wherein DEGs were enriched between the high-risk and low-risk subgroups.[19]

2.5 | Construction of the prognostic risk model

The "survival" package in R was used to perform univariate Cox proportional hazard regression analysis to screen for immune-related genes that are significantly linked to the OS of TSCC patients in the TCGA cohort. The "Glmnet" package was used to perform LASSO Cox regression analysis.[20] We calculated the risk score to construct the risk model using the following formula: risk score =expressed $mRNA_1 \times$ coefficient $mRNA_1$ + expressed $mRNA_2 \times$ coefficient $mRNA_2$ + ...expressed $mRNA_n \times$ coefficient $mRNA_n$. Patients were divided into high-risk and low-risk groups according to the value of the risk coefficient. The Kaplan-Meier survival package in R language was used to perform univariate and multivariate Cox proportional hazard regression analyses on the risk value. The Cox analysis signature included gender, age, grade, and risk score, as well as T, N, and M stages. The sensitivity and specificity of the receiver-operating characteristic (ROC) curve were used to evaluate the prognostic performance of the signature.

2.6 | Identification and verification of immune-related hub genes

To understand the underlying interaction of immune-related mRNAs, the STRING website (https://string-db.org/) was employed, and all PPI pairs with a combined score of >0.4 were extracted. Next, we utilized the Cytoscape (v3.6.1) plugin cytoHubba to calculate all degrees of nodes.[21] In the present study, genes with the highest degree values were considered hub genes. The Human Protein Atlas (HPA) online database (http://www.proteinatlas.org/) was used to validate the expression of hub genes at the translational level.

2.7 | Identification of the interaction of immune-related hub genes

To further explore the interaction of hub genes with miRNAs, lncRNAs, and circRNAs, with the criteria of |log2(fold change)| >1 and $p < 0.05$, we identified differentially expressed lncRNAs and miRNAs in the TCGA–TSCC cohort.

Furthermore, we used the miRcode database (mircode.org/index.php) to target miRNA, and we utilized miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/) and miRDB (http://www.mirdb.org/) to identify the correlation between miRNA and mRNA. Subsequently, we screened circRNA–miRNA interaction pairs using the circBank database (http://www.circbank.cn/index.html). Finally, after integrating circRNA–miRNA, lncRNA–miRNA, and miRNA–mRNA regulatory relationships, a competing endogenous RNA (ceRNA) network was established and visualized.

2.8 | Statistical analysis

All analyses were performed using R. Univariate and multivariate Cox proportional hazards regression analyses were also used to assess the relationship between the risk score and OS of patients. ROC analysis was used to detect the sensitivity and specificity of the genetic signature risk score to predict survival. The area under the ROC curve (AUC) was used as an index to assess the accuracy of the prognosis. In all analyses, a p value of <0.05 was considered to be statistically significant.

3 | RESULTS

3.1 | Construction of TSCC groupings based on ssGSEA

In total, 147 TSCC samples and 15 paracancerous samples were obtained from the TCGA database. The ssGSEA

method was applied to the transcriptome of the TSCC samples to evaluate the distribution of 29 immune cell types. As shown in Figure 1A, over half of the immune cell types were found to be upregulated in the TSCC tissue than in normal tissue. The Pearson correlation analysis showed that these immune cell types in the cancer tissue intimately reinforced each other (Figure 1B), suggesting that TSCC is heterogeneous cancer with high immunogenicity.

To further understand the correlations among these immune cell types, we implemented a consensus clustering analysis to classify the TSCC samples into three clusters based on the ssGSEA using the Consensus Cluster Plus package with $k = 3$ (Figure 2A). The Kaplan–Meier method showed that there was a significant difference among clusters, particularly between clusters A and C (Figure 2B). Subsequently, we plotted a boxplot and heatmap to screen the relationship between immune-related groupings and expression of gene sets from ssGSEA. The expression of most immune cell types was higher in cluster A with the low OS than in clusters B and C (Figures 2C and 3A).

3.2 | Immune characterization in the TSCC clusters

To verify the feasibility of the grouping strategy, we conducted an ESTIMATE analysis to profile the immune characteristics of TSCC based on the expression of immune cell types. Stromal, immune, and ESTIMATE scores had a similar trend as the expression of immune cell types,
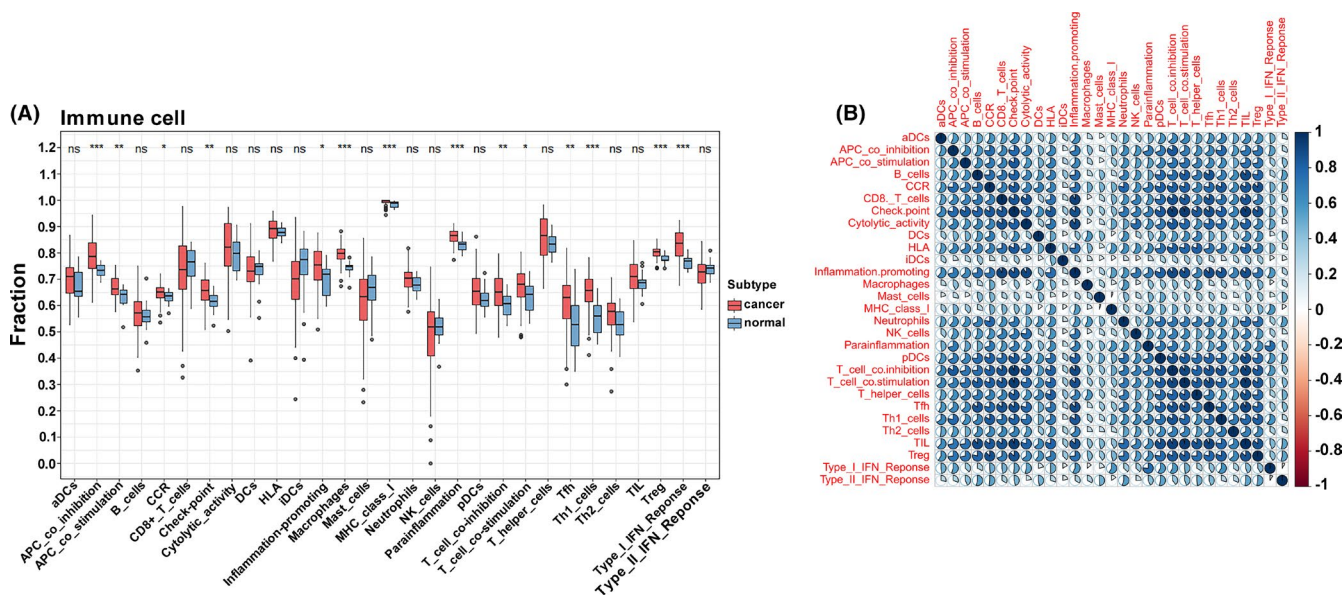

