8-Gene signature related to CD8+ T cell infiltration by integrating single-cell and bulk RNA-sequencing in head and neck squamous cell carcinoma

Shoujing Zhang1, Wenyi Zhang2 and Jian Zhang1*

1Department of Oral and Maxillofacial Surgery, Tianjin Medical University School and Hospital of Stomatology, Tianjin, China, 2Department of Prosthodontics, Tianjin Medical University School and Hospital of Stomatology, Tianjin, China

Background: CD8+ T cells, a critical component of the tumor immune microenvironment, have become a key target of cancer immunotherapy. Considering the deficiency of robust biomarkers for head and neck squamous cell carcinoma (HNSCC), this study aimed at establishing a molecular signature associated with CD8+ T cell infiltration.

Methods: Single-cell RNA sequencing data retrieved from the Gene Expression Omnibus (GEO) database was analyzed to obtain the different cell types. Next, the cell proportions were investigated through deconvolution of RNA sequencing in the Cancer Genome Atlas (TCGA) database, and then the immune-related genes (IRGs) were identified by weighted gene co-expression network analysis (WGCNA). LASSO-Cox analysis was employed to establish a gene signature, followed by validation using a GEO dataset. Finally, the molecular and immunological properties, and drug responses between two subgroups were explored by applying “CIBERSORT”, “ESTIMATE”, and single sample gene set enrichment analysis (ssGSEA) methods.

Results: A total of 215 differentially expressed IRGs were identified, of which 45 were associated with the overall survival of HNSCC. A risk model was then established based on eight genes, including DEFB1, AICDA, TYK2, CCR7, SCARB1, ULBP2, STC2, and LGR5. The low-risk group presented higher infiltration of memory activated CD4+ T cells, CD8+ T cells, and plasma cells, as well as a higher immune score, suggesting that they could benefit more from immunotherapy. On the other hand, the high-risk group showed higher abundance of activated mast cells and M2 macrophages, as well as a lower immune score.

Conclusion: It was evident that the 8-gene signature could accurately predict HNSCC prognosis and thus it may serve as an index for clinical treatment.

KEYWORDS
CD8+ T cells, head and neck squamous cell carcinoma, immunotherapy, prognosis, weighted gene co-expression network analysis
Introduction

Head and neck squamous cell carcinoma (HNSCC) is the seventh most common malignancy worldwide (Siegel et al., 2020). Despite the effective and aggressive treatment strategies involving surgery combined with radio- and chemotherapy, patients with advanced stage HNSCC only have a 50% five-year survival rate (Vigneswaran and Williams, 2014). In recent years, immunotherapy involving checkpoint inhibitors blocking cell death protein 1 (PD-1) or programmed death ligand-1 (PD-L1) has been approved for clinical use, with preliminary results showing that the strategy significantly improves the overall survival of recurrent or metastatic HNSCC patients. However, several clinical trials have demonstrated that anti-PD-1/PD-L1 therapy is only beneficial to a few patients (Ferris et al., 2016; Siu et al., 2019). Studies have suggested that CD8+ T lymphocytes substantially express PD-1 and may play an important role in the efficacy of immunotherapy (Jia et al., 2020). It is worth noting that high dense infiltration of CD8+ T cells in HNSCC patients is generally associated with a good prognosis (Fridman et al., 2017). Moreover, PD-1+ CD8+ T cells showed excellent antitumor effect in an anti-PD1-resistant murine HNSCC model (Xu et al., 2020). Therefore, there is an urgent need to explore the molecular mechanisms associated with CD8+ T cells infiltration.

Single-cell RNA sequencing (scRNA-seq) has been the subject of rapid technological developments in the last decade, thereby resulting in significant improvements in describing and defining the tumor heterogeneity at a single-cell level (Qi et al., 2019). Besides, application of scRNA-seq to characterize the tumor microenvironment (TME) may provide valuable insights into immune landscapes and even effective immunotherapy strategies (Kurten et al., 2021). Similarly, the gene signature identified based on immune molecular characteristics might be a strong predictor of clinical outcome and immunotherapy response (Song et al., 2022). However, the predictive potential of the molecular mechanisms describing immunophenotypic features in HNSCC have not yet been elucidated.

This study explored the mechanism associated with infiltration of CD8+ T cells through integrating bulk and scRNA sequencing. Specifically, a LASSO-Cox regression risk model was built and verified based on the hub immune-related genes (IRGs) identified by weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008). Next, we comprehensively represented the various immune features of an 8-gene signature using "ESTIMATE" (Yoshihara et al., 2013), "CIBERSORT" (Newman et al., 2015), single sample gene set enrichment analysis (ssGSEA) approaches, and immunophenoscore (IPS) data. It is expected that the identified risk score will not only be used as an efficient indicator for HNSCC prognosis, but also as a potential therapeutic target.

