

# A novel comprehensive immune-related gene signature as a promising survival predictor for the patients with head and neck squamous cell carcinoma

Ruihua Fang<sup>1,2,3,\*</sup>, Muhammad Iqbal<sup>1,2,3,\*</sup>, Lin Chen<sup>1,2,3,\*</sup>, Jing Liao<sup>4</sup>, Jierong Luo<sup>5</sup>, Fanqin Wei<sup>1,2,3</sup>, Weiping Wen<sup>1,2,3,4</sup>, Wei Sun<sup>1,2,3</sup>

<sup>1</sup>Department of Otorhinolaryngology Head and Neck Surgery, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou 510080, Guangdong, P.R. China

<sup>2</sup>Institute of Otorhinolaryngology Head and Neck Surgery, Sun Yat-Sen University, Guangzhou 510080, Guangdong, P.R. China

<sup>3</sup>Guangzhou Key Laboratory of Otorhinolaryngology, Guangzhou 510080, Guangdong, P.R. China

<sup>4</sup>Guangdong Institute of Gastroenterology, The Sixth Affiliated Hospital, Sun Yat-Sen University, Guangzhou 510655, Guangdong, P.R. China

<sup>5</sup>Department of Anesthesia, Guangzhou First People's Hospital, The Second Affiliated Hospital of South China University of Technology, Guangzhou 510080, Guangdong, P.R. China

\*Equal contribution

**Correspondence to:** Wei Sun, Weiping Wen; email: [sunwei26@mail.sysu.edu.cn](mailto:sunwei26@mail.sysu.edu.cn), [wenwp@mail.sysu.edu.cn](mailto:wenwp@mail.sysu.edu.cn)

**Keywords:** head and neck squamous cell carcinoma, immune-related signature, prognosis, immune checkpoint molecules, bioinformatics

**Received:** December 18, 2020

**Accepted:** March 3, 2021

**Published:** April 17, 2021

**Copyright:** © 2021 Fang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Head and neck squamous cell carcinoma (HNSCC), the most frequent subtype of head and neck cancer, continues to have a poor prognosis with no improvement. The TNM stage is not satisfactory for individualized prognostic assessment and it does not predict response to therapy. In the present study, we downloaded the gene expression profiles from TCGA database to establish a training set and GEO database for a validation set. In the training set, we developed an 10 immune-related genes signature which had superior predictive value compared with TNM stage. A nomogram including clinical characteristics was also constructed for accurate prediction. Furthermore, it was determined that our prognostic signature might act as an independent factor for predicting the survival of HNSCC patients. As for the immune microenvironment, our results showed higher immune checkpoint expression (CLTA-4 and PD-1) in low-risk group which might reflect a positive immunotherapy response. Thus, our signature not only provided a promising biomarker for survival prediction, but might be evaluated as an indicator for personalized immunotherapy in patients with HNSCC.

## INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common malignant tumor. Worldwide, approximately 830,000 patients suffer from head and neck cancer and about 430,000 people die from this disease annually [1]. Approximately 75% of these

cases are attributable to tobacco smoking and alcohol abuse, which are the two major risk factors for head and neck cancer [2]. Human papillomavirus (HPV) infection is also a significant etiological factor for HNSCC [3]. However, despite advances in surgery, radiotherapy and chemotherapy for the treatment of HNSCC, the 5-year overall survival rate is only 40%-

50% [4]. In addition, the survival rate of advanced cancer patients is only 34.9%, primarily due to metastases and recurrence [5]. Accurately assessing the prognosis for individual patients and performing personalized treatment is of vital importance. Traditionally, American Joint Committee on Cancer (AJCC) staging system-TNM stage is the most important prognostic indicator for predicting postoperative outcome of HNSCC [6]. Nevertheless, the limitation to TNM staging in evaluating patient prognosis is gradually emerging. For example, patients with the same TNM stage and treatment regimen have different clinical outcomes and it does not predict the effectiveness of patient's treatment [7]. Therefore, it is necessary to identify a novel biomarker that can accurately predict patient prognosis and to stratify high-risk HNSCC patients for more intensive treatment regimens.

An increasing body of evidence suggests that the immune system plays a vital role in patient outcome and tumor molecular profiles may be useful for predicting clinical outcome, as well as identifying therapeutic targets [8, 9]. It is also revealed that a prognostic signature containing several to dozens of genes have laid a foundation for predicting the survival of HNSCC [10–12]. In recent years, cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and programmed cell death protein-1 (PD-1)/programmed cell death-ligand 1 (PD-L1) were found to be important immune checkpoint components, that enable tumors to escape from immune surveillance [13, 14]. With the emergence of checkpoint immunosuppressive therapy, the treatment regimens have completely changed for many advanced malignant tumors including HNSCC. However, the checkpoint blockade immune therapeutics does not respond to all patients and the observed objective response rates are in the range of only 16% to 25% [15, 16]. Therefore, an immune-related prognostic signature which can not only predicts survival, but predicts the immunotherapy response for different groups of patients is urgently need.

In this study, an immune-related prognostic signature was constructed with The Cancer Genome Atlas (TCGA) dataset and further validated for its prognostic value using the GSE41613 dataset. Moreover, the relationship between our signature, infiltrating immune cells, and immune checkpoints was determined. Finally, the mutation characteristics and Gene Set Enrichment Analysis (GSEA) of our gene signature were established. Our goal is to provide a novel molecular biomarker that more effectively predicts prognosis and is strongly

associated with the immune microenvironment in HNSCC patients.

## RESULTS

### Differentially expressed genes and functional enrichment analysis

For the TCGA dataset, a total of 400 immune-related genes (IRGs) (305 upregulated and 95 downregulated) and 63 transcription factors (TFs) (46 upregulated and 17 downregulated), which were differentially expressed in HNSCC tissues (n=502) compared with adjacent normal tissues (n=44), were identified and presented in heat maps and volcano plots (Supplementary Figure 1). The 400 IRGs were further analyzed by the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) algorithms and the top 10 results were shown in Figure 1. The GO functional analyses consisted of the following three parts: biological process (BP), molecular function (MF), and cell component (CC). From Figure 1, we identified eight pathways that were enriched in the immune-related gene signature by KEGG analysis. These primarily included genes involved in MAPK signaling, EGFR tyrosine kinase inhibitor resistance, and Ras signaling, and might participate in the development of HNSCC.

### Regulatory network of TFs

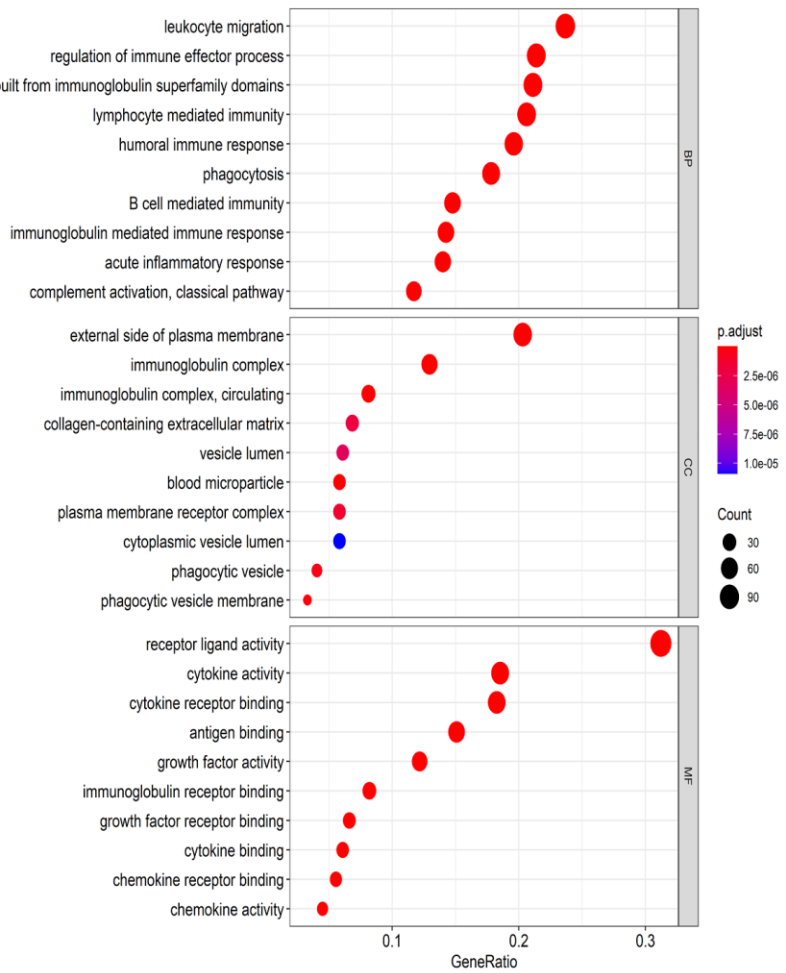
A total of 51 IRGs that significantly associated with overall survival (OS) were identified by using a log-rank test with a univariate Cox analysis ( $p < 0.05$ ). Then, we identified 51 genes with a  $p$ -value  $< 0.05$  for analysis with differentially expressed TFs using a Pearson's correlation test. As a result, correlation coefficients  $> 0.5$  with  $p < 0.05$  were used to establish a network, which was done using the Cytoscape software (Figure 2A). As shown in the network, we found that Foxp3 occupied a major position and positively regulated most of the low-risk IRGs including LTA, CXCR4, CXCR3, IL21R, CD247, ZAP70, SH2D1A, ICOS, and CTLA4. This suggests that Foxp3 may play an important role in the immune regulation of HNSCC.

### Construction and validation of the 10 immune-related gene signature

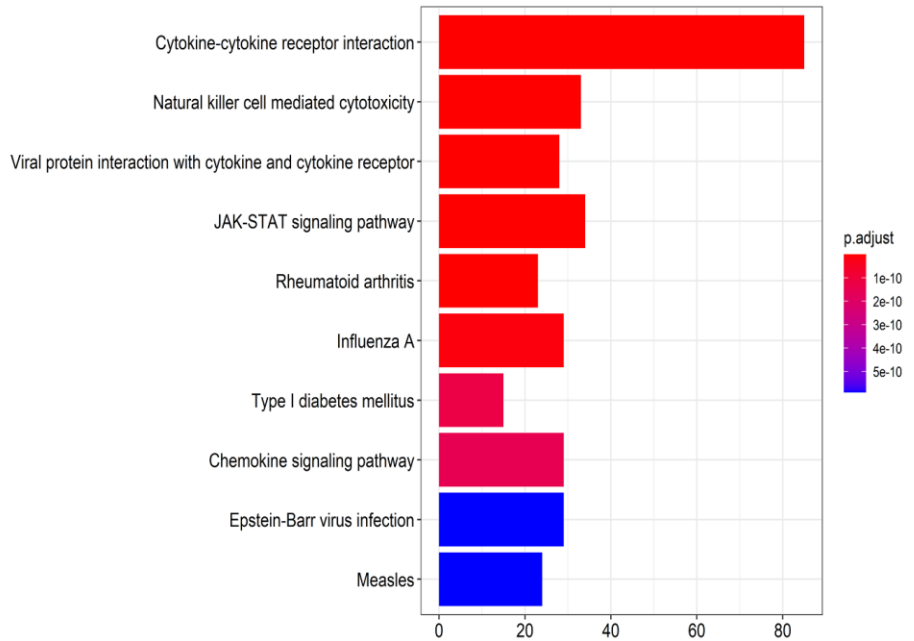
The 24 prognostic IRGs, which were obtained from the univariate Cox analysis with  $p$  value  $< 0.01$  (Figure 2B), were subjected to Lasso analysis and 18 genes were screened out (Figure 3A, 3B). The multivariate Cox regression model was then applied to select the final gene set. As a result, 10 immune-related genes were filtered to establish a prognostic model. The formula for the prognostic model was as follows: risk

**A**

adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains



**B**



**Figure 1. Functional enrichment analysis of differentially expressed IRGs. (A)** The top 10 most significant categories as determined by Gene ontology analysis. The figure represents biological process (BP), cellular component (CC) and molecular function (MF) genes from top to bottom. **(B)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

score = (-0.319322778 \* DEFB1 status) + (-0.417843642 \* EDNRB status) + (0.24520493 \* ADM status) + (0.709588827 \* BTC status) + (0.477760262 \* DKK1 status) + (-0.25962002 \* FAM3D status) + (-0.453322755 \* GNRH1 status) + (0.416649823 \* STC2 status) + (0.2610282 \* TNFRSF12A status) + (-0.357569677 \* CTLA4 status). Based on the prognostic model formula, we calculated each patient's risk score and divided them into high and low risk groups based on the median value of risk scores in the training and validation sets. Then, the risk score and survival status of patients for the 10-gene signature model were determined for the training and validation sets (Figure 3C–3F). Finally, the survival analysis of high and low risk groups in the training and validation sets was presented. Patients in the high risk group had significantly poorer prognosis compared with those in the low risk group ( $p < 0.001$ ; Figure 3G, 3H).