FIGURE 1 The identification of immune gene sets by ssGSEA. (A) The expression of immune gene sets by ssGSEA in cancer and normal tissue. (B) The Pearson correlation of immune cell types in cancer tissue based on ssGSEA
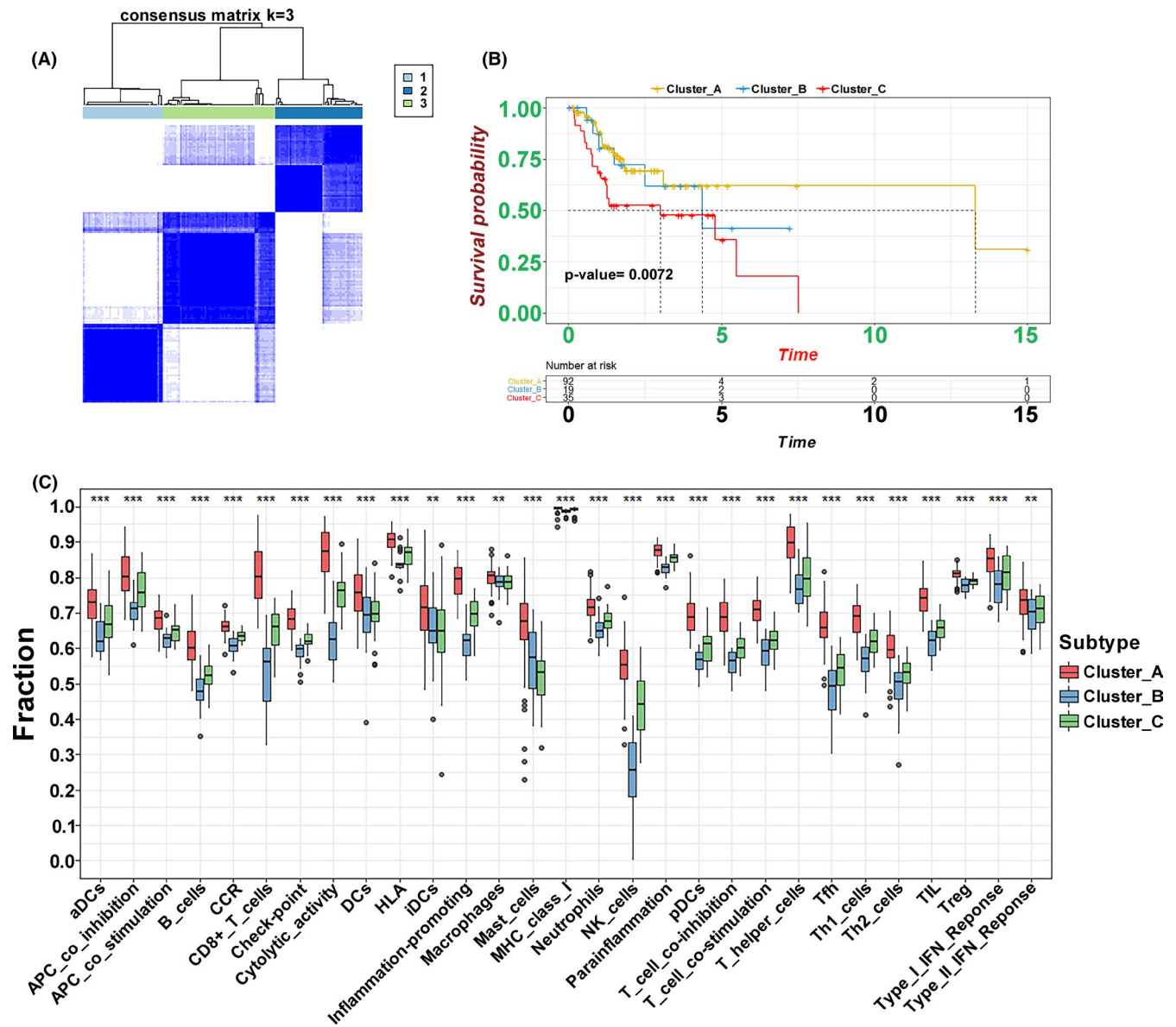
FIGURE 2 The classification of clusters identified by ssGSEA. (A) The Consensus Cluster package-based clustering of patients identified by ssGSEA with $k = 3$. (B) The Kaplan–Meier curves of three clusters regarding OS. (C) The different expression of immune cell sets in the three clusters (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$)

which were upregulated in the relatedly high immune-cell infiltration group (cluster A) (Figure 3A). The box chart also further verified that there the high immune-cell infiltration cluster was significantly positively correlated with the ESTIMATE and immune and stromal scores and negatively correlated with tumor purity (Figure 3B). In addition, we used the CIBERSORT method to quantify the level of immune cell infiltration to carefully evaluate the immune landscape. The results showed that the types of B cells, T cells, and macrophages contributed to immune cell infiltration. We observed that the levels of naïve B cells, CD4 memory-activated T cells, CD8 T cells, and follicular helper T cells were remarkably increased in cluster C

(Figure 3C). These results indicate that this ssGSEA-based clustering strategy can be used to decipher the heterogeneous immune landscape of TSCC, and specific multiple immune-related modulators may have an integrated effect in the mediation of the tumor microenvironment.

3.3 | Interaction of immune-related subgroups with different survivals

To investigate the potential mechanisms involved in the heterogeneity of TSCC, due to the small sample size, we included the samples of two clusters (clusters A and
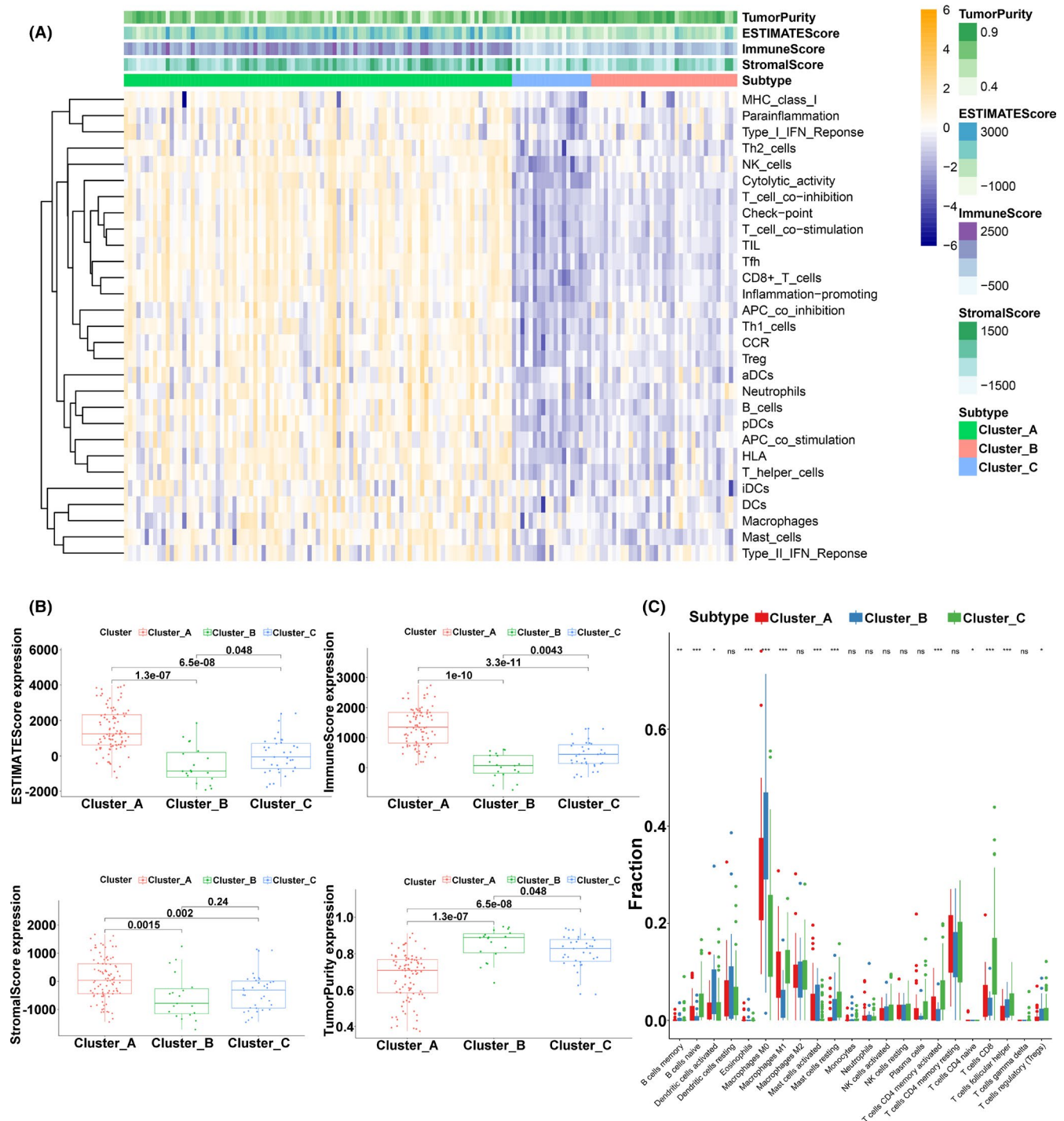
(A)