Materials and methods

The study design is illustrated using a flow diagram (Figure 1).

Data acquisition

The single cell RNA-sequencing profile of GSE103322 dataset (Puram et al., 2017), comprising 5,902 single cells of 18 patients, was downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) (accessed date 13 October 2021). HNSCC RNA-sequencing, clinical and mutation data were downloaded from The Cancer Genome Atlas (TCGA) database using the GDC Data Portal (https://portal.gdc.cancer.gov/) (accessed date 13 October 2021). The Fragments per Kilobase per Million (FPKM) values were first converted to transcripts per million kilobase
(TPM) values. To validate the prognostic power of the model, the transcriptome and clinical files of the GSE65858 dataset, containing 270 HNSCC samples, were obtained from the GEO database (Wichmann et al., 2015).

Notably, a total of 2,720 IRGs were obtained from the ImmPort (https://www.immport.org/home) and InnateDB (https://www.innatedb.com/) databases (accessed date 13 October 2021).
Processing of single-cell and bulk RNA-seq files

The “Seurat” (version 4.1.1) package in R (version 4.1.2) was applied to group 5,902 cells into appropriate clusters, with the resolution set to 0.8. Results were presented by employing the T-distributed stochastic neighbor embedding (t-SNE) for dimension reduction. Next, diverse cell types, B/plasma cells, endothelial cells, regulatory T cells (Treg cells), mast cells, CD8+ T cells, epithelial cells, dendritic cells, macrophages, fibroblasts, and CD4+ T cells were identified based on their specific markers. The “Cellchat” (version 1.1.3) package was used to analyze the cell–cell communication, and then deconvolution was performed using the “BisqueRNA”
(version 1.0.5) method (Jew et al., 2020) to calculate the cells fractions of TCGA bulk profiles. Based on the TCGA RNA-seq profiles, differentially expressed genes (DEGs) were identified with FDR < 0.05 and |log2FC| > 1 set as the cutoff values.

**Determination of immune-related candidate genes**

The differential IRGs were determined by overlapping DEGs and IRGs, and then used to screen the hub genes by WGCNA (version 1.7.0). First, Pearson correlation coefficient was determined for every gene, and a suitable soft threshold $\beta$ was automatically selected through the pick Soft Threshold function. Next, gene expression similarity matrix was transformed into an adjacency matrix using a network type of signed and soft powers $\beta = 3$, followed by employing TOM (topological overlap measure) to cluster genes into network modules. The 1-TOM (dissimilarity TOM) was then applied as the input for hierarchical clustering and the “DynamicTreeCut” algorithm was employed to detect modules (clusters of highly interconnected genes) as branches of the dendrogram. Finally, we identified and selected a module (215 genes) that significantly correlated with CD8+ T cells content. Kaplan-Meier (KM) survival and univariate Cox analysis were utilized to determine the hub genes associated with survival at a threshold of $p < 0.05$.

**Development of a prognostic signature in TCGA ($n = 498$)**

LASSO-Cox analysis was performed using “glmnet” package to determine the optimal prognostic gene set. The risk score of each HNSCC patient was determined as the sum of normalized gene expression values weighted by their LASSO-Cox coefficients in accordance with the following formula:

$$\text{risk score} = \sum_i \text{Coef}_i \cdot \text{Exp}_i$$

Where $\text{Coef}_i$ indicates the calculated regression coefficient of each gene in the LASSO-Cox model and $\text{Exp}_i$ represents the mRNA expression value. Kaplan-Meier (KM) analysis, receiver operating characteristic (ROC) curves, and univariate and multivariate Cox regression analyses were employed to validate the independent prognostic factors in TCGA-HNSC and GSE65858 datasets. For better clinical prediction of HNSCC patient survival probabilities, a nomogram was constructed using the “rms” R package based on multivariate Cox analysis results. The concordance index (C-index) of the nomogram was calculated to assess the discriminative ability.