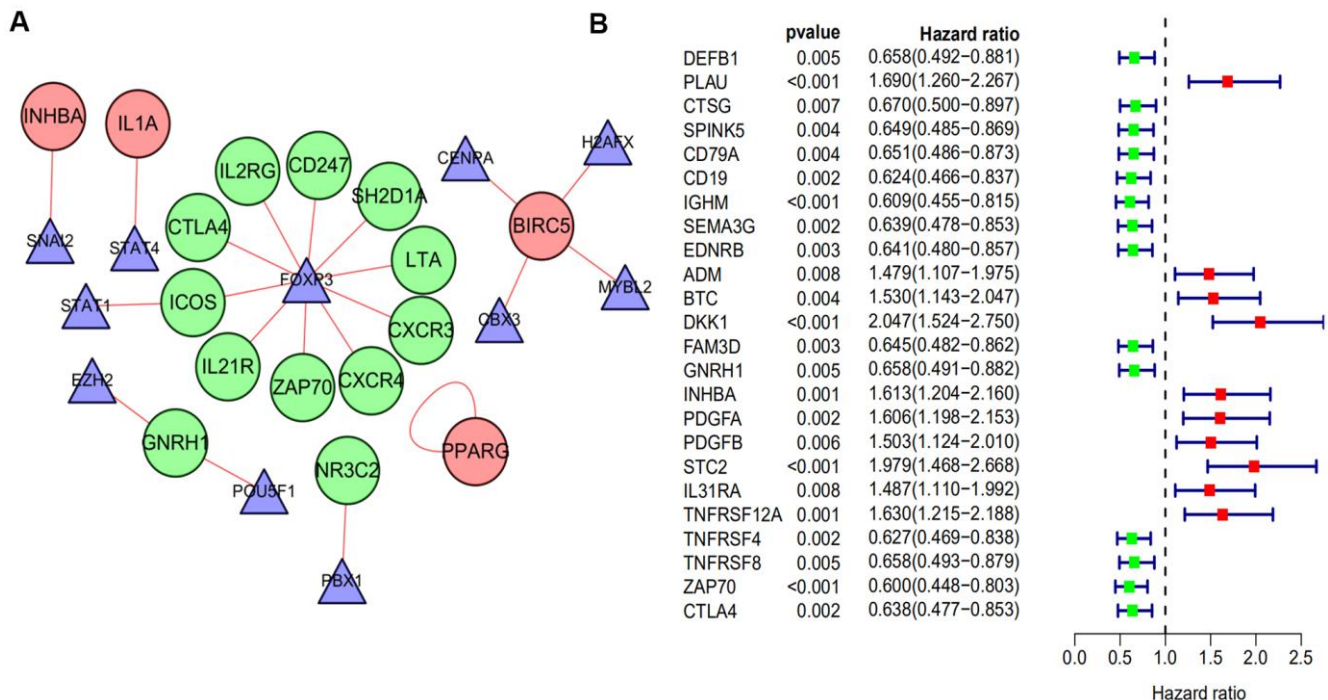
### A 10-gene immune-related signature is an independent prognostic factor

To determine whether the 10-gene immune-related signature was an independent prognostic factor for patient survival in the training and validation sets,

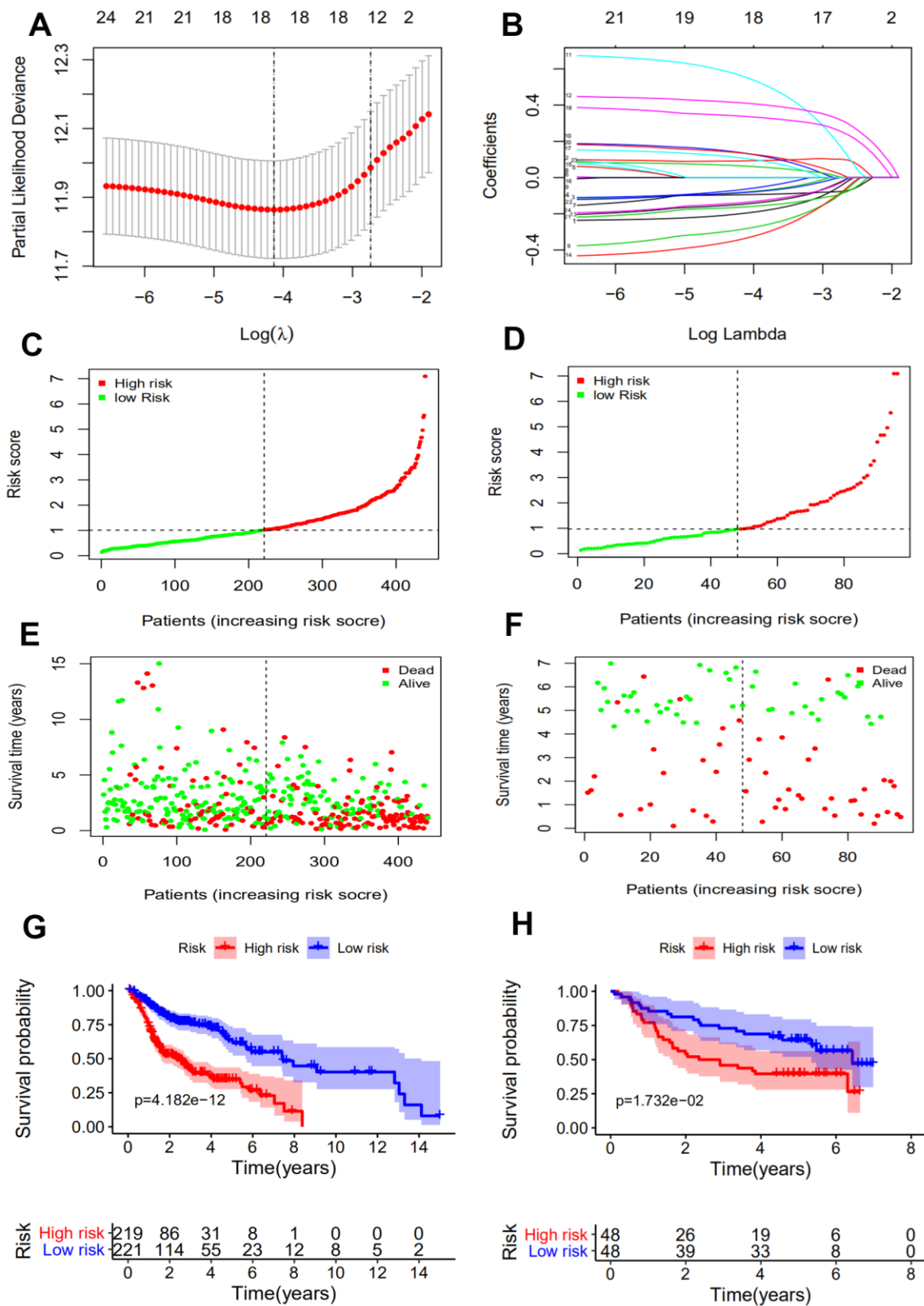
univariate and multivariate Cox models were established (Table 1). The results of the univariate Cox analysis showed that age, N stage, M stage, recurrence and risk score were the factors associated with patient prognosis in the training set. TNM stage and risk score were the factors associated with patient prognosis in the validation set. In addition, our immune-related signature was an independent prognostic risk factor in both the training set (461 cases: HR=1.495, 95% CI (1.349–1.657),  $p < 0.001$ ) and the validation set (96 cases: HR=1.347, 95% CI (1.107–1.638),  $p = 0.003$ ) by multivariate analysis. The association of our signature with clinicopathologic factors was carried out and the results indicated that only T stage, TNM stage, recurrence, and human papillomavirus (HPV) exhibited statistical significance ( $p < 0.05$ , Figure 4).

### External validation of the immune-related gene signature

The area under the ROC curve (AUCs) were 0.715, 0.757 and 0.718 for 2-, 3- and 5-year survival times, respectively, for the training set. The AUCs for the validation set were 0.679, 0.653, 0.645 for 2-, 3- and 5-year survival times, respectively, suggesting that our



**Figure 2. The regulatory network of TF and univariate cox analysis of differentially expressed IRGs and OS of HNSCC. (A)** The regulatory network constructed based on prognosis-related IRGs and TFs. The red circle represents the high-risk IRGs and the green circular represents the low-risk IRGs. The triangle represents differentially expressed TFs. The red link line represents positive regulation, while there was no observed negative regulation. **(B)** A total of 24 differentially expressed IRGs found to be significantly associated with the OS of HNSCC patients ( $p$ -value  $< 0.01$ ). Red points represent positive correlations, while green points represent negative correlations.



**Figure 3. Construction of an immune-related prognostic signature for HNSCC.** (A) Eighteen immune-related genes selected by LASSO Cox analysis. (B) A 10-fold cross-validation for the optimal penalty parameter lambda. (C, D) The risk score distribution of HNSCC patients in the training and validation sets. (E, F) Survival status and duration of patients in the training and validation sets. (G, H) Survival curves for the low and high risk groups in the training and validation sets.

**Table 1. Univariate and multivariate analyses of overall survival in HNSCC patients in training set and validation set.**

Dataset	Characteristics	Univariate		Multivariate	
		HR (95%CI)	<i>p</i> -value	HR (95%CI)	<i>p</i> -value
Training set	Age	1.018(1.004–1.031)	<b>0.012</b>	1.021(1.007–1.035)	<b>0.003</b>
	Gender	0.811(0.593–1.109)	0.189		
	Grade	1.108(0.891–1.377)	0.357		
	Smoking	0.987(0.870–1.120)	0.841		
	Alcohol	1.106(0.809–1.514)	0.527		
	TNM Stage	1.110(0.942–1.308)	0.211		
	T stage	1.115(0.958–1.298)	0.159		
	N stage	1.201(1.027–1.403)	<b>0.022</b>	1.173(1.001–1.374)	<b>0.049</b>
	M stage	7.816(2.865–21.320)	<b>&lt;0.001</b>	4.672(1.641–13.302)	<b>0.004</b>
	Recurrence	3.253(2.430–4.355)	<b>&lt;0.001</b>	2.618(1.935–3.543)	<b>&lt;0.001</b>
	Risk score	1.603(1.454–1.766)	<b>&lt;0.001</b>	1.495(1.349–1.657)	<b>&lt;0.001</b>
Validation set	Age	1.005(0.607–1.662)	0.986		
	Gender	1.092(0.602–1.980)	0.771		
	TNM Stage	3.757(1.919–7.357)	<b>&lt;0.001</b>	3.469(1.764–6.825)	<b>&lt;0.001</b>
	Risk score	1.383(1.152–1.661)	<b>&lt;0.001</b>	1.347(1.107–1.638)	<b>0.003</b>

\*In Cox regression analysis, gender was defined as female=0, male=1; grade, TNM stage and T, N, M stage were defined as 1-4 in accordance with the increase of stage, respectively; smoking, alcohol and recurrence were defined as No=0 and Yes=1, respectively; HR, hazard ratio; CI, confidence interval. Bold values indicate  $p < 0.05$ , which represents statistical significance.

prognostic signature exhibited better sensitivity and specificity (Figure 5A, 5B). Compared with clinicopathologic factors including gender, TNM stage, T stage, N stage, and recurrence, the AUCs of our signature were the highest (Figure 5C). This demonstrated that the 10-gene signature model had a better ability to predict patient prognosis. Furthermore, the nomogram calibration curves for the possibility of 3- and 5-year OS showed consistency between predicted and actual survival in the training set (Figure 5D, 5E). A nomogram with clinicopathologic factors and risk score was established to quantitatively predict the prognosis of HNSCC patients. It revealed that our prognostic signature was a key factor for predicting the survival, while the clinicopathological characteristics showed an inferior impact (Figure 5F).