(B)

(C)

FIGURE 3   Immune landscape of three clusters. (A) The heatmap of immune gene sets identified by ssGSEA and ESTIMATE scores for the three clusters. (B) Different ESTIMATE scores, immune scores, stromal scores, and tumor purity in three clusters. (C) Different distributions of tumor-infiltrating cells in three clusters (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$)

C) in the next analysis. Kaplan–Meier analysis revealed that the OS was significantly different in the two clusters ($p = 0.0065$) (Figure 4A). Based on both clusters, we identified 626 DEGs with the criteria of |log2(fold change)| >2 and adj.$p < 0.05$ in the TCGA-TSCC cohort. Compared with cluster C, cluster A included DEGs associated with T cell activation and regulation of leukocyte activation,

according to the biologic-process GO analysis. Cellular component analysis showed that these DEGs were linked to the side of the membrane and the plasma membrane protein complex, and the molecular function analysis showed that cytokine receptor activity and cytokine binding were crucial (Figure 4B). The KEGG analysis indicated that some immune-related pathways, such as
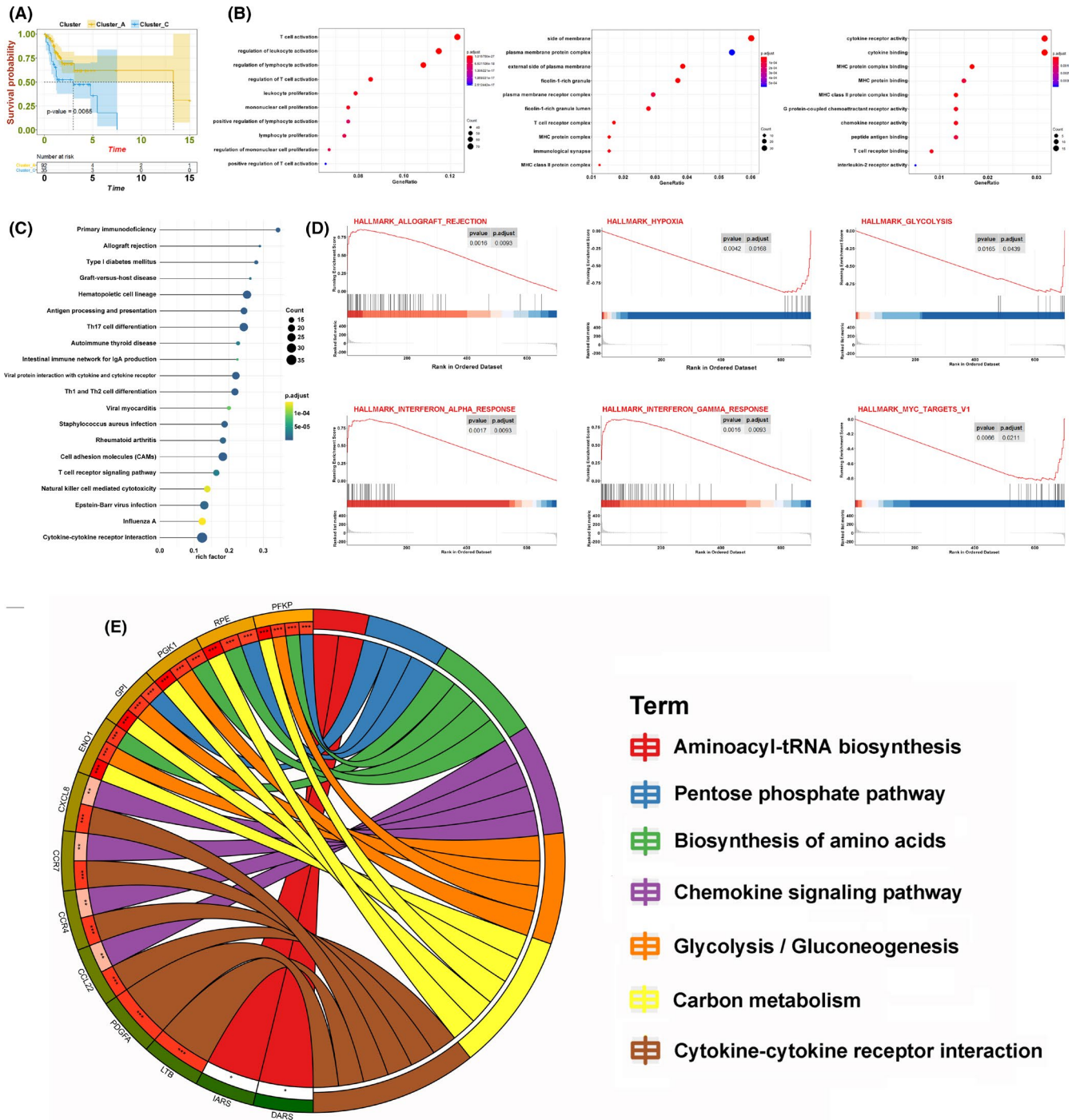
FIGURE 4 Relationship between cluster A and cluster C. (A) The OS Kaplan–Meier curves of cluster A and cluster C. (B) OS analysis of all DEGs from the two clusters. (C) KEGG analysis of all DEGs from the two clusters. (D) GSEA of all DEGs from the two clusters. (E) Circos plot of 41 prognostic mediators from the two clusters

primary immunodeficiency, Th17 cell differentiation, and Th1/Th2 cell differentiation, were enriched in both clusters (Figure 4C). Using GSEA analysis, these signaling pathways, including interferon-gamma response, hypoxia, and glycolysis, were viewed to be core biologic carcinogenic processes involved in the regulation of the immune microenvironment (Figure 4D). Additionally,

we performed univariate Cox analysis to select prognostic regulators from these DEGs, and 41 mRNAs were identified (Table S1). Finally, we continuously iterated the enrichment analysis of these 41 mediators and found that multiple pathways, especially glycolysis, not only played important roles in the regulation of the immune landscape but also altered clinical survival in TSCC (Figure 4E).

3.4 | Construction of the immune risk model

Based on the above findings, we applied Lasso Cox regression analysis to the 41 immune-related mRNAs. Thereafter, we built a risk model with 15 mRNAs, and the coefficients of these genes were used to calculate the risk score as follows (Figure 5A): risk score = CCDC43 × 0.0034 + CCL22 × (−0.0436) + KLHL2 × 0.0429 + SH2D3C × (−0.0052) + ANKRD22 × (−0.0129) + SUN1 × 0.0215 + IARS × 0.0269 + NTMT1 × 0.0069 + IER3 × 0.0042 + CCR7 × (−0.089) + PGK1 × 0.0024 + CTSG × (−0.0038) + GPI × 0.0043 + SECTM1 × 0.0364 + RPE × 0.0119. Patients were divided into low-risk and high-risk groups according to the cutoff value (median risk score). The OS Kaplan-Meier curve showed that the patients in the high-risk group had worse OS than those in the low-risk group ($p = 2.303e^{-8}$) (Figure 5B). The risk coefficient and mortality of samples in the high-risk group were higher than those in the low-risk group (Figure 5C). Univariate and multivariate analyses showed that the risk score can be an independent prognostic biomarker. Unfortunately, after adjusting for all risk factors, only the T stage was slightly associated with OS (Figure 5D). To evaluate the prognostic effectiveness of the risk score, the ROC curve showed that the AUCs of 1-year, 3-year, and 5-year survival were 0.802, 0.832, and 0.819, respectively, suggesting that the risk model was significantly effective and applicable (Figure 5E).

