

# A novel immune-related long non-coding RNA signature improves the prognosis prediction in the context of head and neck squamous cell carcinoma

Lin Chen<sup>a\*</sup>, Zhimou Cai<sup>a\*</sup>, Kexing Lyu<sup>a\*</sup>, Zhiwei Cai<sup>b</sup>, and Wenbin Lei<sup>a</sup>

<sup>a</sup>Department of Otolaryngology, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, P.R. China; <sup>b</sup>Guangzhou Brain Hospital, Guangzhou Medical University, Guangzhou, Guangdong, P.R. China

## ABSTRACT

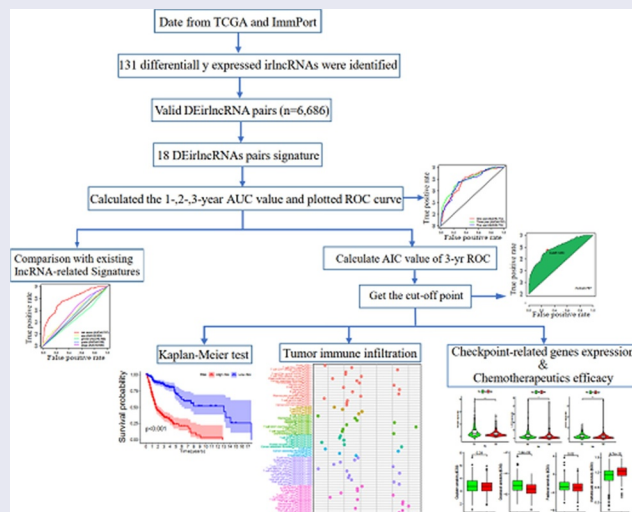
The tumor immune microenvironment plays an important role in head and neck squamous cell carcinoma (HNSCC). Reliable prognostic signatures able to accurately predict the immune landscape and survival rate of HNSCC patients are crucial to ensure an individualized/effective treatment. Here, we used HNSCC transcriptomic and clinical data retrieved from The Cancer Genome Atlas and identified differentially expressed immune-related long non-coding RNAs (DElncRNAs). DElncRNA pairs were recognized using univariate analysis. Cox and Lasso regression analyses were used to determine the association between DElncRNA pairs and the patients' overall survival and build the prediction model. Receiver operating characteristic curves and Kaplan–Meier survival curves were used to validate the prediction model. We then reevaluated the model based on the clinical factors, tumor-infiltrating immune cells, chemotherapeutic efficacy, and immunosuppression biomarkers. We built a risk score model based on 18 DElncRNA pairs, closely related to the overall survival of patients (hazard ratio: 1.376; 95% confidence interval: 1.302–1.453;  $P < 0.0001$ ). Compared with two recently published lncRNA signatures, our DElncRNA pair signature had a higher area under the curve, indicating better prognostic performance. Additionally, the signature score positively correlated with aggressive HNSCC outcomes (low immunity score, significantly reduced CD8 + T cell infiltration, and low expression of immunosuppression biomarkers). However, high-risk patients might have high chemosensitivity. Overall, the lncRNAs signature established here shows promising clinical prediction and the effective disclosure of the tumor immune microenvironment in HNSCC patients; therefore, such signature might help distinguish patients that could benefit from immunotherapy.



## ARTICLE HISTORY

Received 20 April 2021  
Revised 3 June 2021  
Accepted 3 June 2021


## KEYWORDS

Head and neck squamous cell carcinoma; long non-coding RNAs; tumor-infiltrating immune cells; prognosis; TCGA



**CONTACT** Wenbin Lei  [leiwb@mail.sysu.edu.cn](mailto:leiwb@mail.sysu.edu.cn)  Department of Otolaryngology, The First Affiliated Hospital of Sun Yat-sen University, No. 58 Zhongshan Er Road, Guangzhou, Guangdong 510080, China

\*These authors contributed equally to this work as co-first authors.