We further determined whether our 10-gene immune-related signature could predict the survival of patient subgroups. We stratified patients on the basis of clinicopathological factors including age, gender, smoking, alcohol, TNM stage, recurrence, and M0 stage. The results revealed that our signature might be an independent and significant prognostic predictor

for clinical outcome in patient subgroups (Figure 6). The patients with M1 stage disease were not included because of the small number of cases.

#### **Association between immune-related gene signature, immune cell infiltration, and immune checkpoint molecules**

To investigate the association between our immune-related signature and 22 immune infiltration cell types, a correlation analysis was performed with  $|\text{spearman coefficients}| \geq 0.2$  and  $p$  values  $< 0.05$  (Figure 7A–7C). The relationship between our signature and immune checkpoint molecules was also established (Figure 7D–7F). The results of the Pearson’s correlation analysis indicated that our prognostic signature was negatively associated with regulatory T cells (Tregs) ( $r = -0.296$ ,  $p < 0.001$ ), while M0 macrophages ( $r = 0.203$ ,  $p < 0.001$ ) and activated Mast cells ( $r = 0.204$ ,  $p < 0.001$ ) were positively associated. With an increase in immune risk score, the expression of Tregs decreased gradually in contrast to the signatures of M0 macrophages and activated Mast cells. The prognostic signature was also negatively correlated with immune checkpoint molecules, including CTLA-4 ( $r = 0.253$ ,  $p < 0.001$ ) and

PD-1 ( $r = 0.198$ ,  $p < 0.001$ ), but not with PD-L1 ( $p > 0.05$ ).

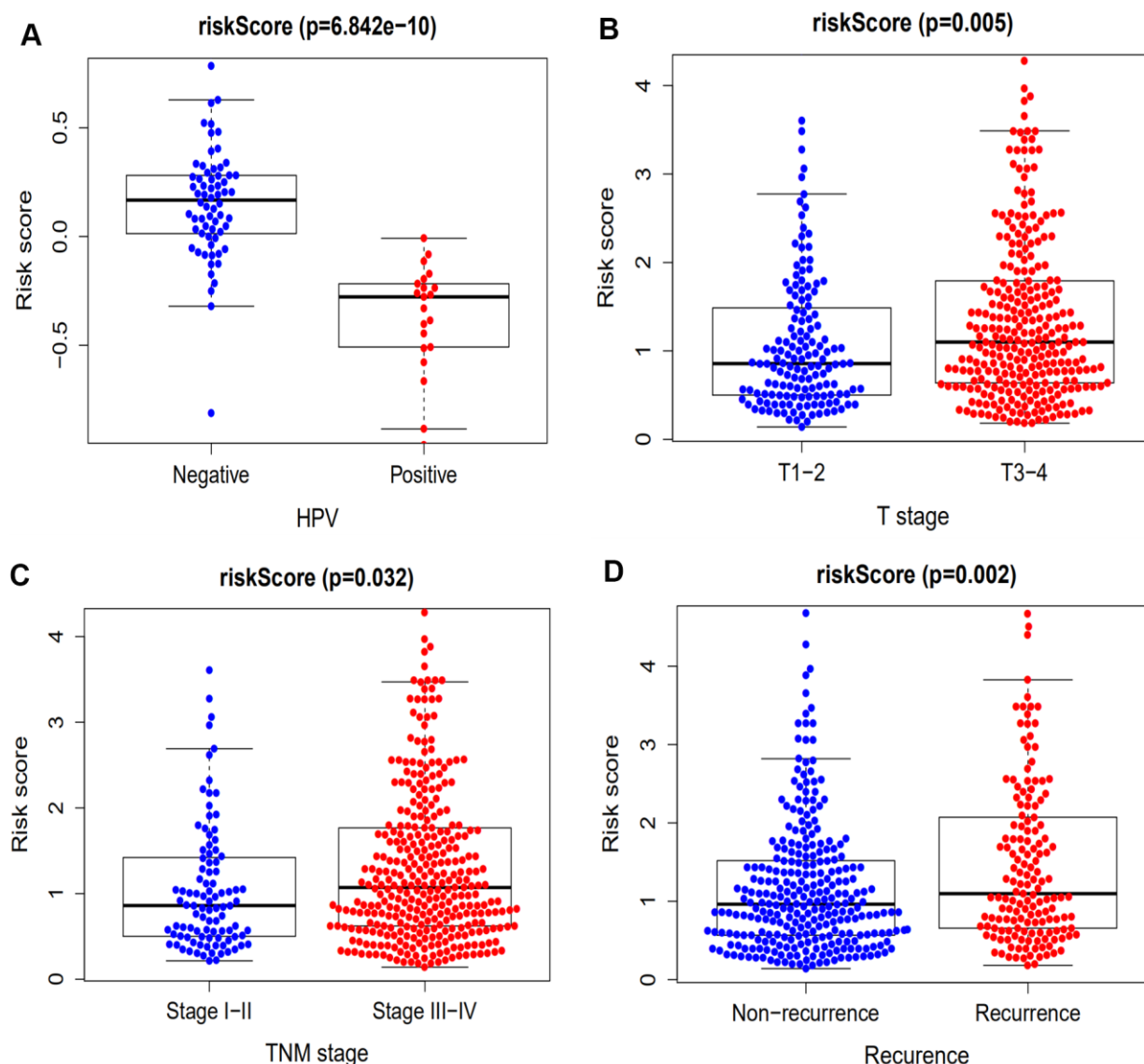
### Genetic alterations and GSEA analysis in the high-risk groups

The genomic alterations of our 10 immune-related genes in each patient were analyzed using the cBioPortal tool (Figure 8A). The genetic alteration percentages ranged from 0.2-3%, and likely have little influence on mRNA expression. The 10 immune-related genes were altered in 62 (13%) of the 496 patients.

DEFB1 and GNRH1 were primarily affected by deep deletion, while the CTLA-4, DKK1 and EDNRB were frequently amplified. The pathways enriched in the high-risk group of the training set were analyzed by GSEA. As a result, there were six pathways that were significantly enriched in the high-risk patients (Table 2 and Figure 8B).

### DISCUSSION

Head and neck squamous cell carcinoma is considered to be a heterogeneous disease and its biological

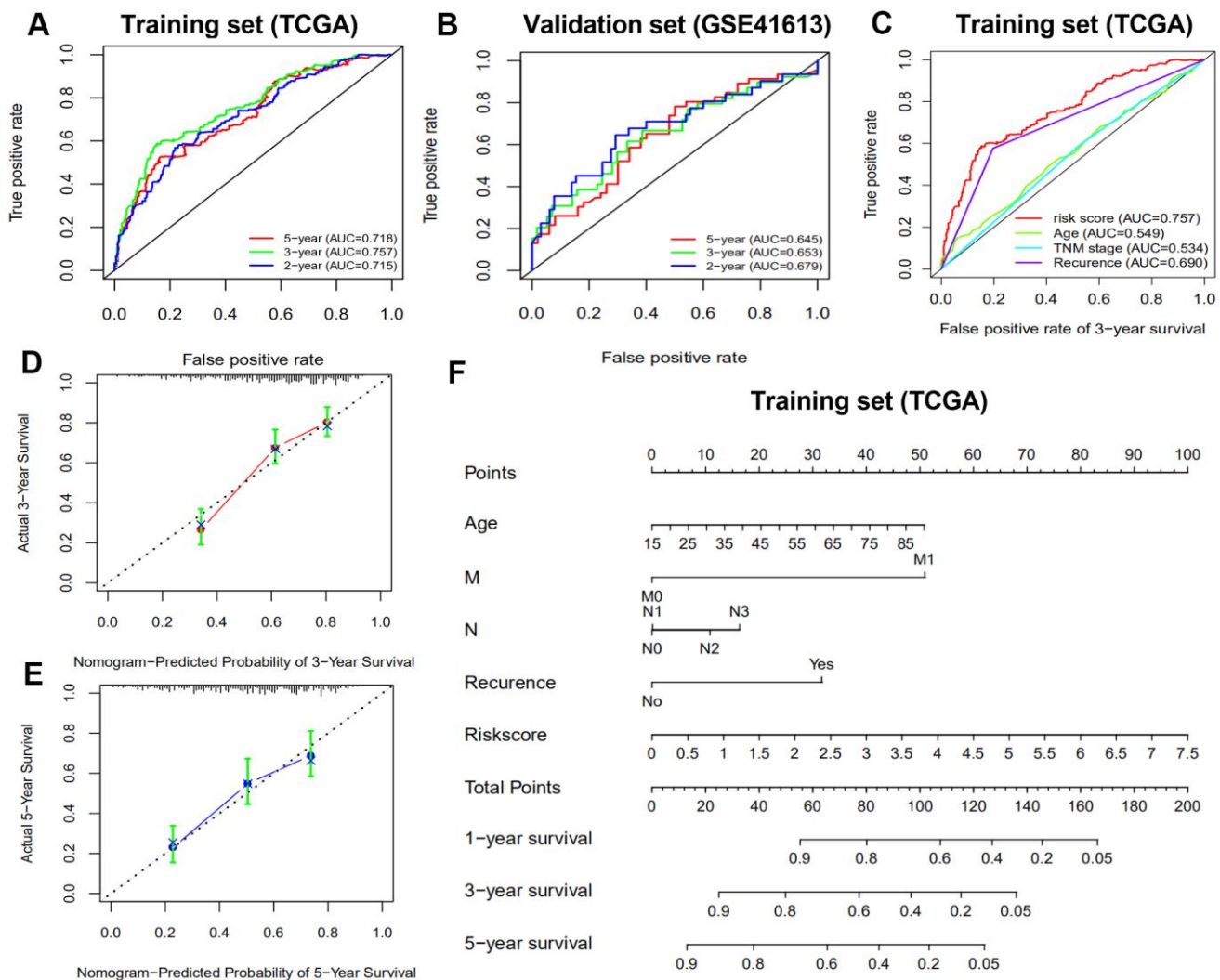


**Figure 4. The association between our prognostic signature and HPV, T stage, TNM stage and recurrence.** (A) The HPV-positive patients exhibited a lower risk score compared with HPV-negative patients ( $p = 6.842 \times 10^{-5}$ ). (B) The immune-related signature risk scores for T3 and 4 stage HNSCC patients were notably higher compared with that of T1 and 2 patients ( $p = 0.005$ ). (C) The immune-related signature risk scores for stage III & IV HNSCC patients were notably higher compared with that of stage I & II patients ( $p = 0.032$ ). (D) Patients with recurrence of HNSCC have higher risk scores compared with those having no recurrence ( $p = 0.002$ ).