**Immune features and therapy prediction in distinct risk groups**

“CIBERSORT” (version 1.03) and “ESTIMATE” (version 1.0.13) analyses were applied to determine the abundance of 22 immune cells and immune infiltration scores. The ssGSEA approach was employed via the “GSVA” (version 1.42.0) package to compute the enrichment scores of 29 immune features (Hänzelmann et al., 2013). To predict the susceptibility of eight common chemotherapeutic drugs (5-Fluorouracil, bleomycin, cetuximab, cisplatin, docetaxel, methotrexate, rapamycin, and sunitinib) for HNSCC, the “pRRophetic” (version 0.5) method was performed to evaluate the half-maximal inhibitory concentration (IC50) of patients in distinct groups (Gelehrer et al., 2014). The immunophenoscore (IPS) of HNSCC patients, which is a scoring scheme that characterizes the determinants of tumor immunogenicity (Charoentong et al., 2017), were downloaded from The Cancer Imaging Archive (TCIA) database (https://tcia.at/home, accessed date 15 November 2021). To predict the anti-CTLA4 and anti-PD1 responses, patients with different IPS were further compared between the two risk groups. Finally, the “Maftools” (version 2.10.05) (Mayakonda et al., 2018) package was used to determine the tumor mutational burden (TMB) and identify the driver genes.

**Enrichment analysis**

The reference gene sets of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (c2.cp.kegg.v7.5.1. symbols.gmt) were obtained from the MSigDB database (https://www.gsea-msigdb.org/gsea/msigdb, accessed date 15 November 2021). GSEA software (version 4.2.3) and Gene Set Variation Analysis (GSVA) were conducted to determine the KEGG pathways with FDR < 0.05.

**Results**

**Cell typing in head and neck squamous cell carcinoma scRNA-seq and deconvolution in the Cancer Genome Atlas-HNSC**

We first collected the Smart-seq2 profile data of 5,902 cells in the GSE103322 dataset. Principal component analysis (PCA) and t-SNE analysis identified 27 cell clusters (Figure 2A). According to expressions of marker genes, 10 distinct cell clusters were identified, including CD8+ T cells, macrophages, CD4+ T cells, fibroblasts, endothelial cells, B/plasma cells, mast cells, Treg cells, epithelial cells, and dendritic cells (Figures 2B, 3A). GSVA results showed that “MYC_TARGETS_V2” and “MYC_TARGETS_V1”
were activated in epithelial cells, whereas “HYPOXIA” was abundant in fibroblasts (Figure 3B). Results obtained after applying the “CellChat” method showed that there was a strong connectivity between different cell types (Figures 3C,D). Next, the BisqueRNA approach was performed to calculate proportions of the 10 cell types by deconvoluting the TCGA bulk profiles. Supplementary Table S1 shows proportion of the 10 cell types in 497 samples. Survival analysis demonstrated that mast cells ($p = 0.001$), CD8$^+$ T cells ($p = 0.011$), and Treg cells ($p = 0.002$) were significantly associated with HNSCC outcome (Figures 2C–E). Moreover, univariate Cox analysis indicated that Treg cells were associated with good outcome ($p = 0.018$), whereas mast cells were intimately linked to poor prognosis ($p = 0.019$) (Figure 2F).

Construction and validation of a gene risk signature associated with CD8$^+$ T cells

First, 9,244 DEGs were obtained from the TCGA-HNSC dataset comprising 501 tumor and 44 normal samples (Figure 4A). Subsequently, 2,720 IRGs from ImmPort and InnateDB databases were matched with DEGs, from which 840 differentially expressed IRGs were obtained for further analysis (Figure 4B). Based on the 840 IRGs and proportions of the 10 cell types in TCGA, the weighted gene co-expression network was generated using the soft-thresholding power $\beta = 3$, which resulted in identification of 10 modules (Figures 5A,B). To further explore the features of CD8$^+$ T cells infiltration, we selected the turquoise module (215 genes) which had the strongest correlation with CD8$^+$ T cells ($r = 0.86, p = 1e-17$). Univariate Cox analysis demonstrated that 45 of the 215 hub genes were closely associated with HNSCC survival (Figure 5C). Therefore, the 45 genes were subjected to LASSO regression analysis to identify the optimal penalty coefficient (Figures 5D,E). The survival analysis identified eight genes, including $DEFB1$, $AICDA$, $TYK2$, $CCR7$, $SCARB1$, $ULBP2$, $STC2$, and $LGR5$, which were significantly associated with HNSCC prognosis (Figures 6A–H). The eight risk regression coefficients were then employed to compute individual risk score of HNSCC patients according to the following formula:

$$\text{Risk score} = (-0.097) \cdot \text{DEFB1} + (-0.444) \cdot \text{AICDA} + (-0.175) \cdot \text{TYK2} + (-0.071) \cdot \text{CCR7} + 0.020 \cdot \text{SCARB1} + 0.079 \cdot \text{ULBP2} + 0.161 \cdot \text{STC2} + (-0.128) \cdot \text{LGR5}$$

Next, the 498 HNSCC patients were stratified into high- and low-risk groups based on the median risk score. KM survival analysis results indicated that the high-risk group patients showed poorer outcomes compared to the low-risk group ($p < 0.001$, Figure 6I). Consistently, similar results were observed in the GSE65858 dataset ($p = 0.005$, Figure 6J).