 Supplemental data for this article can be accessed [here](#)

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Introduction

Head and neck tumors are the sixth most common type of cancer worldwide; head and neck squamous cell carcinoma (HNSCC) is the most common histological type [1]. Of note, HNSCC encompasses a heterogeneous group of tumors that arise in the oral cavity, larynx, and pharynx, mainly associated with tobacco and alcohol consumption, and human papillomavirus infection [2–4]. However, because the early symptoms of HNSCC are not obvious, the 5-year survival rate is below 50%. Furthermore, HNSCC is characterized by a high lymph node metastasis rate, and local recurrence and metastasis can reduce the survival rate to 35% [5]. Currently, HNSCC is estimated to have a global annual incidence and mortality rate of 900000 cases and 450000 deaths, respectively [1].

HNSCC is a disease characterized by profound immunosuppression [6]. Thus, immune checkpoint inhibitor (ICI) therapy is a new hope for HNSCC patients. In fact, recently, considerable progress has been made in ICI-based HNSCC treatment. However, the response rate of recurrent or metastatic HNSCC to PD-1/PD-L1 inhibitors is still low, in the range of 13.3–22% as per previous clinical trials [6]. Therefore, the identification of biomarkers that can effectively predict the efficacy of ICIs is crucial for patient selection. Generally, the expression of PD-L1 is used as the representative marker to predict the efficacy of ICIs. However, for most tumors, the detection of PD-L1 expression alone is not sufficient [7–9].

Long non-coding RNAs (lncRNAs) are functional RNA molecules with a length of more than 200 nucleotides that have minimal or no protein-encoding role [10]. It is increasingly recognized that lncRNAs can interact with DNA, RNA, and the proteins that regulate gene expression at the transcriptional and post-transcriptional levels. They are also associated with various types of cancer [10,11]. Recently, several lncRNA signatures have been established for prognostic prediction in some cancers, including HNSCC. For instance, Mao et al. [12] identified an eight-lncRNA signature that is an independent prognostic factor in HNSCC patients. Additionally, Bao

et al. [13] identified genome instability-associated lncRNAs and constructed a signature to evaluate the prognosis of breast cancer patients.

The diagnosis, evaluation, and treatment of cancer are closely related to the tumor immune microenvironment, especially to the infiltrating immune cells [14–16]. Recent research has confirmed that lncRNAs are critical regulators of gene expression in the immune system; lncRNAs direct the expression of genes related to the development/activation of diverse immune cells, impacting tumor immune cell infiltration and, consequently, the immune microenvironment [17,18]. Therefore, we hypothesized that immune-related lncRNAs (irlncRNAs) influence immune cell infiltration in HNSCC, and, therefore, the treatment response and prognosis. In this study, we aimed to identify irlncRNAs to i) investigate the relationship between irlncRNAs and immune cell infiltration, ii) construct a powerful prognostic model in order to assess the risk in HNSCC, and iii) predict the response of HNSCC patients to immunotherapy.

## Materials and methods

### Data collection

HNSCC RNAseq expression profile data were downloaded from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>); tumor and normal tissue sequencing information from 501 patients with HNSCC and 44 healthy individuals were retrieved, together with the clinical and follow-up data in the context of the HNSCC patients. Valid data were extracted after the elimination of data from patients with a follow-up time of no more than 30 days and of duplicated data. ‘Ensembl\_Stable\_id’ was converted to ‘Gene Symbol’ in the RNAseq expression profile according to the GTF file downloaded from Ensembl (<http://asia.ensembl.org>) and annotated to distinguish the mRNAs and lncRNAs for subsequent analysis. A list of immune-related genes (ir-genes) was downloaded from the ImmPort database (<http://www.immport.org>), and correlation analyses were conducted between the ir-genes and all lncRNAs to screen for irlncRNAs.