behavior is frequently aggressive. The high mortality rate observed for HNSCC is primarily due to the frequent recurrence of advanced tumors [17]. There is a significant need for clinicians to give personalized and realistic prognostic prediction as TNM stage is no longer an accurate prognostic indicator. Therefore, it is crucial to identify new markers that predict clinical outcome, achieve personalized approaches to therapy, and establish early intervention treatments. To date, many studies have tried to establish prognostic signatures, including gene sets [18], miRNAs [19], lncRNAs [20] and methylation analyses [21], as promising predictors of prognosis for HNSCC. In recent years, the immune system has been recognized as playing an important role in cancer development and progression [22]. Nevertheless, the

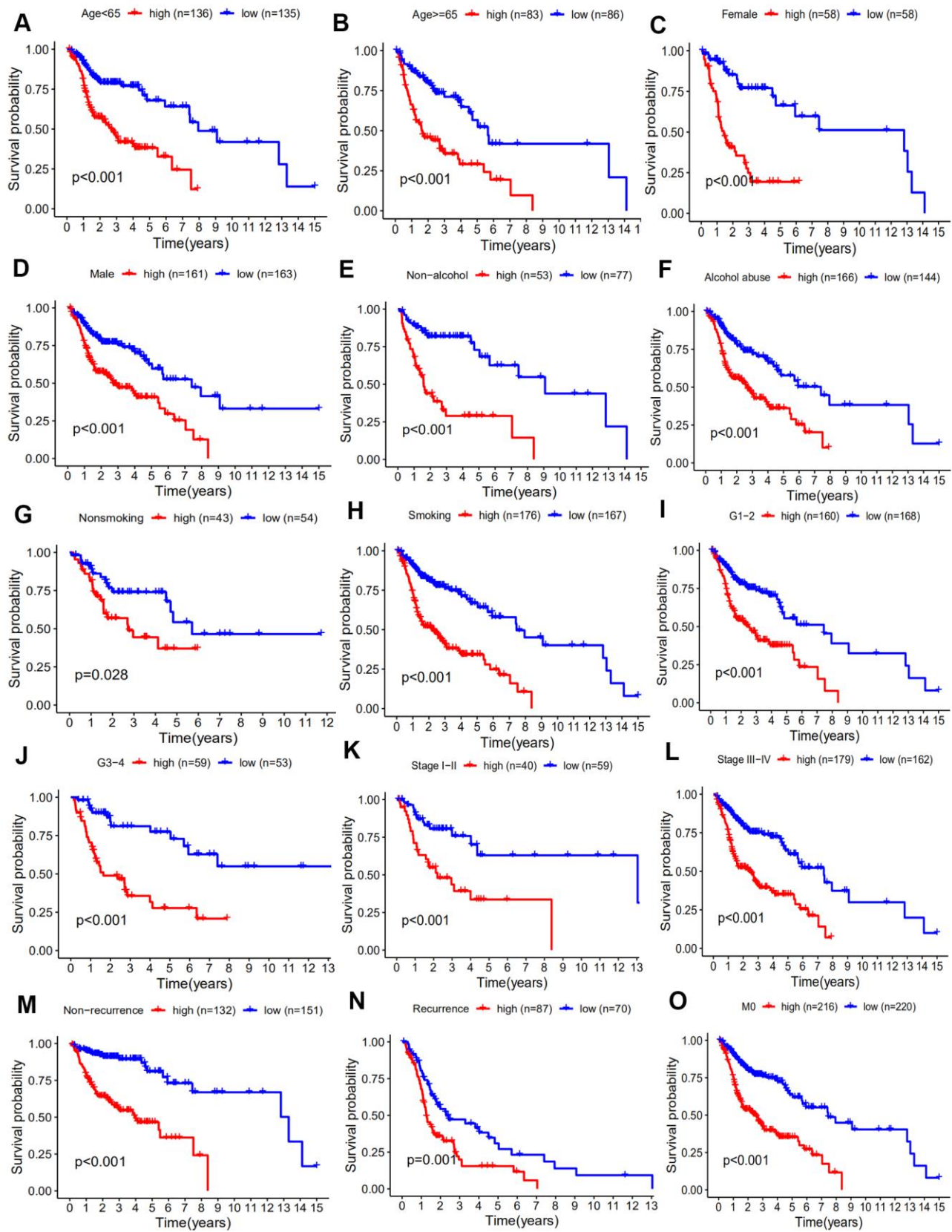
contribution of immune-related molecular mechanisms to HNSCC remain unclear.

In the present study, we screened 400 differentially expressed immune-related genes from the TCGA dataset. Using GO and KEGG enrichment analysis, we found that the immune-related genes were primarily associated with immune response, cancer, and drug resistance pathways (i.e. MAPK signaling pathway, EGFR tyrosine kinase inhibitor resistance, Ras signaling pathway and endocrine resistance). Similar studies have demonstrated that the MAPK signaling pathway participates in cancer progression, including proliferation, apoptosis and immune escape, and it is fundamental to cancer control [23].



**Figure 5. Receiver operating characteristic curve (ROC) analysis and nomogram construction predicted overall survival using risk score. (A, B) ROC analysis for predicting 2-, 3- and 5-year survival times in the training and validation sets. (C) ROC analysis between risk score and clinicopathological characteristics in the training set. (D, E) Calibration curves of the nomogram for predicting the probability of 3- and 5-year survival. (F) A nomogram to quantitatively predict 1-, 2-, and 3-year survival for HNSCC patients based on the prognostic signature and clinicopathological characteristics of the training set.**



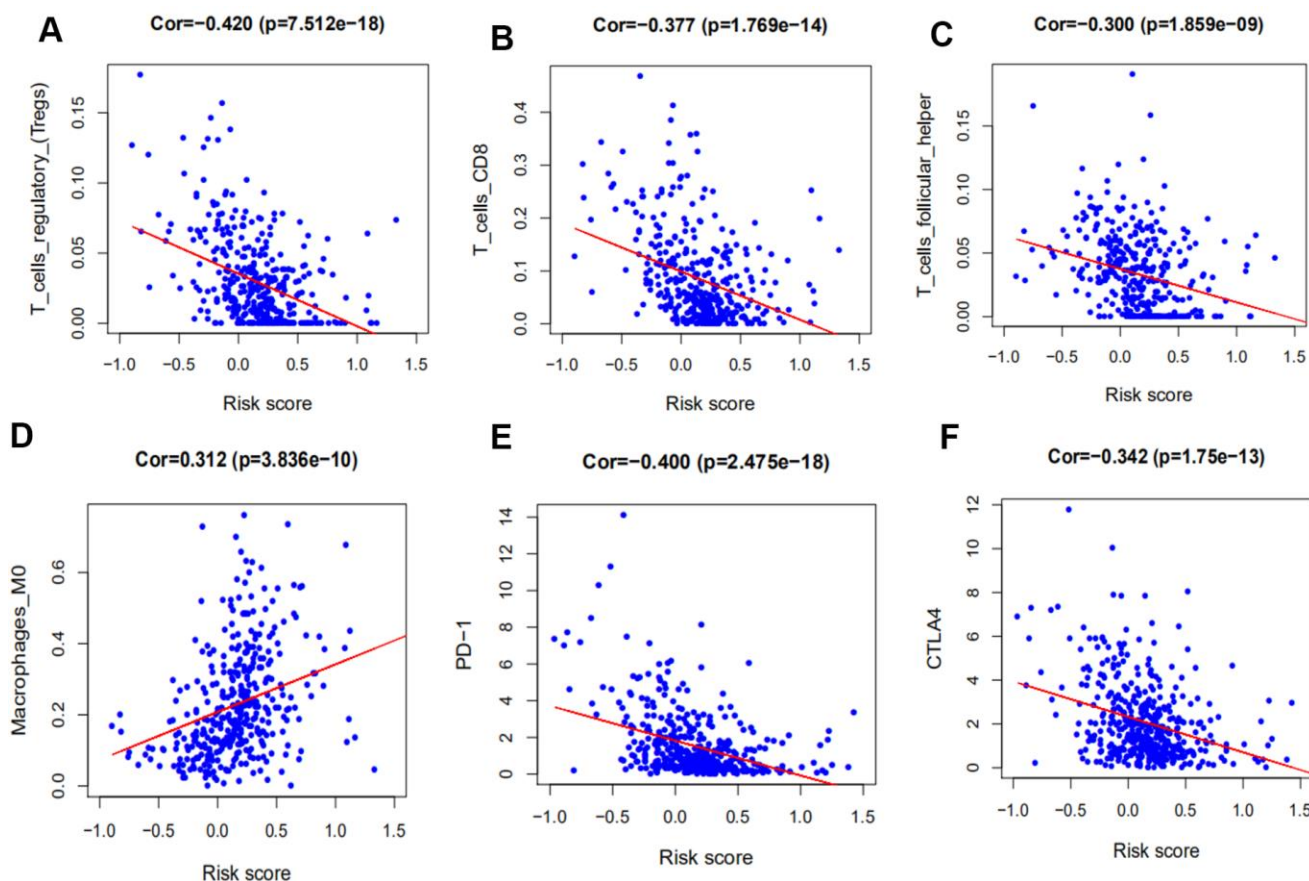


**Figure 6. Kaplan-Meier analyses of HNSCC patient subgroups and clinicopathology factors including. (A) age  $\geq 65$  years, (B) age < 65 years, (C) female, (D) male, (E) non-alcohol, (F) alcohol, (G) Nonsmoking, (H) smoking, (I) G1 and 2, (J) G3 and 4, (K) stage I-II, (L) stage III-IV, (M) non-recurrence, (N) recurrence, (O) M0 stage ( $p < 0.05$ ).**

Transcription factors regulate gene expression and their dysregulation or mutation is well known to contribute to tumorigenesis [24]. Foxp3, a member of the forkhead transcription factor family, is one of the key transcription factors that controls the development and function of Treg cells [25]. An analysis of the TF-mediated network was done to reveal the regulatory mechanisms of prognostic immune-related genes. The results indicated that Foxp3 was a crucial TF that upregulated most of the low-risk prognostic immune-related genes. This suggested that Foxp3 might be a key factor in the immune regulatory mechanism of HNSCC. Foxp3 might also control the immune microenvironment by regulating the expression of genes that contribute to the immunotherapy of HNSCC. Foxp3 and CTLA4 were also determined to be positively correlated in our study, consistent with that of previous studies [26, 27].