3.5 | The immune landscape and verification of hub genes

To precisely select the core regulators in the immune signature, we utilized Cytoscape based on the STRING website to calculate all degrees of nodes among the 15 regulators and identified the hub genes, _PGK1_, _GPI_, and _RPE_, which had the top three highest degree values (Table S2). To further confirm the key role of these hub genes in the immune mediation process, as shown in Figure 6A, _PGK1_ expression was reevaluated and was found to be significantly highly correlated with clinical stage and tumorigenesis in TSCC. Moreover, the results revealed that the _PGK1_ expression had an evident negative correlation with tumor purity ($p = 0.002$) and a positive correlation with the immune score ($p = 0.002$), stromal score ($p = 0.020$), and ESTIMATE score ($p = 0.002$). Additionally, we noted
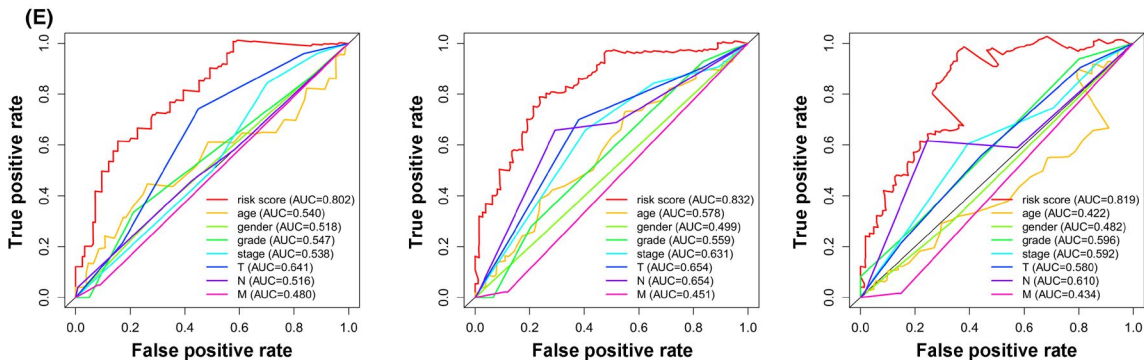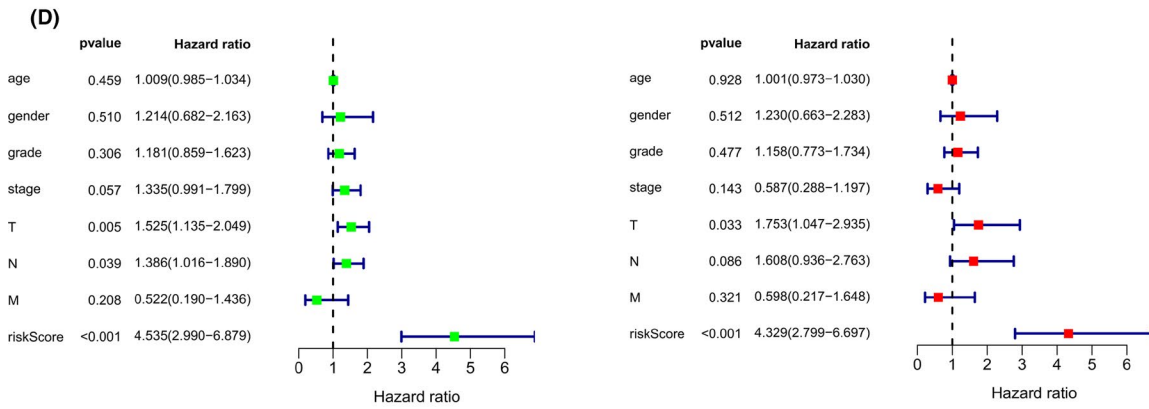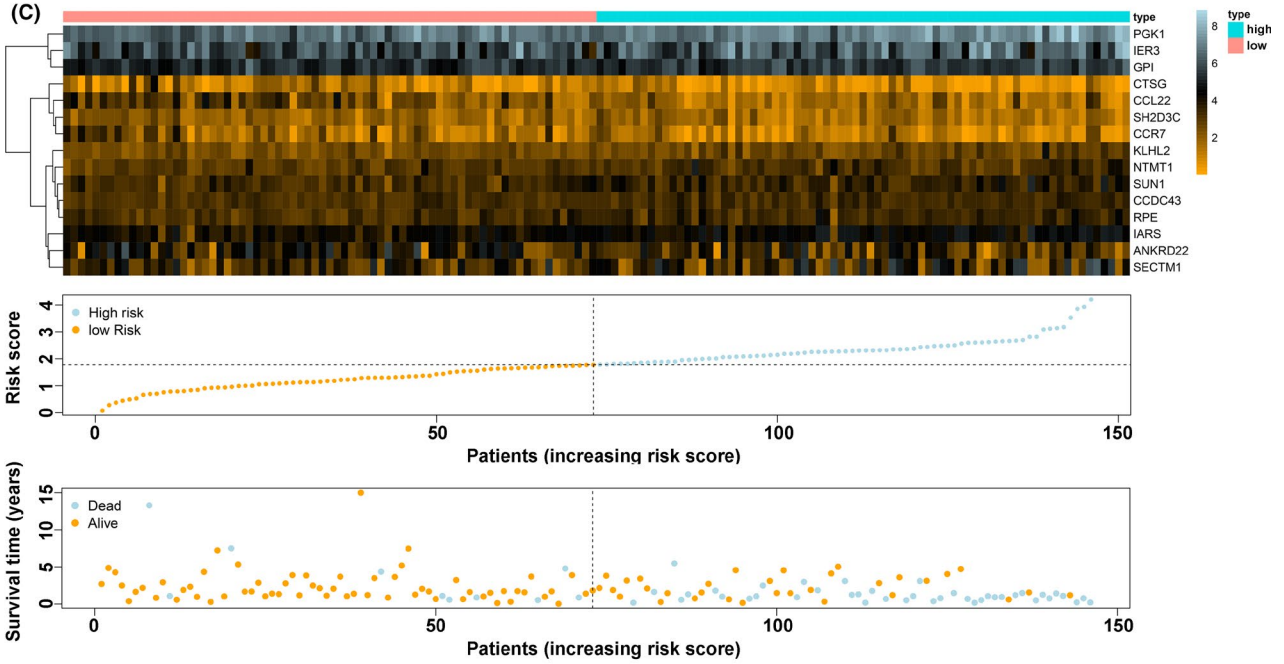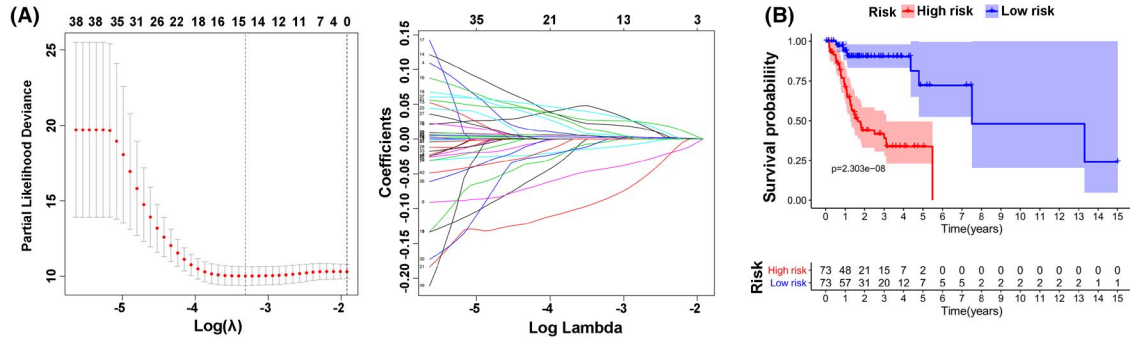
that the expression of _PGK1_ was closely related to that of _PD1_ (also known as _PDCD1_ or _CD279_) (Figure 6B). Six types of immune cells that had an influence on TME, such as CD8 T cells and follicular helper T cells, were verified to be regulated by the hub gene _PGK1_ (Figure 6C). The findings for _GPI_ and _RPE_ are presented in Figure S1. In accordance with the immunohistochemistry results from the Human Protein Atlas database, the expression of the hub genes _PGK1_, _GPI_, and _RPE_ was obviously elevated in TSCC tissues than in normal tissues (Figure 7A).