Immune gene correlation coefficients higher than 0.4 or lower than -0.4 and with a *P* value lower than 0.001 were considered as indicative of irlncRNAs. The differentially expressed irlncRNAs (DEirlncRNAs) were identified through the ‘limma’ R package; irlncRNAs with a false discovery rate < 0.05 and  $|\log_2FC| > 1.5$  were defined as DEirlncRNAs.

### **DEirlncRNAs pairing**

We used a novel modeling algorithm (with pairing and iteration), to construct a two-biomarker combinations-based irlncRNA signature to improve the accuracy of the cancer prognostic model. The DEirlncRNAs were cyclically singly paired, and a 0-or-1 matrix was constructed. If the expression of the former lncRNA was higher than that of the latter, the matrix was defined as 1; otherwise, it was defined as 0. The constructed 0-or-1 matrix was further screened. A valid match was defined when the number of DEirlncRNA pairs, of which the expression was set as 0 or 1, accounted for more than 20% or less than 80% of the total number of pairs.

### **Building the prognostic DEirlncRNA pair-based signature**

Univariate Cox analysis was performed using the ‘survival’ R package to screen DEirlncRNA pairs of prognostic value. DEirlncRNA pairs with *P* < 0.01 were considered significantly related to the prognosis of HNSCC. Next, Lasso regression analysis, performed using the ‘glmnet’ and ‘survival’ R packages, was used to select the minimum error value to eliminate overfitting of the data. The optimal penalty parameter ‘lambda’ value was calculated via 1000 cross-validations [19]. Then, multivariate analyses were performed using the Cox regression model to identify the prognostic DEirlncRNA pairs in HNSCC. Based on the coefficients from the multivariate regression analysis and the status of the DEirlncRNA pairs, we constructed a DEirlncRNA pair signature (DEirlnc-PSig) for the prediction of the clinical outcome, as follows:

$$\text{DEirlnc} - \text{PSig}(\text{patient}) = \sum_{i=1}^n \beta_i * P_i$$

where  $\beta_i$  indicates the coefficient for each DEirlncRNA pair and  $P_i$  indicates the DEirlncRNA pair status on the 0-or-1 matrix.

### **Validation of the constructed risk-assessment model**

A time-dependent receiver operating characteristic (ROC) curve and the respective area under the curve (AUC) were obtained using the ‘survivalROC’ R package to predict the prognosis accuracy of DEirlnc-PSig in patients with HNSCC. The 1-, 2-, and 3-year ROC curves and the AUCs of the models were plotted. The Akaike information criterion (AIC) values for each data point of the 3-year ROC curve were also estimated to identify the maximum inflection point, considered as the cutoff value for the separation of patients into the high-risk and low-risk groups, with high or low DEirlnc-PSig expression, respectively [20].

The Kaplan–Meier method was used to construct survival curves, and the log-rank test was used to assess the difference in the survival between the high- and low-risk groups with a significance level of 0.05. The R packages used in these steps were ‘survival’ and ‘survminer’. To verify the clinical application of DEirlnc-PSig, multivariate Cox regression and stratified analyses were further used, to assess the independence of DEirlnc-PSig from other key clinical factors. A forest map was drawn to demonstrate the results using again the R packages ‘survival’ and ‘survminer’. In addition, we compared our defined DEirlnc-PSig with two previously published lncRNA signatures to further assess the performance of our prognostic model.

### **RNA preparation, cDNA synthesis, and qRT-PCR validation**

Total RNA from frozen tissue specimens was extracted using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer’s protocol. The RNA quantity and quality were determined using a NanoDrop 2000

spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Reverse transcription was conducted using the Prime Script RT Master Mix Kit (TAKARA BIO INC, Japan) according to the manufacturer's protocol. qRT-PCR was performed using the SYBR Green Premix Pro Taq HS qPCR Kit (Agbio, Hunan, China) on the LightCycle480 II system (Roche Applied Science, Basel, Switzerland). The relative quantification of each gene was achieved after normalization to the expression levels of *GAPDH* using the comparative CT method. The mean Ct value of each gene minus the mean Ct value of *GAPDH* was defined as the  $\Delta$  Ct. The primer sets used in this study are listed in Table S1. Since, as mentioned above, the risk model in this study was constructed based on two-biomarker combinations, we evaluated patient risk via the comparison of the  $\Delta$  Ct value of the matched genes.