It has been demonstrated that immune gene signatures can predict prognosis in many solid tumors including

ovarian cancer [28], clear cell renal cell carcinoma [29], cervical cancer [30], lung adenocarcinoma [31], and hepatocellular carcinoma [32]. In the present study, we developed a prognostic signature based on 10 immune-related genes from TCGA dataset and validated them using GSE41613 dataset. The patients in the high risk group were considered to have short survival times in both datasets, in accordance with previous studies [11]. ROC analysis indicated that our immune signature exhibited better sensitivity and specificity for survival prediction at 2-, 3- and 5-years, even exceeding the predictive ability of TNM stage. Moreover, multivariate Cox analysis indicated that the 10-gene immune-related signature was an independent prognosis risk factor for HNSCC. Patients exhibiting recurrence had higher risk scores compared with patients without recurrence, suggesting that the prognostic signature had a broader predictive value for recurrence. A previous study demonstrated that HPV-positive HNSCC patients had improved survival compared with HPV-negative

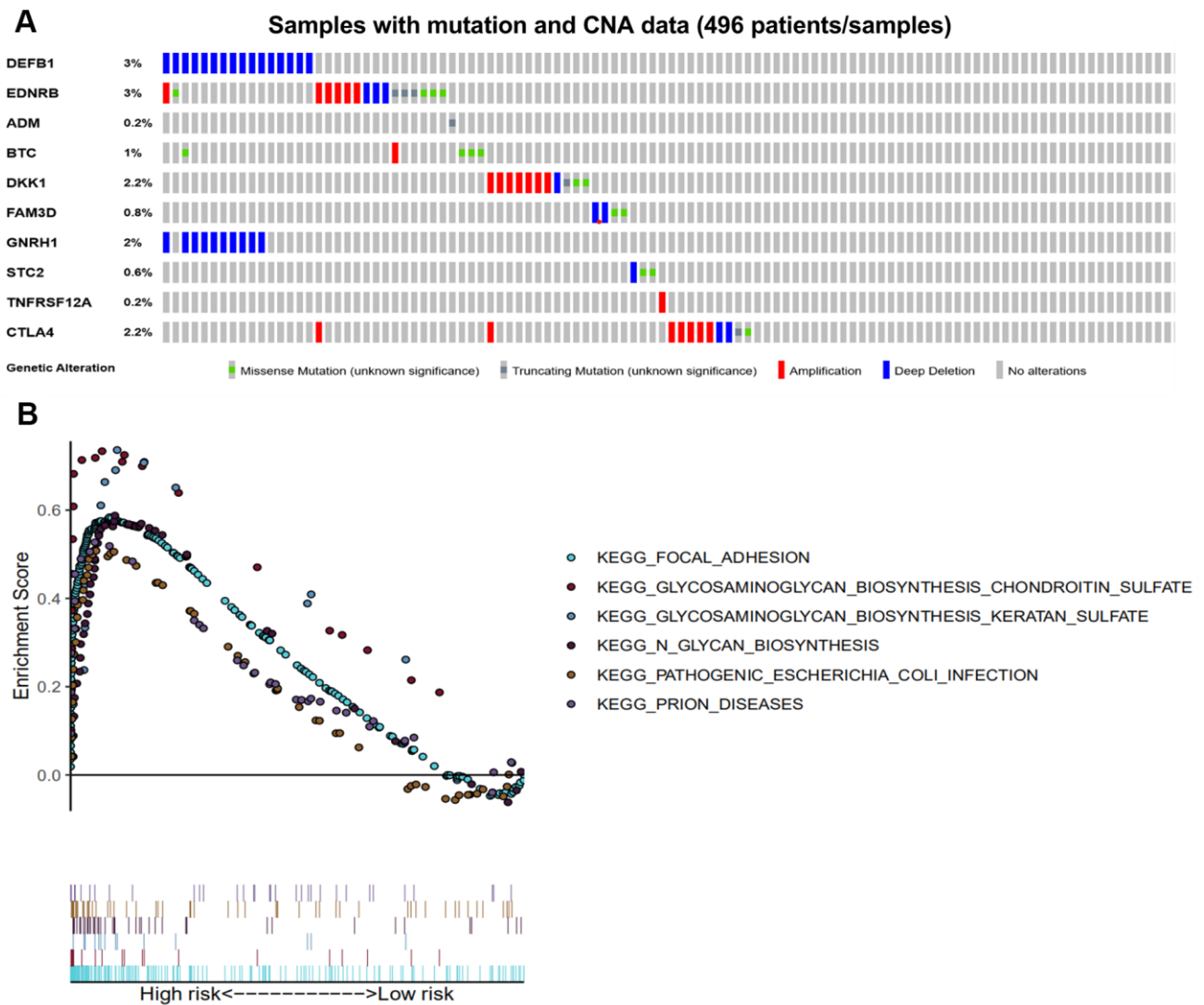


**Figure 7. Correlation between our prognostic signature and that of tumor-infiltrating immune cells and immune checkpoint molecules.** (A) Association between risk score and T regulatory cells (Tregs). (B) Association between risk score and activated Mast cells. (C) Association between risk score and M0 macrophages. (D) Association between risk score and CTLA-4. (E) Association between risk score and PD-1. (F) Association between risk score and PD-L1.

patients [33]. This is consistent with our findings that HPV-positive patients had a lower risk score. We further predicted survival time of patients with different clinical factors based on our 10 immune-related gene signature. This demonstrated that our signature could act as accurately and strongly biomarkers for predicting prognosis in HNSCC patients with various clinicopathologic factors.

Other studies corroborate that our 10 immune-related gene signature is closely related to the development of cancer. For example, DEFBI (encoding human  $\beta$ -defensin-1 [HBD-1]), is a potential tumor suppressor which has been shown to participate in the innate immune response and can suppress tumor migration and invasion in oral squamous cell carcinoma [34]. EDNRB

promoter methylation, which is associated with the histologic diagnosis of premalignancy and the presence of malignancy, may be a promising marker for the early detection of premalignant lesions in oral cavity cancer [35]. The ADM gene, which plays a role in carcinogenesis by regulating cellular processes including proliferation, differentiation, migration, growth, immunosuppression and hypoxia, increases lymph node metastasis risk in oral and oropharynx cancer [36]. Increased BTC mRNA expression is associated with worse survival in HNSCC [37]. Dkk-1, a tumor suppressor gene, is associated with distant metastasis in HNSCC patients, when happens to allelic loss at Dkk-1 locus frequently [38]. Overexpression of FAM3D-AS1 is demonstrated to inhibit cell proliferation, invasion, EMT, and cell survival rate in colorectal cancer [39]. The level



**Figure 8. Genetic alterations of the 10 immune-related genes using GSEA. (A)** Genetic alterations of 10 immune-related genes in HNSCC samples. The rows and columns indicate the genes and tumor samples, respectively. **(B)** The six enriched pathways in the high-risk groups based on the prognostic signature in HNSCC.

**Table 2. The six enriched pathways in the low-risk group.**

NAME	SIZE	ES	NES	NOM <i>p</i> -val	FDR <i>q</i> -val	FWER <i>p</i> -val
KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_KERATAN_SULFATE	15	0.737	1.941	0.000	0.071	0.085
KEGG_PRION_DISEASES	35	0.527	1.753	0.004	0.136	0.349
KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION	56	0.508	1.835	0.004	0.091	0.204
KEGG_FOCAL_ADHESION	199	0.584	1.956	0.004	0.116	0.070
KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_CHONDROITIN_SULFATE	22	0.734	1.927	0.006	0.052	0.090
KEGG_N_GLYCAN_BIOSYNTHESIS	46	0.588	1.846	0.008	0.104	0.193

expression of GNRH1 has been shown to indicate metastatic spread of tumor cells in gynecological malignancies [40]. The downregulation of STC2 plays a vital role in the metastasis and progression of HNSCC [41]. TNFRSF12A contributes to carcinogenesis by promoting angiogenesis, proliferation, apoptosis, migration and inflammation in tumors, can cause cachexia, and is a promising therapeutic target to prolong survival [42]. CTLA-4, a checkpoint for tumor immunotherapy, can induce T cells to be nonreactive and participate in the repression of T cell proliferation, cell cycle progression, and the immune response [43].

To further understand the relationship of immune-related prognostic signature and immune microenvironment, 22 immune infiltrating cells, derived from CIBERSORT, and 3 immune checkpoint molecules were selected for analysis. In cancer, Tregs contribute to tumor immune escape by suppressing the antitumor response. Some studies had reported that higher Foxp3+ Treg cell infiltration was associated with poorer patient survival in most tumors, but not in HNSCC [44, 45]. Our results showed that the high level of Tregs infiltration was significantly associated with the low risk score, which was associated with favorable prognosis. Macrophages that infiltrate into the tumor microenvironment may facilitate tumor growth, angiogenesis, invasion, and metastasis, and are associated with poor prognosis in HNSCC [46]. Our study indicated that they were positively associated with risk score and an increase in M0 macrophages portended a poor prognosis.

Recently, the immun checkpoints involving PD-1, PD-L1, and CTLA-4 represent promising immunotherapy targets for antitumor therapy [47]. PD-1 is a transmembrane protein that is mainly expressed on the surface of T lymphocytes. CTLA-4 is a membrane glycoprotein that is frequently expressed on Tregs. The mechanism of action of CTLA-4 and PD-1 remain controversial. Our study revealed that high CTLA-4 expression was significantly associated with a lower risk score. This may be caused by the upregulation of

Tregs in HNSCC, suggesting that CTLA-4 may be involved in some aspect of the antitumor effect. Studies have reported that HNSCC patients exhibiting high PD-L1/PD-1 expression tend to have prolonged survival outcomes and a lower probability of recurrence [48, 49]. In addition, our signature was negatively correlated with PD-1, but not PD-L1.

Studies have found that the genetic variation of HBD-1 contributes to lower RNA expression and may be involved in carcinogenesis of oral squamous cell carcinoma [50]. The genetic alterations of our 10 immune-related genes may help explain the aberrant expression of these genes to some extent in tumors, and patients that carry such genetic alterations may be more responsive to immunotherapy. Meanwhile, our GSEA results indicated that six enriched pathways in the high-risk immune group were significantly correlated with the biological processes in HNSCC progression. For example, the pathway of focal adhesion in HNSCC participates in the development of distant metastasis to lymph nodes [51]. The keratan sulphate in the tumor environment plays a vital role in the promotion or regulation of tumor development [52]. The chondroitin sulfate glycosaminoglycan biosynthesis pathway can promote and regulate tumor progression and metastasis by influencing important biological processes such as cell growth, adhesion, signal transduction, and lipid metabolism [53].

The above results demonstrate that our signature has potential clinical prognostic value and is associated with response regulated by the immune microenvironment. This may provide a potential target for diagnosis and for development of new targeted therapies.

There are several limitations to our study. Firstly, our study was performed using public databases and had a retrospective design, so further studies should be done with prospective clinical data sets to validate our results. Secondly, although we established and verified our model using different gene expression datasets, concerns

regarding sample bias and model over-fitting still remained. Furthermore, the infiltrating cell populations were calculated by an analytical tool (CIBERSORT) using gene expression data. This is different from the patient's tumor cell infiltration and thus have the false discovery rate. Finally, the biological functions of our immune signature should be further validated in biological experiments.

## MATERIALS AND METHODS

### Study samples

The training set was acquired from the TCGA data portal (<https://portal.gdc.cancer.gov/cart>; up to March 22, 2020), and consisted of processed RNA-Seq FPKM data for HNSCC patients (n=461). The corresponding clinical data, including age, gender, smoking, alcohol abuse, differentiation grade, clinical TNM stage, T stage, N stage, M stage, recurrence, and survival information were completely provided. Next, data from 96 cases (GSE41613) were downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) along with clinical data and follow-up time as a validation set. Only patients with complete clinical and expression data available at that time were included in this study and the survival time threshold in our study was greater than one month. The clinicopathological characteristics of HNSCC patients from the training and validation sets were showed (Supplementary Table 1). The OS was defined as the date of the study enrollment to the last follow-up time.

### Differentially expressed genes of IRGs and TFs

We obtained IRGs from the Immunology Database and Analysis Portal (ImmPort) (<https://www.immport.org/>) and TFs data were downloaded from the Cistrome Cancer resource (<http://cistrome.org/CistromeCancer/CancerTarget/>). To establish a training set, we used the R package Limma to identify differentially expressed genes (DEGs) for IRGs and TFs from 502 HNSCC tissues and 44 adjacent normal tissues in TCGA dataset, where  $FDR < 0.05$  and  $|\log(FC)| \geq 1$  were set as the screening criteria [54]. Heat maps and volcano plots of IRGs and TFs were also generated with R software. Furthermore, we assessed the biologic functions of differentially expressed IRGs using the GO and KEGG pathway databases. Enrichment analysis was done with the Cluster Profiler package [55] in R and functional categories with  $p < 0.05$  were shown.