Validation in the Cancer Genome Atlas-HNSC and GSE65858 cohorts, and scRNA-seq data

The risk score, survival status distributions of HNSCC patients, and correlation analysis are displayed in Figure 7. Results demonstrated that survival reduced with rising risk score, and there was a significant correlation between risk
score and survival in TCGA cohort ($r = -0.2$, $p = 6.4e^{-06}$). Time-dependent ROC and calibration curves at one-, three-, and five-years were then constructed (Figures 8A,B). In the TCGA cohort, the areas under the ROC curves (AUCs) were 0.679, 0.703, and 0.644 for 1-, 3-, and 5-years survival, respectively. In both the TCGA and GSE65858 cohorts, univariate and multivariate Cox analyses demonstrated that the risk score was an independent predictor for prognosis (Figures 8C–F). To determine the cells that these eight genes were enriched, the distribution plots for expressions of the eight genes in the 10 cell types identified in the GSE103322 dataset were generated and are shown in Figures 9A–I. Results showed that the expression levels of DEFBI and ULBP2 were higher in epithelial cells, whereas TYK2 and CCR7 levels were abundant in dendritic cells. In addition, the endothelial cells had higher expressions of SCARB1 and STC2, and LGR5 was highly expressed in both dendritic cells and
Based on proportions of the 10 cell types obtained after deconvolution, correlation analysis was performed to evaluate the association among proportion of CD8+ T cells and risk score. Obtained results revealed that fractions of CD8+ T cells declined as the risk score increased (r = −0.41, p < 2.2 e-16, Figures 9J,K).

**Construction of a nomogram for clinical practice**

A heatmap was generated to depict the changes in expression of the eight genes between different clinical subgroups (Figure 10A). The performance of the risk score was then explored in different clinicopathological subgroups, including clinical stage (stage I-III and stage IV), age (<60 and >60), grade (G1-2 and G3-4), T stage (T0-2 and T3-4), N stage (N0-1 and N2-3), and gender (female and male). According to the survival analysis results, HNSCC patients with high-risk scores consistently had a poorer outcome in all subgroups (Figures 10B–G). Next, the three remarkable variables in the multivariate analysis, including age, N stage, and risk score, were selected and used to build a nomogram (C-index: 0.676) for estimating the 1-, 3-, and 5-year survival rate (Figure 11A). By drawing a vertical line to the axis points, we could estimate patient survival based on total points. Overall, the calibration curves and the AUCs (1-, 3-, and 5-year: 0.733, 0.749, and 0.691, respectively) suggested that the risk model could accurately predict the HNSCC survival rate (Figures 11B,C).
The immune landscape of the two risk groups