### **Investigation of tumor-infiltrating immune cells**

To explore the differences in immune cell infiltration between the different risk groups, we used the xCell [21], TIMER [22], QUANTISEQ [23], MCPcounter [24], EPIC [25], CIBERSORT-abs [26], and CIBERSORT [27] methods in the context of the HNSCC dataset (TCGA project). Differences in the numbers of distinct immune-infiltrating cell types between the high- and low-risk groups in the constructed model were analyzed using the Wilcoxon's signed-rank test. The correlation between the risk scores and immune cell infiltration was evaluated using Spearman's correlation analysis. The significance threshold was set as  $P < 0.05$ . The 'limma', 'scales', 'ggplot2', 'ggtext', and 'ggpubr' R packages were used for the above operations.

### **Exploration of the significance of the model for clinical treatment**

Using the 'pRRophetic' R package, we calculated the half inhibitory concentration ( $IC_{50}$ ) of the commonly administered chemotherapeutic drugs in the HNSCC dataset (TCGA project) to evaluate the utility of the model for the clinical treatment of HNSCC. Anti-tumor drugs such as cisplatin, docetaxel, methotrexate, and paclitaxel are

recommended for the treatment of HNSCC as per the National Comprehensive Cancer Network guidelines [28]. The difference in the  $IC_{50}$  between the high- and low-risk groups was compared using the Wilcoxon's signed-rank test.

### **Analysis of the expression of immunosuppression biomarkers in the context of ICIs**

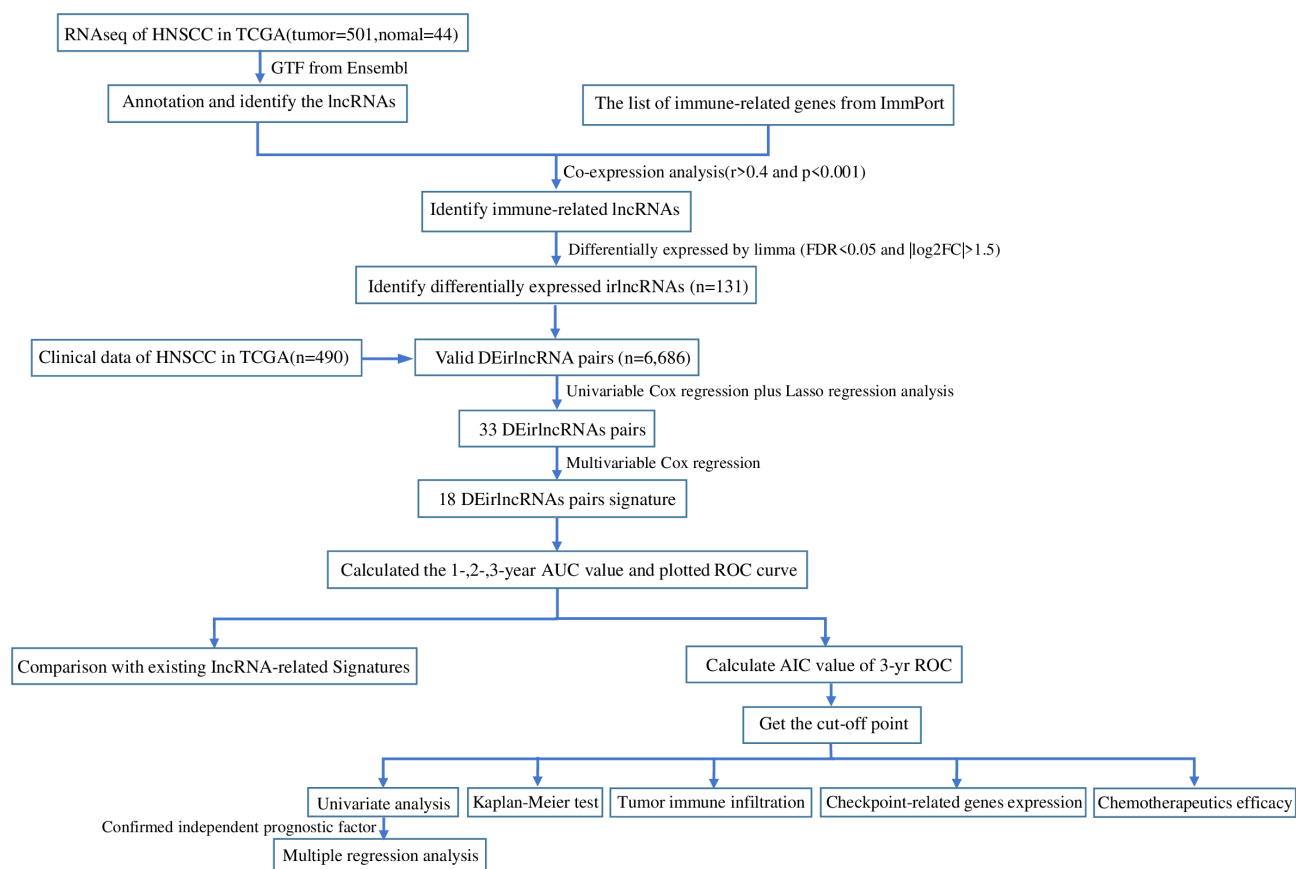
The relationship between the constructed model and the expression of genes related to ICIs was analyzed using the 'limma' and 'ggpubr' R packages and visualized in the form of violin plots.