### Regulatory TF networks

To further investigate the relationship between DEGs of IRGs and HNSCC prognosis, we used the univariate

Cox proportional hazard model. Genes with a  $p$ -value  $< 0.05$  were selected for further analysis and genes with a hazard ratio (HR) value  $> 1$  were defined as high-risk IRGs, whereas the remainder were considered low-risk. Finally, the association between the above prognosis IRGs and differentially expressed TFs was analyzed by Pearson's correlation test. The cut-off criteria included correlation coefficients  $> 0.5$  and  $p$ -values  $< 0.05$ , which were determined by the `cor.test` function in R (Supplementary Table 2). To more clearly express the relationships, we used Cytoscape for constructing and visualizing the regulatory network [56].

### Construction and validation of an immune-related signature

First, to normalize the differentially expressed IRGs values in the training and validation sets, gene expression values lower than the median were defined as 0, otherwise a value of 1 was assigned. Second, to construct an immune-related prognostic signature, we selected the differentially expressed IRGs with a  $p$ -value  $< 0.01$  by univariate cox analysis. We then used LASSO Cox regression and multivariate Cox regression model to assess the relationship between prognostic immune-related gene expression and OS in the training set using the survival and glmnet R packages. The smallest parameter model for the immune-related prognostic signature was constructed with a 10-fold cross-validation and used one standard error of the best penalty parameter  $\lambda$  to prevent overfitting [57]. Finally, the risk score for each patient was calculated by gene expression and its corresponding coefficients from the multivariate Cox regression analysis. Patients were then divided into high and low risk groups based on the median risk score. To validate the immune-related prognostic signature, we used the same formula as the training set to calculate each patient risk score followed by classification into high and low risk groups.

A prognostic nomogram combined with prognostic clinicopathological factors, including age, N stage, M stage, recurrence, and immune-related gene signature was constructed using the rms R package. To further validate the prognostic value of our signature, Kaplan-Meier analysis of OS in HNSCC patients with subgroup clinicopathological factors was performed.

### Correlation analysis of infiltrating immune cells and immune checkpoint genes

CIBERSORT is an analytical tool that can provide an estimation of the abundances of member cell types in a mixed cell population by using gene expression data. In this study, we identified 22 immune infiltrating cell types by uploading the gene expression training set to

the CIBERSORT webpage (<https://cibersort.stanford.edu/>) and using a reference LM22 expression signature with 100 permutations. The infiltrating immune cells derived from CIBERSORT included T cells (CD4+ T cells, CD8+ T cells, naïve CD4+ T cells, resting memory CD4+ T cells,  $\gamma\delta$  T cells, regulatory T cells, follicular helper T cells and regulatory T cells), B cells (naïve B cells, memory B cells and plasma cells), myeloid subsets (M0 macrophages, M1 macrophages, M2 macrophages, activated and resting dendritic cells, activated and resting mast cells, monocytes, neutrophils and eosinophils), and NK cells (activated and resting NK cells). In addition, only the results of the infiltrating immune cell fractions with a  $p$ -value  $< 0.05$  were considered for further analysis. Finally, the correlations between the immune-related signature and tumor-infiltrating immune cells, immune checkpoint modulators, such as PD-1, PD-L1 and CTLA-4, were determined.

### **Mutation characteristics and gene set enrichment analysis of the immune-related signature**

The mutation characteristics of our immune-related signature in all HNSCC patients from the TCGA dataset were obtained using cBioPortal (<http://www.cbioportal.org/>). GSEA was performed to identify the pathways that were significantly enriched between high risk groups based in the immune-related signature. An FDR  $< 0.25$  and nominal  $p < 0.05$  were used as the screening criteria to identify significant gene sets by the GSEA software.

### **Statistical analysis**

Kaplan–Meier analysis was done to compare the OS between high and low-risk groups using the log-rank test. Meanwhile, ROC curve was used to evaluate the accuracy of the immune-related gene signature. The clinicopathological characteristics on OS was determined by univariate and multivariate analyses on the basis of the Cox proportional hazards model for both training and validation sets. Furthermore, the correlation between immune-related signature and tumor-infiltrating immune cells and immune checkpoint molecules was evaluated by Pearson's correlation test. All analyses were conducted using R software (version 3.6.2) and the results were considered significant when corresponding  $p$ -values  $< 0.05$ .

### **Abbreviations**

HNSCC: Head and neck squamous cell carcinoma; AJCC: American Joint Committee on Cancer; TFs: Transcription factors; LASSO: Least absolute shrinkage and selector operation; HPV: Human papillomavirus;

qRT-PCR: quantitative real-time polymerase chain reaction; CTLA-4: Cytotoxic T-lymphocyte-associated protein 4; PD-1: Programmed cell death protein-1; PD-L1: Programmed cell death-ligand 1; TCGA: The Cancer Genome Atlas; GEO: Group on Earth Observations; GSEA: Gene Set Enrichment Analysis; OS: Overall survival; IRGs: Immune-related genes; ImmPort: Immunology Database and Analysis Portal; DEGs: Differentially expressed genes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; HR: Hazard Ratio; ROC: Operating characteristic curve; BP: Biological process; MF: Molecular function; CC: Cell component; AUCs: Area under the ROC curve; Tregs: Regulatory T cells.

### **AUTHOR CONTRIBUTIONS**

RF, WS and WW participated in the design of the study. FW and MI were responsible for collecting the clinical information and gene expression data, samples and interpreted the data. RF, JL and LC performed the literature review and statistical analyses of data. RF and MI drafted the manuscript. WW and WS revised the manuscript. All authors read and approved the final manuscript.

### **ACKNOWLEDGMENTS**

We would like to appreciate The Cancer Genome Atlas (TCGA) database and Group on Earth Observations (GEO) database for the data they share, as well as the Edanz Group (<https://en-author-services.edanzgroup.com/>) for editing a draft of this manuscript.

### **CONFLICTS OF INTEREST**

The authors declare that they have no conflicts of interest.

### **FUNDING**

The study was supported by the National Natural Science Foundation of China (NSFC) grants 81870696, 81670902 (W.P.W.), 81602365, 81972527 (W.S.), 81902753, Guangdong Natural Science Foundation of China grants 2018B030312008 (W.P.W.) and 2016A030310167, Guangzhou Science and Technology Project of China grant 201704020098, 201605030003.

### **REFERENCES**

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68:394–424.

- <https://doi.org/10.3322/caac.21492>  
PMID:[30207593](https://pubmed.ncbi.nlm.nih.gov/30207593/)
2. Farsi NJ, Rousseau MC, Schlecht N, Castonguay G, Allison P, Nguyen-Tan PF, Soulières D, Coutlée F, Hier M, Madathil S, Franco EL, Nicolau B. Aetiological heterogeneity of head and neck squamous cell carcinomas: the role of human papillomavirus infections, smoking and alcohol. *Carcinogenesis*. 2017; 38:1188–95.  
<https://doi.org/10.1093/carcin/bgx106>  
PMID:[29029021](https://pubmed.ncbi.nlm.nih.gov/29029021/)
  3. Jou A, Hess J. Epidemiology and molecular biology of head and neck cancer. *Oncol Res Treat*. 2017; 40:328–32.  
<https://doi.org/10.1159/000477127>  
PMID:[28531899](https://pubmed.ncbi.nlm.nih.gov/28531899/)
  4. Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. *Nat Rev Cancer*. 2011; 11:9–22.  
<https://doi.org/10.1038/nrc2982>  
PMID:[21160525](https://pubmed.ncbi.nlm.nih.gov/21160525/)
  5. Chauhan SS, Kaur J, Kumar M, Matta A, Srivastava G, Alyass A, Assi J, Leong I, MacMillan C, Witterick I, Colgan TJ, Shukla NK, Thakar A, et al. Prediction of recurrence-free survival using a protein expression-based risk classifier for head and neck cancer. *Oncogenesis*. 2015; 4:e147.  
<https://doi.org/10.1038/oncsis.2015.7>  
PMID:[25893634](https://pubmed.ncbi.nlm.nih.gov/25893634/)
  6. García J, López M, López L, Bagué S, Granell E, Quer M, León X. Validation of the pathological classification of lymph node metastasis for head and neck tumors according to the 8th edition of the TNM classification of Malignant tumors. *Oral Oncol*. 2017; 70:29–33.  
<https://doi.org/10.1016/j.oraloncology.2017.05.003>  
PMID:[28622888](https://pubmed.ncbi.nlm.nih.gov/28622888/)
  7. Angell H, Galon J. From the immune contexture to the Immunoscore: the role of prognostic and predictive immune markers in cancer. *Curr Opin Immunol*. 2013; 25:261–67.  
<https://doi.org/10.1016/j.coi.2013.03.004>  
PMID:[23579076](https://pubmed.ncbi.nlm.nih.gov/23579076/)
  8. Mirza AH, Thomas G, Ottensmeier CH, King EV. Importance of the immune system in head and neck cancer. *Head Neck*. 2019; 41:2789–800.  
<https://doi.org/10.1002/hed.25716> PMID:[30821023](https://pubmed.ncbi.nlm.nih.gov/30821023/)
  9. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, Diehn M, West RB, Plevritis SK, Alizadeh AA. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med*. 2015; 21:938–45.  
<https://doi.org/10.1038/nm.3909> PMID:[26193342](https://pubmed.ncbi.nlm.nih.gov/26193342/)
  10. Wang J, Chen X, Tian Y, Zhu G, Qin Y, Chen X, Pi L, Wei M, Liu G, Li Z, Chen C, Lv Y, Cai G. Six-gene signature for predicting survival in patients with head and neck squamous cell carcinoma. *Aging (Albany NY)*. 2020; 12:767–83.  
<https://doi.org/10.18632/aging.102655>  
PMID:[31927533](https://pubmed.ncbi.nlm.nih.gov/31927533/)
  11. She Y, Kong X, Ge Y, Yin P, Liu Z, Chen J, Gao F, Fang S. Immune-related gene signature for predicting the prognosis of head and neck squamous cell carcinoma. *Cancer Cell Int*. 2020; 20:22.  
<https://doi.org/10.1186/s12935-020-1104-7>  
PMID:[31988638](https://pubmed.ncbi.nlm.nih.gov/31988638/)
  12. Yang B, Fu L, Xu S, Xiao J, Li Z, Liu Y. A nomogram based on a gene signature for predicting the prognosis of patients with head and neck squamous cell carcinoma. *Int J Biol Markers*. 2019; 34:309–17.  
<https://doi.org/10.1177/1724600819865745>  
PMID:[31452437](https://pubmed.ncbi.nlm.nih.gov/31452437/)
  13. Miyauchi S, Kim SS, Pang J, Gold KA, Gutkind JS, Califano JA, Mell LK, Cohen EE, Sharabi AB. Immune modulation of head and neck squamous cell carcinoma and the tumor microenvironment by conventional therapeutics. *Clin Cancer Res*. 2019; 25:4211–23.  
<https://doi.org/10.1158/1078-0432.CCR-18-0871>  
PMID:[30814108](https://pubmed.ncbi.nlm.nih.gov/30814108/)
  14. Azoury SC, Gilmore RC, Shukla V. Molecularly targeted agents and immunotherapy for the treatment of head and neck squamous cell cancer (HNSCC). *Discov Med*. 2016; 21:507–16.  
PMID:[27448787](https://pubmed.ncbi.nlm.nih.gov/27448787/)
  15. Haddad R, Cohen EE, Venkatachalam M, Young K, Singh P, Shaw JW, Korytowsky B, Abraham P, Harrington KJ. Cost-effectiveness analysis of nivolumab for the treatment of squamous cell carcinoma of the head and neck in the United States. *J Med Econ*. 2020; 23:442–47.  
<https://doi.org/10.1080/13696998.2020.1715414>  
PMID:[31928375](https://pubmed.ncbi.nlm.nih.gov/31928375/)
  16. Seiwert TY, Burtneß B, Mehra R, Weiss J, Berger R, Eder JP, Heath K, McClanahan T, Lunceford J, Gause C, Cheng JD, Chow LQ. Safety and clinical activity of pembrolizumab for treatment of recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-012): an open-label, multicentre, phase 1b trial. *Lancet Oncol*. 2016; 17:956–65.  
[https://doi.org/10.1016/S1470-2045\(16\)30066-3](https://doi.org/10.1016/S1470-2045(16)30066-3)  
PMID:[27247226](https://pubmed.ncbi.nlm.nih.gov/27247226/)
  17. Pipkorn P, Licata J, Kallogjeri D, Piccirillo JF. Association of symptoms and clinical findings with anticipated outcomes in patients with recurrent head and neck cancer. *JAMA Otolaryngol Head Neck Surg*. 2018; 144:738–45.