To elucidate the biological characteristics activated in distinct risk groups, KEGG pathway enrichment analysis was performed using GSVA and GSEA methods. By setting the adjusted p value (FDR) < 0.05, a total of 51 and 16 pathways were obtained in GSVA and GSEA, respectively (Figures 12A,B). Several overlapping immunoregulatory processes were enhanced in the low-risk group, including "hematopoietic cell lineage", "T cell receptor signaling pathway", "antigen processing and presentation" and "natural killer cell-mediated cytotoxicity". To describe the patterns of immune infiltrations, CIBERSORT and ESTIMATE methods were implemented for calculating the cell fractions and immune-related scores of HNSCC samples (Figures 13A,B). The low-risk group showed more significant infiltrations of CD8+ T cells, M1 macrophages, follicular helper T cells, plasma cells, regulatory T cells, and memory activated CD4+ T cells, as well as a higher immune score. With regard to the high-risk group, abundant infiltrations of activated mast cells, M2 macrophages, resting NK cells, and low immune score were observed. The ssGSEA approach was then applied to estimate the scores of specific immune functions and cells. Results revealed significant differences of most immune cells and functions between high- and low-risk groups (Figure 13C). Besides, 15 immune checkpoint molecules (IFNG, GZMB, HAVCR2, CD274, CD8A, PDCD1,
TBX2, IDO1, GZMA, LAG3, CXCL10, CTLA4, PRF1, CXCL9, and TNF) were selected and their expressions were compared between the two risk groups (Figure 13D). Based on the correlation analysis results, it was evident that the expressions of CD274 and CTLA4 in the two groups were significantly different (CD274: \( p = 0.0006 \); CTLA4: \( p = 2.5 \times 10^{-14} \)), and decreased as the risk score rose (CD274: \( r = -0.18, p = 3.6 \times 10^{-5} \); CTLA4: \( r = -0.43, p < 2.2 \times 10^{-16} \)) (Figures 13E–H). Next, the pRRophetic algorithm was applied to predict the IC50 of eight common chemotherapeutic drugs between the two groups. Patients with a high-risk score showed an increased susceptibility to bleomycin (\( p = 0.00014 \)), cisplatin (\( p = 3.2 \times 10^{-5} \)), and methotrexate (\( p = 0.039 \)). On the other hand, low-risk group patients showed increased sensitivity to rapamycin (\( p = 5.6 \times 10^{-6} \)) (Figures 14A–H). To forecast the response to anti-PD1 and anti-CTLA4 immunotherapy, the IPS scores of HNSCC patients were used to compare the two risk groups (Figure 14I–L). Results indicated that patients in the low-risk group exhibited higher IPS scores and showed greater response to anti-PD1 therapy and anti-PD1 plus anti-CTLA4 therapy (ips_ctla4_neg_pd1_pos: \( p = 0.0054 \), ips_ctla4_pos_pd1_pos: \( p = 1.6 \times 10^{-5} \)) relative to patients in the high-risk group. Given the important role of TMB in prognosis, the intrinsic connection
between TMB and risk score was explored to assess genetic signature. It was found that the high-risk group exhibited higher TMB (Figure 15A). A significant correlation was observed between TMB and risk score ($r = 0.22$, $p = 1.3e-06$, Figure 15B). Survival curve suggested that a low TMB/low risk group showed a great outcome compared with the other groups ($p < 0.001$, Figure 15C). The top 20 driver genes with the highest alteration frequency were analyzed (Figures 15D,E) and four genes ($TP53$, $PKHD1L1$, $DNAH9$, $FAT1$) were significantly different between high- and low-risk groups (Supplementary Table S2).
FIGURE 10
The relationship between risk signature and the clinical characteristics. (A) The heatmap depicting eight gene expressions among distinct clinical patterns. (B–G) Kaplan-Meier survival analysis according to the 8-gene signature stratified by clinicopathological factors. (B) Age < 60: p = 0.011, age ≥ 60: p < 0.001. (C) Female: p = 0.001, Male: p < 0.001. (D) G1–2: p = 0.002, G3–4: p < 0.001. (E) N0–1: p < 0.001, N2–3: p = 0.009. (F) Stage I–III: p = 0.004, Stage IV: p = 0.001. (G) T0–2: p = 0.006, T3–4: p < 0.001.
Immunotherapy has been successful used to treat cancer patients in the advanced tumor stage. Nevertheless, clinical application of the strategy is hampered by several limitations, including low response rates, development of serious side effects, and drug resistance (Sacco et al., 2021). One of the key reasons for these limitations is the paucity of potential predictive markers. In the present study, we calculated the proportion of CD8+ T cells, and selected IRGs-related to CD8+ T cells in filtration by integrating scRNA and bulk sequencing profiles. As a result, 215 differential IRGs were identified by WGCNA, of which 45 genes were significantly associated with HNSCC survival. Subsequently, we developed and validated an 8-gene risk model which may be useful for predicting prognosis and immunotherapeutic effect.