3.6 | Interaction and pan-cancer analysis of _PGK1_

To further unravel the interaction of _PGK1_ in the immune-related pathway with miRNAs, lncRNAs, and circRNAs, we first identified 13 lncRNAs and 18 miRNAs with |logFC(fold change)| ≥1 and FDR <0.05, and targeted 915 mRNAs (miRTarBase, miRDB). We identified the intersection of these 915 mRNAs and the top 100 _PGK1_-related genes. Finally, we visualized and constructed a ceRNA regulatory network containing nine _PGK1_-related mRNAs, three miRNAs, and 10 lncRNAs (Figure 7B). After determining the crucial molecular relationships of _PGK1_, we explored this oncogene in a pan-cancer analysis. We downloaded all pan-cancer data from the UCSC Cancer Genomics Browser (https://genome-cancer.ucsc.edu). The results indicated that the expression of _PGK1_ was distinctly upregulated in most types of cancers, including bladder urothelial carcinoma (BLCA), head and neck squamous cell carcinoma (HNSCC), and lung adenocarcinoma (LUAD) (Figure 8A). Furthermore, high _PGK1_ expression was correlated with decreased OS ($p < 0.001$), disease-specific survival (DSS) ($p < 0.001$), and progression-free interval (PFI) ($p = 0.004$) (Figure 8B). Similarly, high _PGK1_ expression was positively associated with tumor mutational burden (TMB), a promising indicator to differentiate responders to immune checkpoint inhibitors (Figure 8C). The pan-cancer analysis indicated that _PGK1_ may activate multiple immune genes and interfere with the tumor environment in 32 types of cancer (Figure 8D).

4 | DISCUSSION

Tongue squamous cell carcinoma, characterized by a notably aggressive biologic behavior and heterogeneous

(A)

(B) Risk — High risk — Low risk

(C) type: high, low

(D)

	pvalue	Hazard ratio
age	0.459	1.009(0.985–1.034)
gender	0.510	1.214(0.682–2.163)
grade	0.306	1.181(0.859–1.623)
stage	0.057	1.335(0.991–1.799)
T	0.005	1.525(1.135–2.049)
N	0.039	1.386(1.016–1.890)
M	0.208	0.522(0.190–1.436)
riskScore	<0.001	4.535(2.990–6.879)

	pvalue	Hazard ratio
age	0.928	1.001(0.973–1.030)
gender	0.512	1.230(0.663–2.283)
grade	0.477	1.158(0.773–1.734)
stage	0.143	0.587(0.288–1.197)
T	0.033	1.753(1.047–2.935)
N	0.086	1.608(0.936–2.763)
M	0.321	0.598(0.217–1.648)
riskScore	<0.001	4.329(2.799–6.697)

(E)

risk score (AUC=0.802)
age (AUC=0.540)
gender (AUC=0.518)
grade (AUC=0.547)
stage (AUC=0.538)
T (AUC=0.641)
N (AUC=0.516)
M (AUC=0.480)

risk score (AUC=0.832)
age (AUC=0.578)
gender (AUC=0.499)
grade (AUC=0.559)
stage (AUC=0.631)
T (AUC=0.654)
N (AUC=0.654)
M (AUC=0.451)

risk score (AUC=0.819)
age (AUC=0.422)
gender (AUC=0.482)
grade (AUC=0.596)
stage (AUC=0.592)
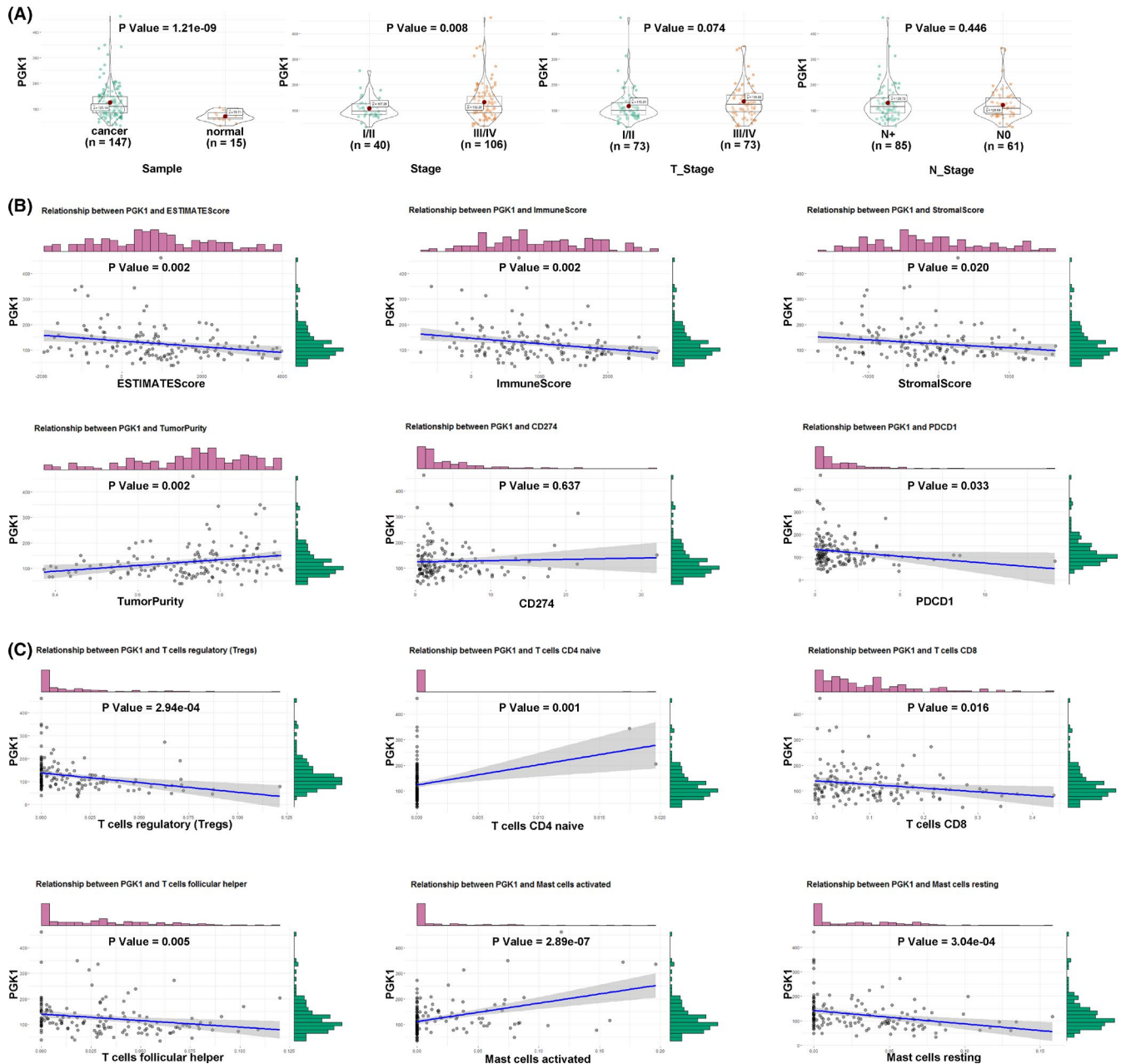T (AUC=0.580)
N (AUC=0.610)
M (AUC=0.434)