- <https://doi.org/10.1001/jamaoto.2018.1230>  
PMID:[30003215](https://pubmed.ncbi.nlm.nih.gov/30003215/)
18. Xing L, Guo M, Zhang X, Zhang X, Liu F. A transcriptional metabolic gene-set based prognostic signature is associated with clinical and mutational features in head and neck squamous cell carcinoma. *J Cancer Res Clin Oncol*. 2020; 146:621–30.  
<https://doi.org/10.1007/s00432-020-03155-4>  
PMID:[32067104](https://pubmed.ncbi.nlm.nih.gov/32067104/)
  19. Liu C, Yu Z, Huang S, Zhao Q, Sun Z, Fletcher C, Jiang Y, Zhang D. Combined identification of three miRNAs in serum as effective diagnostic biomarkers for HNSCC. *EBioMedicine*. 2019; 50:135–43.  
<https://doi.org/10.1016/j.ebiom.2019.11.016>  
PMID:[31780396](https://pubmed.ncbi.nlm.nih.gov/31780396/)
  20. Yang B, Shen J, Xu L, Chen Y, Che X, Qu X, Liu Y, Teng Y, Li Z. Genome-wide identification of a novel eight-lncRNA signature to improve prognostic prediction in head and neck squamous cell carcinoma. *Front Oncol*. 2019; 9:898.  
<https://doi.org/10.3389/fonc.2019.00898>  
PMID:[31620361](https://pubmed.ncbi.nlm.nih.gov/31620361/)
  21. Zhao X, Cui L. Development and validation of a m<sup>6</sup>A RNA methylation regulators-based signature for predicting the prognosis of head and neck squamous cell carcinoma. *Am J Cancer Res*. 2019; 9:2156–69.  
PMID:[31720080](https://pubmed.ncbi.nlm.nih.gov/31720080/)
  22. Ferris RL. Immunology and immunotherapy of head and neck cancer. *J Clin Oncol*. 2015; 33:3293–304.  
<https://doi.org/10.1200/JCO.2015.61.1509>  
PMID:[26351330](https://pubmed.ncbi.nlm.nih.gov/26351330/)
  23. Peluso I, Yarla NS, Ambra R, Pastore G, Perry G. MAPK signalling pathway in cancers: olive products as cancer preventive and therapeutic agents. *Semin Cancer Biol*. 2019; 56:185–95.  
<https://doi.org/10.1016/j.semcancer.2017.09.002>  
PMID:[28912082](https://pubmed.ncbi.nlm.nih.gov/28912082/)
  24. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. *Cell*. 2013; 152:1237–51.  
<https://doi.org/10.1016/j.cell.2013.02.014>  
PMID:[23498934](https://pubmed.ncbi.nlm.nih.gov/23498934/)
  25. Barbi J, Pardoll DM, Pan F. Ubiquitin-dependent regulation of Foxp3 and Treg function. *Immunol Rev*. 2015; 266:27–45.  
<https://doi.org/10.1111/imr.12312> PMID:[26085205](https://pubmed.ncbi.nlm.nih.gov/26085205/)
  26. Jie HB, Gildener-Leapman N, Li J, Srivastava RM, Gibson SP, Whiteside TL, Ferris RL. Intratumoral regulatory T cells upregulate immunosuppressive molecules in head and neck cancer patients. *Br J Cancer*. 2013; 109:2629–35.  
<https://doi.org/10.1038/bjc.2013.645>  
PMID:[24169351](https://pubmed.ncbi.nlm.nih.gov/24169351/)
  27. Chen C, Rowell EA, Thomas RM, Hancock WW, Wells AD. Transcriptional regulation by Foxp3 is associated with direct promoter occupancy and modulation of histone acetylation. *J Biol Chem*. 2006; 281:36828–34.  
<https://doi.org/10.1074/jbc.M608848200>  
PMID:[17028180](https://pubmed.ncbi.nlm.nih.gov/17028180/)
  28. Shen S, Wang G, Zhang R, Zhao Y, Yu H, Wei Y, Chen F. Development and validation of an immune gene-set based Prognostic signature in ovarian cancer. *EBioMedicine*. 2019; 40:318–26.  
<https://doi.org/10.1016/j.ebiom.2018.12.054>  
PMID:[30594555](https://pubmed.ncbi.nlm.nih.gov/30594555/)
  29. Shen C, Liu J, Wang J, Zhong X, Dong D, Yang X, Wang Y. Development and validation of a prognostic immune-associated gene signature in clear cell renal cell carcinoma. *Int Immunopharmacol*. 2020; 81:106274.  
<https://doi.org/10.1016/j.intimp.2020.106274>  
PMID:[32044664](https://pubmed.ncbi.nlm.nih.gov/32044664/)
  30. Yang S, Wu Y, Deng Y, Zhou L, Yang P, Zheng Y, Zhang D, Zhai Z, Li N, Hao Q, Song D, Kang H, Dai Z. Identification of a prognostic immune signature for cervical cancer to predict survival and response to immune checkpoint inhibitors. *Oncoimmunology*. 2019; 8:e1659094.  
<https://doi.org/10.1080/2162402X.2019.1659094>  
PMID:[31741756](https://pubmed.ncbi.nlm.nih.gov/31741756/)
  31. Guo D, Wang M, Shen Z, Zhu J. A new immune signature for survival prediction and immune checkpoint molecules in lung adenocarcinoma. *J Transl Med*. 2020; 18:123.  
<https://doi.org/10.1186/s12967-020-02286-z>  
PMID:[32143735](https://pubmed.ncbi.nlm.nih.gov/32143735/)
  32. Wang Z, Zhu J, Liu Y, Liu C, Wang W, Chen F, Ma L. Development and validation of a novel immune-related prognostic model in hepatocellular carcinoma. *J Transl Med*. 2020; 18:67.  
<https://doi.org/10.1186/s12967-020-02255-6>  
PMID:[32046766](https://pubmed.ncbi.nlm.nih.gov/32046766/)
  33. Koneva LA, Zhang Y, Virani S, Hall PB, McHugh JB, Chepeha DB, Wolf GT, Carey TE, Rozek LS, Sartor MA. HPV integration in HNSCC correlates with survival outcomes, immune response signatures, and candidate drivers. *Mol Cancer Res*. 2018; 16:90–102.  
<https://doi.org/10.1158/1541-7786.MCR-17-0153>  
PMID:[28928286](https://pubmed.ncbi.nlm.nih.gov/28928286/)
  34. Han Q, Wang R, Sun C, Jin X, Liu D, Zhao X, Wang L, Ji N, Li J, Zhou Y, Ye L, Liang X, Jiang L, et al. Human beta-defensin-1 suppresses tumor migration and invasion and is an independent predictor for survival of oral squamous cell carcinoma patients. *PLoS One*. 2014; 9:e91867.