The eight critical genes, including DEFB1, AICDA, TYK2, CCR7, SCARB1, ULBP2, STC2, and LGR5, play essential roles in tumor progression and immune-modulatory effects. For example, DEFB1, the human antimicrobial peptide defensin β 1, is considered as a potential tumor suppressor gene and has been shown to mediate PI3K/mTOR signaling, thereby leading to death of tumor cells (Sun et al., 2006; Lee et al., 2015). DEFB1 was also found to be theoretically useful as a prognostic biomarker for HNSCC (Han et al., 2014). Moreover, DEFB1 was commonly detected in epithelial cells, which is consistent with our results. UL16-binding protein 2 (ULBP2), a ligand of the activating NK cell receptor NKG2D, was found to be engaged in target recognition by NK cells (Textor et al., 2011). A previous study confirmed that the soluble ULBP2 secreted by cancer cells contributed to the immune escape (Waldhauer and Steinle, 2006). Herein, we observed that ULBP2 was upregulated in epithelial cells. Meanwhile, ULBP2 has been shown to be a prognosis indicator for several cancers, such as lung cancer and pancreatic cancer (Chang et al., 2011; Yamaguchi et al., 2012). The activation-induced cytidine deaminase (AICDA) is an essential enzyme of the adaptive immune system. A recent study found that elevated expression of AICDA regulates the function of B cells in regional lymph nodes and significantly improves prognosis of HNSCC patients (Pylaeva et al., 2021). Tyrosine kinase 2 (TYK2), a member of the Janus kinase (JAK) family, has emerged as both a promising biomarker and a target for anticancer therapies (Borcherding et al., 2021). It has been reported that high expression of TYK2 is associated with better prognosis of HNSCC (Fang et al., 2021). A recent review concluded that CC motif chemokine receptor (CCR7) is correlated with good outcomes of HNSCC patients (Korbecki et al., 2020). However, if located on cancer cells, CCR7 and its ligands (CCL19/CCL21) is a vital axis for carcinogenic properties, such as epithelial-mesenchymal transition (EMT) tumor invasion and migration (Chen et al., 2020; Korbecki et al., 2020). Notably, the present study found that CCR7 was predominantly expressed in dendritic cells. SCARB1 has been demonstrated to be involved in cholesterol metabolism, thereby facilitating cancer progression (Gutierrez-Pajares et al., 2016). In addition, stanniocalcin-2 (STC2) exerted a significant role in a wide variety of signaling pathways in HNSCC apoptosis and autophagy (Li et al., 2020). Studies have revealed that
downregulated expression of STC2 can suppress growth of HNSCC cells (Li et al., 2019; Li et al., 2020). Moreover, the leucine-rich repeat-containing G protein-coupled receptor LGR5 participated in Wnt signaling and was intimately linked to the severity of HNSCC (Dalley et al., 2015).

Given the important role of immune cell infiltrations in the diagnosis and treatment of diseases, we further explored the immune landscape in different HNSCC groups. Based on the degree of immune cell infiltrations, particularly CD8+ T cells, the tumor phenotypes can be defined as two major patterns.
“hot” and “cold”, which are associated with good and poor antitumor immune responses, respectively (Galon and Bruni, 2019). This study explored the abundance of immune cells and functions using CIBERSORT, ESTIMATE, and ssGSEA methods. According to the obtained results, the low-risk group exhibited more infiltration of CD8+ T cells, memory activated CD4+ T cells, and plasma cells, as well as higher immune score, and thus can be categorized as “hot” tumor phenotype. On the other hand, the high-
FIGURE 14
Drug response prediction between the two risk groups. (A–H) The IC50 of eight common chemotherapeutic agents (5-Fluorouracil, Bleomycin, Cetuximab, Cisplatin, Docetaxel, Methotrexate, Rapamycin, and Sunitinib) and correlation analysis with risk score. (I–L) The difference of immunophenoscore (IPS) scores among high- and low-risk groups.
risk group showed greater abundance of activated mast cells, resting NK cells, and M2 macrophages, and lower immune score, suggesting the "cold" tumor phenotype. Furthermore, the immune checkpoint-related genes exhibited relatively high expressions in the low-risk group, including IFNG, PRF1, GZMA, GZMB, CXCL10, CXCL9, CD8A, CD274 (PD-L1), HAVCR2, IDO1, LAG3, CTLA4, and PDCD1. Studies have confirmed that infiltration of M2 macrophages is associated with tumorigenic chronic inflammation with secretion of protumorigenic factors, such as IL-6, VEGF, and TGFβ (Ruffell and Coussens, 2015). Accumulating evidence suggests that preexisting CD8+ T cells and PD-L1 expression are generally correlated with improved efficacy of immunotherapy (Farhood et al., 2019; Gavrielatou et al., 2020). Consistently, our results suggested that patients with low-risk score, as a consequence of higher IPS scores, had more vigorous immune responses to anti-PD1 therapy and anti-PD1 plus anti-CTLA4 therapy. Moreover, patients in the two groups exhibited varying sensitivity to four common chemotherapeutic drugs, including bleomycin, cisplatin, methotrexate, and rapamycin (Cramer et al., 2019). Notably, previous studies have verified the therapeutic safety and effectiveness of chemotherapy in combination with PD-L1 blockade (Burtness et al., 2019; Cohen et al., 2019). Nevertheless, different sensitivities to 5-Fluorouracil, cetuximab, docetaxel, and sunitinib were not observed in this study. TMB level was considered to be an indicator of immunotherapy response (Rizvi et al., 2015). We then examined the relationship between TMB and the risk score. The alteration frequency of TP53, PKHD1L1, DNAH9 and
**Conclusion**

In conclusion, by comprehensively analyzing the single-cell and bulk RNA sequencing of HNSCC, this study developed and externally validated a novel and robust model based on eight CD8+ T cells-related genes. It is expected that the 8-gene signature will facilitate understanding of HNSCC immune characteristics, predict prognosis of HNSCC patients, and guide the clinical use of immunotherapy.