FIGURE 6 Comprehensive evaluation of *PGK1*. (A) The relationship between *PGK1* expression and clinical characteristics. (B) The relationship between *PGK1* expression and ESTIMATE scores, PD1 (PDCD1)/PD-L1 (*CD274*). (C) The relationship between *PGK1* expression and tumor-infiltrating cells

survival, is a clinically, gnomically, and immunologically distinct subgroup of head and neck tumors that arise from the base of the tongue.[2] Current treatments for TSCC, including surgical techniques and chemoradiotherapy combined with cetuximab have failed to achieve satisfactory therapeutic effects; thus, immune checkpoint inhibitors are expected to prolong OS in advanced TSCC.[22] It is generally agreed that PD1 or PDL1 acted as major indicators to guide the selection of immune checkpoint inhibitors.[23] However, there are conflicting opinions on these indicators in the literature. Sanni Tervo reported that the higher density and intensity of PD1 were significantly correlated with better survival.[24] Yet Tomofumi Naruse and Naoki Akisada demonstrated that patients with high PD-L1 expression had a remarkably shorter DFS and increased local recurrence.[25,26] Currently, PD1/PDL1 remains inapplicable in the prediction of TSCC prognosis; thus, it is urgent to elucidate the underlying genetic pathogenesis in TSCC and to identify reliable prognostic biomarkers based on the immune landscape.

In this study, we performed an ssGSEA of immune genes from the ImmPort database to calculate an

FIGURE 7 Interaction and verification of *PGK1* in TSCC. (A) The verification of *PGK1*, *GPI*, and *RPE* expression in immunohistochemistry. (B) The interactive network of *PGK1*
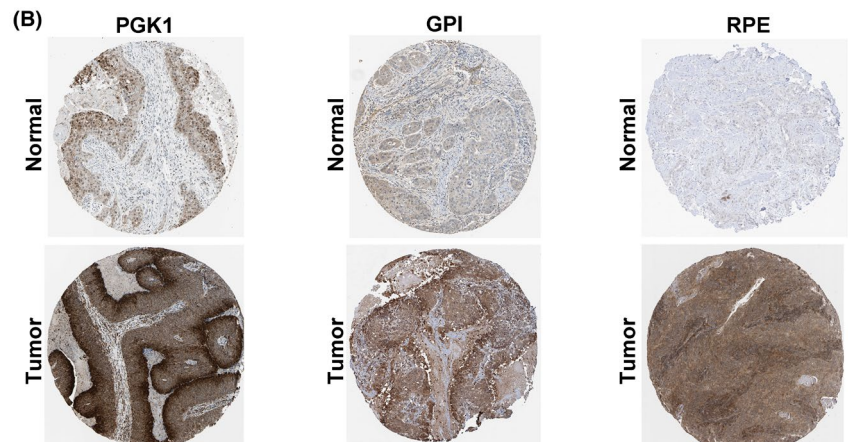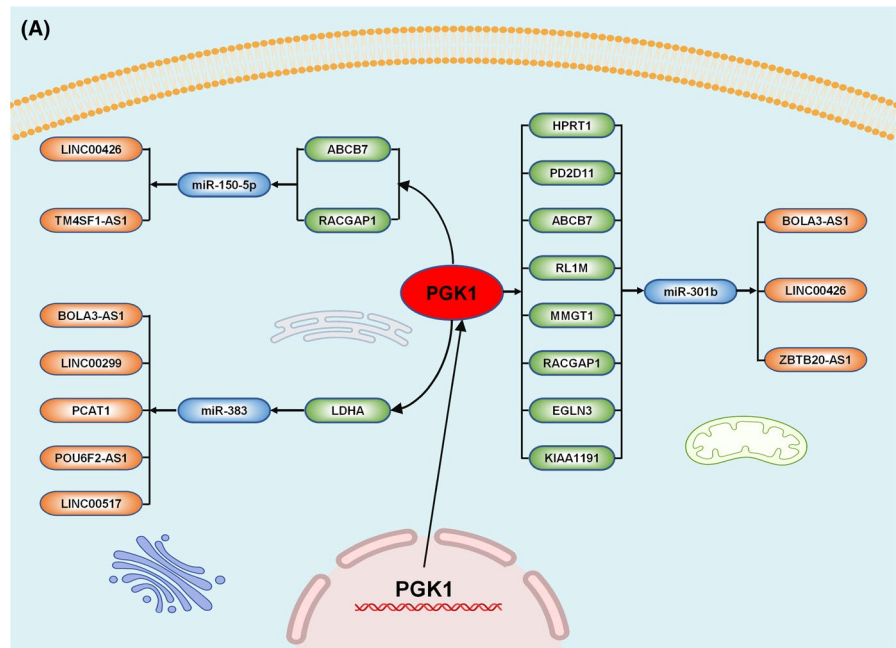
immune-based prognostic score and determine the specific immune infiltration, which is more generic than the normal gene signature. First, we conducted the ESTIMATE and CIBERSORT analyses to find the ssGSEA-based clusters that were highly associated with ESTIMATE scores and immune scores. Furthermore, T cells, CD4 memory-activated T cells, CD8 T cells, and follicular helper T cells were found to be remarkably increased in the cluster with a better prognosis. Many studies have reported that a higher degree of T cell infiltrates, especially CD8+, improved OS in HPV-positive and HPV-negative oral squamous carcinoma,[27] and DFS in patients with respectable laryngeal cancer.[28,29] Moreover, some researchers found that the frequency of PD1+ CD8+ T cells in the tumor microenvironment, which negatively impacts effector and immunosuppressive functions, can predict the clinical efficacy of PD1 inhibitors better than PD1 or PD-L1 expression or TMB.[30] Next, we explored specific mRNAs that could potentially mediate CD8+ T cells and constructed

a prognostic signature with a broad scope of applications that could accurately identify cases with worse survival. The results indicated that the constructed risk model was much better than the current TNM classification system, but its utility needs to be reverified with other TSCC groups.