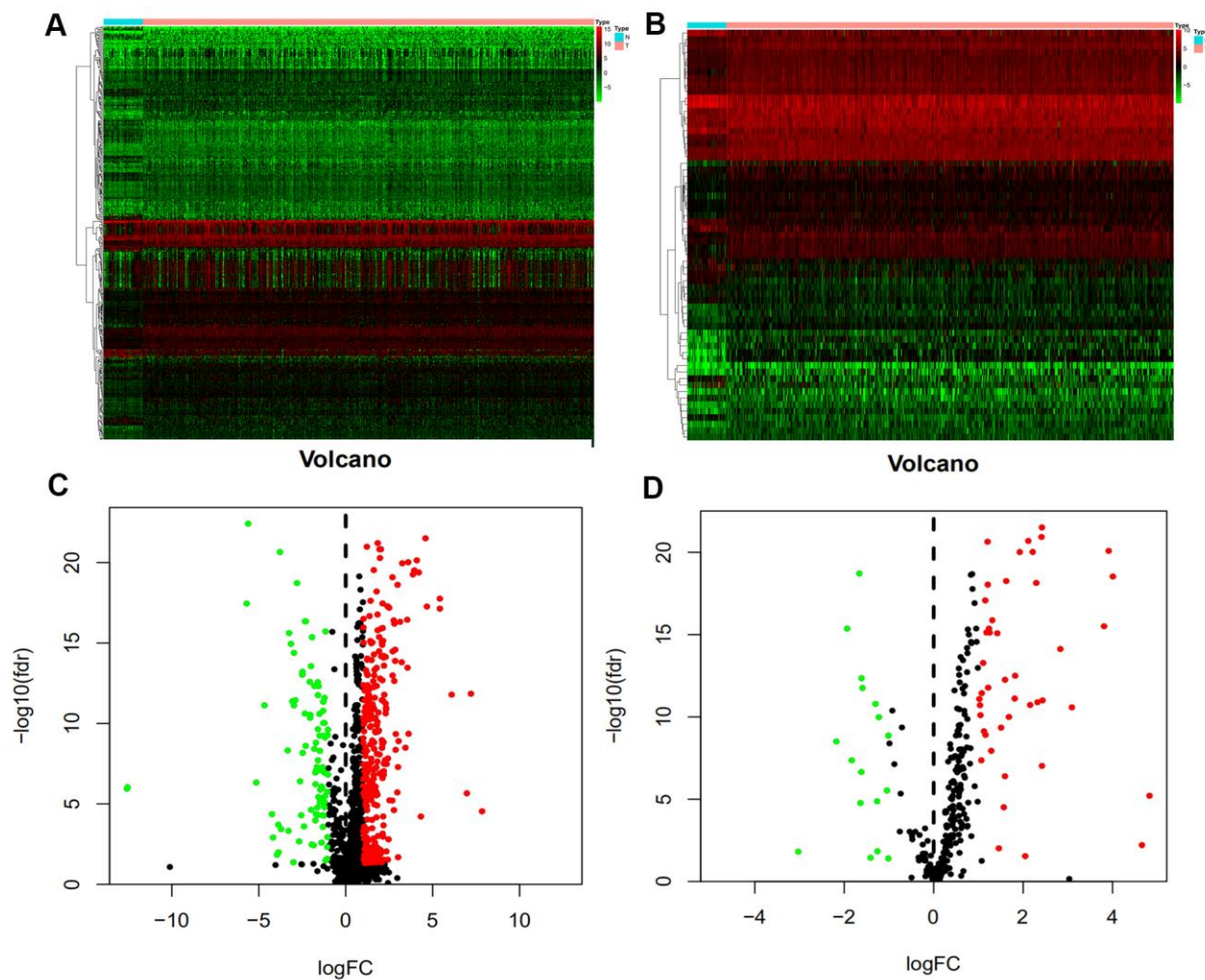


- <https://doi.org/10.1371/journal.pone.0091867>  
PMID:24658581
35. Pattani KM, Zhang Z, Demokan S, Glazer C, Loyo M, Goodman S, Sidransky D, Bermudez F, Jean-Charles G, McCaffrey T, Padhya T, Phelan J, Spivakovsky S, et al. Endothelin receptor type B gene promoter hypermethylation in salivary rinses is independently associated with risk of oral cavity cancer and premalignancy. *Cancer Prev Res (Phila)*. 2010; 3:1093–103.  
<https://doi.org/10.1158/1940-6207.CAPR-10-0115>  
PMID:20798208
36. Maia LL, Peterle GT, Dos Santos M, Trivilin LO, Mendes SO, de Oliveira MM, Dos Santos JG, Stur E, Agostini LP, Couto CV, Dalbó J, de Assis AL, Archanjo AB, et al. JMJD1A, H3K9me1, H3K9me2 and ADM expression as prognostic markers in oral and oropharyngeal squamous cell carcinoma. *PLoS One*. 2018; 13:e0194884.  
<https://doi.org/10.1371/journal.pone.0194884>  
PMID:29590186
37. Gao J, Ulekleiv CH, Halstensen TS. Epidermal growth factor (EGF) receptor-ligand based molecular staging predicts prognosis in head and neck squamous cell carcinoma partly due to deregulated EGF-induced amphiregulin expression. *J Exp Clin Cancer Res*. 2016; 35:151.  
<https://doi.org/10.1186/s13046-016-0422-z>  
PMID:27669890
38. Katase N, Gunduz M, Beder LB, Gunduz E, Al Sheikh Ali M, Tamamura R, Yaykasli KO, Yamanaka N, Shimizu K, Nagatsuka H. Frequent allelic loss of Dkk-1 locus (10q11.2) is related with low distant metastasis and better prognosis in head and neck squamous cell carcinomas. *Cancer Invest*. 2010; 28:103–10.  
<https://doi.org/10.3109/07357900903095680>  
PMID:19995224
39. Meng Y, Yu F. Long noncoding RNA FAM3D-AS1 inhibits development of colorectal cancer through NF- $\kappa$ B signaling pathway. *Biosci Rep*. 2019; 39:BSR20190724.  
<https://doi.org/10.1042/BSR20190724>  
PMID:31147452
40. Andrusiewicz M, Szczerba A, Wołuiń-Cholewa M, Warchoł W, Nowak-Markwitz E, Gąsiorowska E, Adamska K, Jankowska A. CGB and GNRH1 expression analysis as a method of tumor cells metastatic spread detection in patients with gynecological malignancies. *J Transl Med*. 2011; 9:130.  
<https://doi.org/10.1186/1479-5876-9-130>  
PMID:21827674
41. Li T, Qin Y, Zhen Z, Shen H, Cong T, Schiferle E, Xiao S. Long non-coding RNA HOTAIR/microRNA-206 sponge regulates STC2 and further influences cell biological functions in head and neck squamous cell carcinoma. *Cell Prolif*. 2019; 52:e12651.  
<https://doi.org/10.1111/cpr.12651>  
PMID:31297902
42. Johnston AJ, Murphy KT, Jenkinson L, Laine D, Emmrich K, Faou P, Weston R, Jayatilake KM, Schloegel J, Talbo G, Casey JL, Levina V, Wong WW, et al. Targeting of Fn14 prevents cancer-induced cachexia and prolongs survival. *Cell*. 2015; 162:1365–78.  
<https://doi.org/10.1016/j.cell.2015.08.031>  
PMID:26359988
43. Mocellin S, Nitti D. CTLA-4 blockade and the renaissance of cancer immunotherapy. *Biochim Biophys Acta*. 2013; 1836:187–96.  
<https://doi.org/10.1016/j.bbcan.2013.05.003>  
PMID:23748107
44. Shang B, Liu Y, Jiang SJ, Liu Y. Prognostic value of tumor-infiltrating FoxP3+ regulatory T cells in cancers: a systematic review and meta-analysis. *Sci Rep*. 2015; 5:15179.  
<https://doi.org/10.1038/srep15179> PMID:26462617
45. Seminerio I, Descamps G, Dupont S, de Marrez L, Laigle JA, Lechien JR, Kindt N, Journe F, Saussez S. Infiltration of FoxP3+ regulatory T cells is a strong and independent prognostic factor in head and neck squamous cell carcinoma. *Cancers (Basel)*. 2019; 11:227.  
<https://doi.org/10.3390/cancers11020227>  
PMID:30781400
46. Kumar AT, Knops A, Swendseid B, Martinez-Outschoom U, Harshyne L, Philp N, Rodeck U, Luginbuhl A, Cognetti D, Johnson J, Curry J. Prognostic significance of tumor-associated macrophage content in head and neck squamous cell carcinoma: a meta-analysis. *Front Oncol*. 2019; 9:656.  
<https://doi.org/10.3389/fonc.2019.00656>  
PMID:31396482
47. Nishino M, Ramaiya NH, Hatabu H, Hodi FS. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol*. 2017; 14:655–68.  
<https://doi.org/10.1038/nrclinonc.2017.88>  
PMID:28653677
48. Chen SW, Li SH, Shi DB, Jiang WM, Song M, Yang AK, Li YD, Bei JX, Chen WK, Zhang Q. Expression of PD-1/PD-L1 in head and neck squamous cell carcinoma and its clinical significance. *Int J Biol Markers*. 2019; 34:398–405.  
<https://doi.org/10.1177/1724600819884722>  
PMID:31674884
49. Schneider S, Kadletz L, Wiebringhaus R, Kenner L, Selzer E, Füreder T, Rajky O, Berghoff AS, Preusser M,

- Heiduschka G. PD-1 and PD-L1 expression in HNSCC primary cancer and related lymph node metastasis - impact on clinical outcome. *Histopathology*. 2018; 73:573–84.  
<https://doi.org/10.1111/his.13646>  
PMID:[29742291](https://pubmed.ncbi.nlm.nih.gov/29742291/)
50. Joly S, Compton LM, Pujol C, Kurago ZB, Guthmiller JM. Loss of human beta-defensin 1, 2, and 3 expression in oral squamous cell carcinoma. *Oral Microbiol Immunol*. 2009; 24:353–60.  
<https://doi.org/10.1111/j.1399-302X.2009.00512.x>  
PMID:[19702947](https://pubmed.ncbi.nlm.nih.gov/19702947/)
51. Canel M, Secades P, Garzón-Arango M, Allonca E, Suarez C, Serrels A, Frame M, Brunton V, Chiara MD. Involvement of focal adhesion kinase in cellular invasion of head and neck squamous cell carcinomas via regulation of MMP-2 expression. *Br J Cancer*. 2008; 98:1274–84.  
<https://doi.org/10.1038/sj.bjc.6604286>  
PMID:[18349846](https://pubmed.ncbi.nlm.nih.gov/18349846/)
52. Hayes AJ, Melrose J. Keratan sulphate in the tumour environment. *Adv Exp Med Biol*. 2020; 1245:39–66.  
[https://doi.org/10.1007/978-3-030-40146-7\\_2](https://doi.org/10.1007/978-3-030-40146-7_2)  
PMID:[32266652](https://pubmed.ncbi.nlm.nih.gov/32266652/)
53. Iida J, Dorchak J, Clancy R, Slavik J, Ellsworth R, Katagiri Y, Pugacheva EN, van Kuppevelt TH, Mural RJ, Cutler ML, Shriver CD. Role for chondroitin sulfate glycosaminoglycan in NEDD9-mediated breast cancer cell growth. *Exp Cell Res*. 2015; 330:358–70.  
<https://doi.org/10.1016/j.yexcr.2014.11.002>  
PMID:[25445787](https://pubmed.ncbi.nlm.nih.gov/25445787/)
54. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015; 43:e47.  
<https://doi.org/10.1093/nar/gkv007>  
PMID:[25605792](https://pubmed.ncbi.nlm.nih.gov/25605792/)
55. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16:284–87.  
<https://doi.org/10.1089/omi.2011.0118>  
PMID:[22455463](https://pubmed.ncbi.nlm.nih.gov/22455463/)
56. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13:2498–504.  
<https://doi.org/10.1101/gr.1239303>  
PMID:[14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
57. Frost HR, Amos CI. Gene set selection via LASSO penalized regression (SLPR). *Nucleic Acids Res*. 2017; 45:e114.  
<https://doi.org/10.1093/nar/gkx291>  
PMID:[28472344](https://pubmed.ncbi.nlm.nih.gov/28472344/)

## SUPPLEMENTARY MATERIALS

### Supplementary Figure



**Supplementary Figure 1. Differentially expressed IRGs and TFs in HNSCC.** (A, B) Heat map of significantly differentially IR and TF genes in HNSCC. The color from green to red refers to the progression from low to high expressed genes. (C, D) Volcano plot of differentially expressed IRGs and TFs. The red and green dots in the plot represent significantly upregulated and downregulated genes, respectively, and no differential expression is depicted by black dots.

## Supplementary Tables

**Supplementary Table 1. The clinicopathological characteristics of HNSCC patients from the training and validation sets.**

<b>Clinicopathologic parameters</b>	<b>Training set (n = 461) number of cases (%)</b>	<b>Validation set (n = 96) number of cases (%)</b>
<b>Gender</b>		
Male	339 (26.46)	65 (67.71)
Female	122 (73.54)	31 (32.29)
<b>Age</b>		
<60 y	202 (43.82)	50 (52.08)
≥60 y	259 (56.18)	46 (47.92)
<b>TNM stage</b>		
I-II	104 (22.56)	41 (42.71)
III-IV	357 (77.44)	55 (57.29)
<b>Drinking</b>		
Yes	328 (71.11)	
No	133 (28.89)	
<b>Smoking</b>		
Yes	106 (23.00)	
No	355 (77.00)	
<b>Grade</b>		
1-2	340 (73.75)	
3-4	121 (26.25)	
<b>Lymph node metastasis</b>		
N0	226 (50.98)	
N+	235 (49.02)	
<b>T classification</b>		
T1-T2	161 (34.92)	
T3-T4	300 (65.08)	
<b>M classification</b>		
M0	456 (98.92)	
M1	5 (1.08)	
<b>Recurrence</b>		
Yes	162 (35.14)	
No	299 (64.86)	

**Supplementary Table 2. The association between the prognosis immune-related genes and differentially expressed TFs.**

<b>TF</b>	<b>ImmuneGene</b>	<b>Correlation coefficients</b>	<b>P-value</b>	<b>Regulation</b>
CBX3	BIRC5	0.518642408820594	1.15520246900326e-31	positive
CENPA	BIRC5	0.69287091986189	3.39740962361839e-64	positive
EZH2	GNRH1	0.523031102481409	2.89994062578473e-32	positive
FOXP3	LTA	0.735231898437069	5.50266657984769e-76	positive
FOXP3	CXCR4	0.602664676810251	7.50120067401527e-45	positive
FOXP3	CXCR3	0.734367694212062	1.00840545610555e-75	positive
FOXP3	IL21R	0.797986384807953	2.2645404903834e-98	positive
FOXP3	IL2RG	0.664185745269767	2.60811386532034e-57	positive
FOXP3	CD247	0.55309782116114	1.27354006034913e-36	positive
FOXP3	ZAP70	0.532205241805523	1.50989089544079e-33	positive
FOXP3	SH2D1A	0.683751210147642	6.39922931062554e-62	positive
FOXP3	ICOS	0.859952634007701	5.02642593010807e-130	positive
FOXP3	CTLA4	0.733274069140433	2.16300325109839e-75	positive
H2AFX	BIRC5	0.534652159139509	6.7600337274877e-34	positive
MYBL2	BIRC5	0.563377513450786	3.24548208832276e-38	positive
PBX1	NR3C2	0.601407493116652	1.26429217741422e-44	positive
POU5F1	GNRH1	0.609650592416114	3.95154070696453e-46	positive