**Data availability statement**

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

**References**

Borcherding, D. C., He, K., Amin, N. V., and Hirbe, A. C. (2021). TYK2 in cancer metastases: Genomic and proteomic discovery. *Cancers (Basel)* 13 (16), 4171. doi:10.3390/cancers13164171

Burtness, B., Harrington, K. J., Greel, R., Soulières, D., Tahara, M., de Castro, G., et al. (2019). Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck. *NEJMoa1602252*. doi:10.1038/nrclinonc.2017.101

Chang, Y. T., Wu, C. C., Shyr, Y. M., Chen, T. C., Hwang, T. L., Yeh, T. S., et al. (2011). Secretome-based identification of ULBP2 as a novel serum marker for pancreatic cancer detection. *PLoS One* 6 (5), e20029. doi:10.1371/journal.pone.0020029

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Chen, Y., Shao, Z., Jiang, E., Zhou, X., Wang, L., Wang, H., et al. (2020). CCL21/CXCR7 interaction promotes EMT and enhances the stemness of OSCC via a JAK2/STAT3 signaling pathway. *J. Cell. Physiol.* 235 (9), 5995–6009. doi:10.1002/jcp.29525

Cohen, E. W. E., Soulières, D., Le Tourneau, C., Dinis, J., Licitra, L., Ahn, M.-J., et al. (2019). 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* 393 (10167), 156–167. doi:10.1016/s0140-6736(18)31999-8

Cramer, J. D., Burtness, B., Le, Q. T., and Ferris, R. L. (2019). The changing therapeutic landscape of head and neck cancer. *Nat. Rev. Clin. Oncol.* 16 (11), 669–683. doi:10.1038/s41571-019-0227-z

Dalley, A. J., Abdul Majeed, A. A., Pitty, I. P., Major, A. G., and Farah, C. S. (2015). LGR5 expression in oral epithelial dysplasia and oral squamous cell carcinoma. *Oncol. Rep.* 34 (4), 1856–1860. doi:10.3892/or.2014.4003

Fang, L., Wang, W., Shi, L., Chen, Q., and Rao, X. (2021). Prognostic values and clinical relationship of TYK2 in laryngeal squamous cell cancer. *Med. Baltim.* 100 (34), e27062. doi:10.1097/MD.00000000000027062

Ferris, R. L., Blumenschein, G., Jr., Fayette, J., Guigay, J., Colevas, A. D., Licitra, L., et al. (2016). Nivolumab for recurrent squamous-cell carcinoma of the head and neck. *N. Engl. J. Med.* 375 (19), 1856–1867. doi:10.1056/NEJMoa1602252

Fridman, W. H., Zitvogel, L., Sauve, R., Frédéric, M., and Kroemer, G. (2017). The immune contexture in cancer prognosis and treatment. *Nat. Rev. Clin. Oncol.* 14 (12), 717–734. doi:10.1038/nrclincanc.2017.101

Galon, J., and Bruni, D. (2019). Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat. Rev. Drug Discov.* 18 (3), 197–218. doi:10.1038/s41573-018-0007-y

**Author contributions**

JZ designed the study and collected the data. SZ and WZ carried out the data analyses, prepared all figures and tables and wrote the manuscript. All authors participated in improving the writing of the manuscript and approved the submitted version.