In our analysis, we identified and selected three hub genes, *PGK1*, *GPI*, and *RPE*, which are involved in the glycolysis pathway. Glycolysis is the metabolism of glucose to lactate despite the presence of adequate oxygen, which is also called the Warburg effect.[31] The Warburg effect has been recognized to be one of the hallmarks of cancer and to provide essential evidence that oncogenes and tumor suppressors alter the regulation of energy metabolism.[32] More recent studies have found that glycolysis and lactate production, in particular, are strongly increased in peripheral T cells.[33] Excessively increased glucose uptake induces T cell activation and proliferation and can lead to increased proliferation and lymphoproliferative disease.[34]
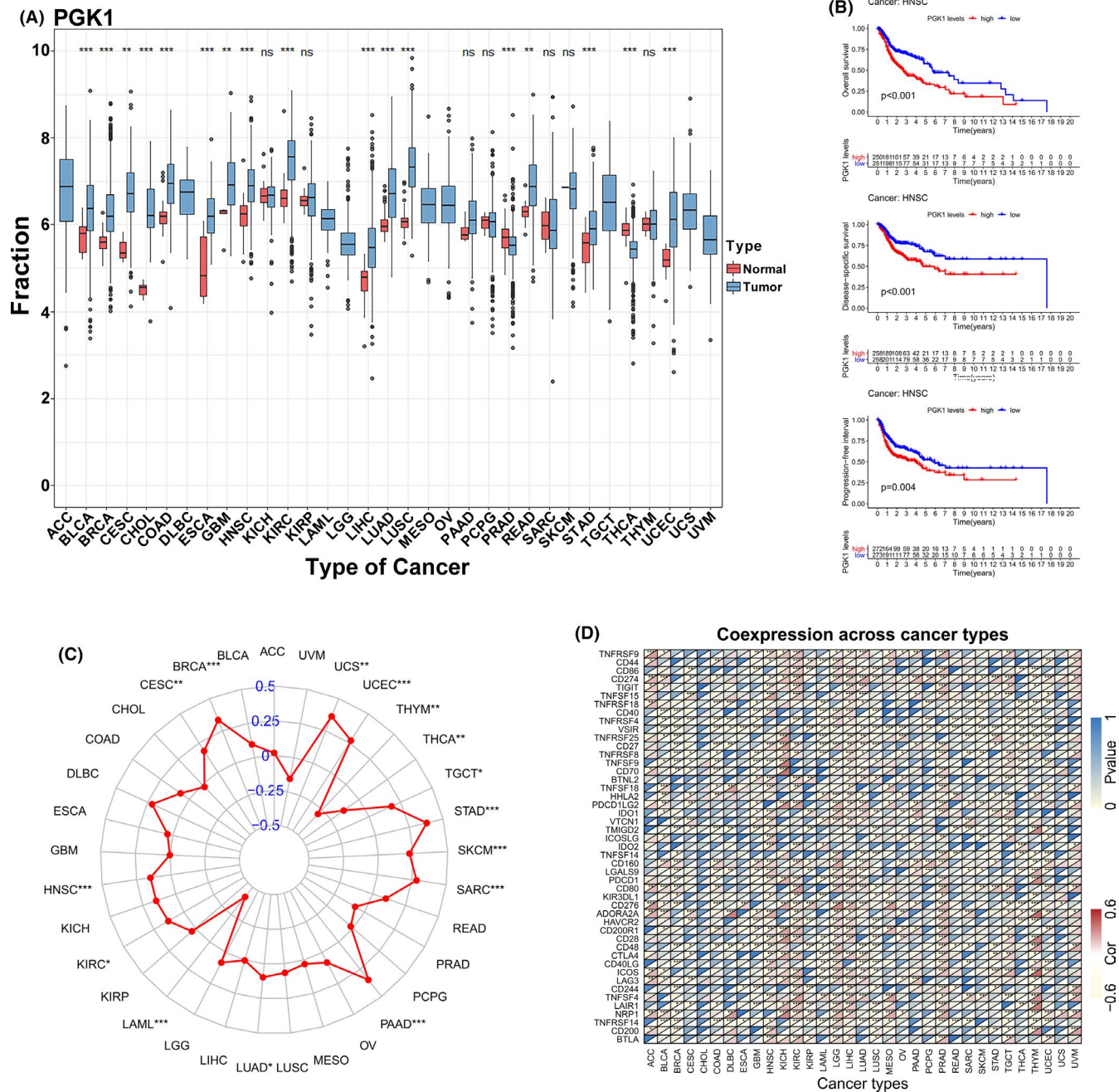
FIGURE 8 Pan-cancer analysis of *PGK1*. (A) The expression of *PGK1* in 32 cancer species. (B) The relationship between *PGK1* expression and OS, DSS, and PFI in HNSCC. (C) The relationship between *PGK1* expression and TMB in 32 cancer species. (D) The relationship between *PGK1* expression and immune-related regulators

Our findings demonstrated a similar relationship between the tumor immune environment and glycolysis in TSCC.

Phosphoglycerate kinase 1 (*PGK1*), an isoform of the PGK family that catalyzes ATP formation in the glycolysis pathway,[35] plays a rate-limiting role in controlling ATP and 3-PG levels[36] and is a mediator in the regulation of autophagy initiation, DNA replication, and repair in mammal cell nuclei.[37-39] More notably, via the pancancer analysis in our study, the mRNA expression level of *PGK1* was significantly associated with the prognosis

of multiple cancers, including head and neck, liver, and breast cancer. *PGK1* has gradually become a novel target in clinical research for various cancers. Researchers have found that enhanced glycolysis activity combined with higher *PGK1* expression in breast cancer was associated with pro-tumor immunity via upregulation of immune/inflammation pathways, especially the IL-17 signaling pathway.[40] However, little is known about the relationship between PGK1 and the tumor immune microenvironment in TSCC. In this study, we first identified the key role of

PGK1 in TSCC and comprehensively evaluated its profile in the immune landscape. Additionally, we constructed an interactive network with *PGK1* at the core and many intersections were verified experimentally. Han RL reported that miR-383 suppressed the expression level of *LDHA*, an mRNA highly correlated with *PGK1*, by directly binding to its 3'-untranslated region and thus silenced aerobic glycolysis in ovarian cancer cells.[41] Similarly, Fang found that overexpression of miR-383 could inhibit cell proliferation and invasion triggered by LDHA in hepatocellular cancer.[42] Furthermore, Li verified that circ_0136666 accelerated the progression of colorectal cancer by directly targeting and downregulating miR-383.[43] Hence, based on these reports, the interactive network constructed in this study was acceptable for the in-depth exploration of in vitro experiments.

Glycosylphosphatidylinositol (*GPI*)-anchored proteins have been regarded as well-established cancer biomarkers that correlate with changes in GPI-T expression. With GPI-T, the importance of these proteins truly came to light only after their discovery in bladder cancer.[44] Several interesting findings have demonstrated the highly intimate relationship between the levels of GPI-T subunits and different cancer types.[45] Ribulose-phosphate 3-epimerase (*RPE*), an important protein involved in carbohydrate degradation can catalyze the reversible epimerization of D-ribulose 5-phosphate to D-xylulose 5-phosphate.[46] However, evidence supporting the oncogenic nature of *RPE*, particularly with respect to tumorigenesis, was minimal. Together, none of the past studies focused on the role of both biomarkers in TSCC; thus, our study is the first to unravel the genomic properties of these biomarkers in the field of immunology. Certainly, our findings in this study may be the tip of an iceberg, and there are some limitations that need to be improved; for instance, the mechanism underlying TSCC progression still has immense potential to be further elucidated, and further clinical trials and molecular experiments are required to verify our results.

5 | CONCLUSION

This study systematically evaluated TSCC patients from the TCGA database and constructed different clusters based on ssGSEA analysis and profiled the immune landscape and immune cell infiltration in the tumor environment. After analyzing the correlation among clusters, we established a signature of 15 immune-related genes that were identified as having independent prognostic significance for TSCC and selected three hub genes, *PGK1*, *GPI*, and *RPE*. Our findings suggest that these immune-related genes may be promising indicators for the mediation of

the immune microenvironment and may provide novel insights into immunotherapy for TSCC.