**Conflict of interest**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s note**

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

**Supplementary material**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.938611/full#supplementary-material
Pylaev, E., Ozel, I., Squire, A., Spyra, I., Wallner, C., Korek, M., et al. (2021). B-helper neutrophils in regional lymph nodes correlate with improved prognosis in patients with head and neck cancer. Cancers (Basel) 13 (12), 3092. doi:10.3390/cancers13123092

Qi, Z., Barrett, T., Parikh, A. S., Tiros, I., and Puram, S. V. (2019). Single-cell sequencing and its applications in head and neck cancer. Oral Oncol. 99, 104441. doi:10.1016/j.oraloncology.2019.104441

Rzvi, N. A., Hellmann, M. D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J. L., et al. (2015). Cancer immunology: Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Sci. (New York, N.Y.) 348 (6230), 124–128. doi:10.1126/science.aaa1348

Ruffell, B. and Coussens, L. M. (2015). Macrophages and therapeutic resistance in cancer. Cell 27 (4), 462–472. doi:10.1016/cell.2015.02.015

Sacco, A. G., Chen, R., Worden, F. P., Wong, D. J. L., Adkins, D., Siewicki, P., et al. (2021). Pembrolizumab plus cetuximab in patients with recurrent or metastatic head and neck squamous cell carcinoma: An open-label, multi-arm, non-randomised, multicentre, phase 2 trial. Lancet. Oncol. 22 (6), 883–892. doi:10.1016/s1470-2045(21)00136-4

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. Ca Cancer J. Clin. 70 (1), 7–30. doi:10.3322/caac.21590

Sia, L. I., Even, C., Mesia, R., Remenar, E., Daste, A., Deleod, J. P., et al. (2019). Safety and efficacy of durvalumab with or without tremelimumab in patients with PD-L1-low/negative or metastatic HNSCC. The phase 2 CONOR randomized clinical trial. JAMA Otolaryngol. Head Neck Surg. 145 (2), 195–203. doi:10.1001/jamaotol.2018.4628

Song, P., Li, W., Wu, X., Qian, Z., Ying, J., Gao, S., et al. (2022). Integrated analysis of single-cell and bulk RNA-sequencing identifies a signature based on B cell marker genes to predict prognosis and immunotherapy response in lung adenocarcinoma. Cancer Immunol. Immunother. 1, 1. doi:10.1007/s00262-022-03143-2

Sun, C. Q., Arnold, R., Fernandez-Golzar, C., Parrish, A. B., Almekinders, T., He, J., et al. (2006). Human beta-defensin-1: a potential chromosome 8p tumor suppressor: Control of transcription and induction of apoptosis in renal cell carcinoma. Cancer Res. 66 (17), 8542–8549. doi:10.1158/0008-5472.CAN-06-0294

Textor, S., Fieger, N., Arnold, N., Porgador, A., Hofmann, T. G., Cerwenka, A., et al. (2011). Human NK cell subsets are induced to detect p53 in cancer cells by upregulation of the NKG2D ligands ULBP1 and ULBP2. Cancer Res. 71 (18), 5998–6009. doi:10.1158/0008-5472.CAN-10-3211

Vigneswaran, N., and Williams, M. D. (2014). Epidemiologic trends in head and neck cancer and aids in diagnosis. Oral Maxillofac. Surg. Clin. North Am. 26 (2), 123–141. doi:10.1016/joms.2014.01.001

Waldhauser, I., and Steinek, A. (2006). Proteolytic release of soluble UL16-binding protein 2 from tumor cells. Cancer Res. 66 (5), 2520–2526. doi:10.1158/0008-5472.CAN-05-2520

Wichmann, G., Rosolowski, M., Krohn, K., Kreuz, M., Boehm, A., Reiche, A., et al. (2015). The role of HPV RNA transcription, immune response-related gene expression and diuuptive TP53 mutations in diagnostic and prognostic profiling of head and neck cancer. Int. J. Cancer 137 (12), 2846–2857. doi:10.1002/ijc.29649

Xu, K., Fu, Y., Han, Y., Xia, R., Xu, S., Duan, S., et al. (2020). Fewer tumour-specific PD-1(+)CD8(+) TILs in high-risk “Infiltrating” HPV(+) HNSCC. Br. J. Cancer 123 (6), 932–941. doi:10.1038/s41416-020-0968-6

Yamaguchi, K., Chikuma, H., Shirizu, A., Takata, M., Kinosita, N., Hashimoto, K., et al. (2012). Diagnostic and prognostic impact of serum-soluble UL16-binding protein 2 in lung cancer patients. Cancer Sci. 103 (8), 1405–1413. doi:10.1111/j.1349-7006.2012.02330.x

Yang, F., Long, N., Anelupiritanang, T., Bottomly, D., Savage, J. C., Lee, T., et al. (2022). Identification and prioritization of myeloid malignancy germline variants in a large cohort of adult patients with AML. Blood 139 (8), 1208–1221. doi:10.1182/blood.2021011334

Yoshikara, K., Shahnordlogoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Incurring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. 4, 2612. doi:10.1038/ncomms3612