AUTHOR CONTRIBUTION
KC designed the study. YJ analyzed, interpreted the data, wrote the original draft. ZWW, and DH wrote this manuscript. YXZ and YXC edited and revised the manuscript. All authors have seen and approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID
Yi Jin   https://orcid.org/0000-0002-8521-634X
Zhanwang Wang   https://orcid.org/0000-0002-3655-0704
Yuxing Zhu   https://orcid.org/0000-0002-8757-0676
Xingyu Chen   https://orcid.org/0000-0002-9284-4950
Ke Cao   https://orcid.org/0000-0001-5392-2306

REFERENCES

1. Gillison ML, Chaturvedi AK, Anderson WF, et al. Epidemiology of human papillomavirus-positive head and neck squamous cell carcinoma. *J Clin Oncol*. 2015;33(29):3235-3242.
2. Bello IO, Soini Y, Salo T. Prognostic evaluation of oral tongue cancer: means, markers and perspectives (II). *Oral Oncol*. 2010;46(9):636-643.
3. van Dijk BAC, Brands MT, Geurts SME, et al. Trends in oral cavity cancer incidence, mortality, survival and treatment in the Netherlands. *Int J Cancer*. 2016;139(3):574-583.
4. Bagan J, Sarrion G, Jimenez Y. Oral cancer: clinical features. *Oral Oncol*. 2010;46(6):414-417.
5. Almangush A, Heikkinen I, Mäkitie AA, et al. Prognostic biomarkers for oral tongue squamous cell carcinoma: a systematic review and meta-analysis. *Br J Cancer*. 2017;117(6):856-866.
6. Poeta ML, Manola J, Goldwasser MA, et al. TP53 mutations and survival in squamous-cell carcinoma of the head and neck. *N Engl J Med*. 2007;357(25):2552-2561.
7. Bonner JA, Harari PM, Giralt J, et al. Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. *N Engl J Med*. 2006;354(6):567-578.
8. Bonner JA, Harari PM, Giralt J, et al. Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. *Lancet Oncol*. 2010;11(1):21-28.
9. Bauml J, Seiwert TY, Pfister DG, et al. Pembrolizumab for Platinum- and Cetuximab-refractory head and neck cancer: results from a single-arm. Phase II Study. *J Clin Oncol*. 2017;35(14):1542-1549.

10. Weiss MA, Nguyen DT, Khait I, et al. Two-dimensional NMR and photo-CIDNP studies of the insulin monomer: assignment of aromatic resonances with application to protein folding, structure, and dynamics. *Biochemistry*. 1989;28(25):9855-9873.

11. Cohen EEW, Soulières D, Le Tourneau C, et al. Pembrolizumab versus methotrexate, docetaxel, or cetuximab for recurrent or metastatic head-and-neck squamous cell carcinoma (KEYNOTE-040): a randomised, open-label, phase 3 study. *Lancet*. 2019;393(10167):156-167.

12. Wang Q, Hu B, Hu X, et al. Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer Cell*. 2017;32(1):42-56.

13. Gonzalez H, Hagerling C, Werb Z. Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes Dev*. 2018;32(19–20):1267-1284.

14. Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016;44(8):e71.

15. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7.

16. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.

17. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-457.

18. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-140.

19. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.

20. Tibshirani R, Bien J, Friedman J, et al. Strong rules for discarding predictors in lasso-type problems. *J R Stat Soc Series B Stat Methodol*. 2012;74(2):245-266.

21. Chin C-H, Chen S-H, Wu H-H, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. 2014;8(Suppl 4):S11.

22. Burtness B, Harrington KJ, Greil R, et al. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. *Lancet*. 2019;394(10212):1915-1928.

23. Seliger B. Basis of PD1/PD-L1 Therapies. *J Clin Med*. 2019;8(12):2168.

24. Tervo S, Seppälä M, Rautiainen M, et al. The expression and prognostic relevance of programmed cell death protein 1 in tongue squamous cell carcinoma. *APMIS*. 2020;128(12):626-636.

25. Naruse T, Yanamoto S, Okuyama K, et al. Immunohistochemical Study of PD-1/PD-L1 axis expression in oral tongue squamous cell carcinomas: effect of neoadjuvant chemotherapy on local recurrence. *Pathol Oncol Res*. 2020;26(2):735-742.

26. Akisada N, Nishimoto K, Takao S, et al. PD-L1 expression in tongue squamous cell carcinoma. *Med Mol Morphol*. 2021;54(1):52-59.

27. Badoual C, Hans S, Merillon N, et al. PD-1-expressing tumor-infiltrating T cells are a favorable prognostic biomarker in HPV-associated head and neck cancer. *Cancer Res*. 2013;73(1):128-138.

28. Wang J, Wang S, Song X, et al. The prognostic value of systemic and local inflammation in patients with laryngeal squamous cell carcinoma. *Onco Targets Ther*. 2016;9:7177-7185.

29. Vassilakopoulou M, Avgeris M, Velcheti V, et al. Evaluation of PD-L1 expression and associated tumor-infiltrating lymphocytes in laryngeal squamous cell carcinoma. *Clin Cancer Res*. 2016;22(3):704-713.

30. Kumagai S, Togashi Y, Kamada T, et al. The PD-1 expression balance between effector and regulatory T cells predicts the clinical efficacy of PD-1 blockade therapies. *Nat Immunol*. 2020;21(11):1346-1358.

31. Koppenol WH, Bounds PL, Dang CV. Otto Warburg's contributions to current concepts of cancer metabolism. *Nat Rev Cancer*. 2011;11(5):325-337.

32. Ganapathy-Kanniappan S, Geschwind JF. Tumor glycolysis as a target for cancer therapy: progress and prospects. *Mol Cancer*. 2013;12:152.

33. Bental M, Deutsch C. Metabolic changes in activated T cells: an NMR study of human peripheral blood lymphocytes. *Magn Reson Med*. 1993;29(3):317-326.

34. Michalek RD, Gerriets VA, Jacobs SR, et al. Cutting edge: distinct glycolytic and lipid oxidative metabolic programs are essential for effector and regulatory CD4+ T cell subsets. *J Immunol*. 2011;186(6):3299-3303.

35. Hu H, Zhu W, Qin J, et al. Acetylation of PGK1 promotes liver cancer cell proliferation and tumorigenesis. *Hepatology*. 2017;65(2):515-528.

36. Wang S, Jiang B, Zhang T, et al. Insulin and mTOR pathway regulate HDAC3-mediated deacetylation and activation of PGK1. *PLoS Biol*. 2015;13(9):e1002243.

37. Qian XU, Li X, Cai Q, et al. Phosphoglycerate Kinase 1 phosphorylates Beclin1 to induce autophagy. *Mol Cell*. 2017;65(5):917-931.

38. Cai Q, Wang S, Jin L, et al. Long non-coding RNA GBCDRlnc1 induces chemoresistance of gallbladder cancer cells by activating autophagy. *Mol Cancer*. 2019;18(1):82.

39. Qian X, Li X, Lu Z. Protein kinase activity of the glycolytic enzyme PGK1 regulates autophagy to promote tumorigenesis. *Autophagy*. 2017;13(7):1246-1247.

40. Li W, Xu M, Li YU, et al. Comprehensive analysis of the association between tumor glycolysis and immune/inflammation function in breast cancer. *J Transl Med*. 2020;18(1):92.

41. Han R L, Wang F P, Zhang P A, et al. miR-383 inhibits ovarian cancer cell proliferation, invasion and aerobic glycolysis by targeting LDHA. *Neoplasma*. 2017;64(2):244-252.

42. Fang Z, He L, Jia H, et al. The miR-383-LDHA axis regulates cell proliferation, invasion and glycolysis in hepatocellular cancer. *Iran J Basic Med Sci*. 2017;20(2):187-192.

43. Li Y, Zang H, Zhang X, et al. circ_0136666 Facilitates the Progression of Colorectal Cancer via miR-383/CREB1 Axis. *Cancer Manag Res*. 2020;12:6795-6806.

44. Montie JE. CDC91L1 (PIG-U) is a newly discovered oncogene in human bladder cancer. *J Urol*. 2005;174(3):869-870.

45. Zhao P, Nairn AV, Hester S, et al. Proteomic identification of glycosyl- phosphatidylinositol anchor-dependent membrane proteins elevated in breast carcinoma. *J Biol Chem*. 2012;287(30):25230-25240.

46. Liang W, Ouyang S, Shaw N, et al. Conversion of D-ribulose 5-phosphate to D-xylulose 5-phosphate: new insights from structural and biochemical studies on human RPE. *FASEB J.* 2011;25(2):497-504.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.