## **Results**

LncRNAs are important factors for the prognosis of HNSCC patients. We hypothesized that immune-related lncRNAs (irlncRNAs) influence immune cell infiltration, treatment response, and prognosis in the context of HNSCC patients. In this study, we aimed to investigate the relationship between irlncRNAs and immune cell infiltration, and to construct a powerful prognostic model to assess HNSCC risk and predict the response of patients to immunotherapy. We identified DEirlncRNAs using an HNSCC cohort retrieved from TCGA. DEirlncRNA pairs were recognized using univariate analysis. Cox and Lasso regression analyses were also used to identify potential DEirlncRNA pairs related to the OS and build the prediction model. Overall, the prediction model established here reflects the HNSCC tumor immune microenvironment and can be used to predict HNSCC sensitivity to different treatment agents.

### **Identification of irlncRNAs and DEirlncRNAs**

Figure 1 presents the study design. A total of 805 irlncRNAs were identified after the co-expression analysis of ir-genes and lncRNAs. Differential expression analysis was subsequently performed using the 'limma' R package, and 132 DEirlncRNAs (116 upregulated and 16 downregulated) were identified between the tumor and normal tissues (Figure S1).



**Figure 1.** Flow chart of this study.

### **Establishment of the DEirlncRNA pairs and DEirlncRNA-pair signature**

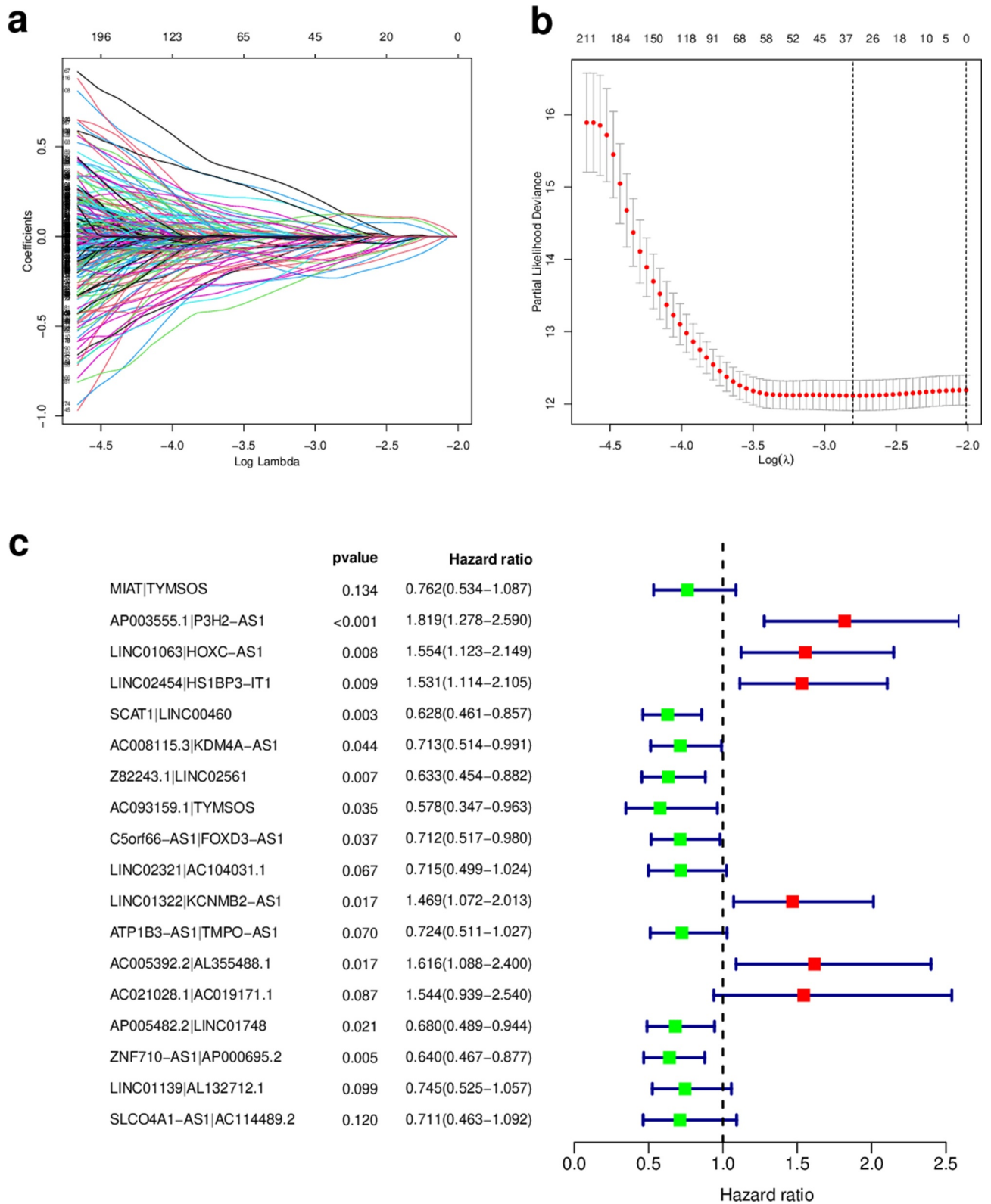
Using an iteration loop and the 0-or-1 matrix screening based on the 132 DEirlncRNAs, 6686 valid DEirlncRNA pairs were identified. After univariate Cox regression analysis, 500 survival-related DEirlncRNA pairs were selected ( $P < 0.01$ ). To further avoid data overfitting, Lasso regression analysis was employed to screen out 33 DEirlncRNA pairs that showed the highest correlation with survival (Figure 2(a,b)). These 33 DEirlncRNA pairs were further subjected to multivariate Cox regression analysis, and the 18 most significant DEirlncRNA pairs were identified and used for the development of the DEirlncRNA pair signature for HNSCC patients using the stepwise elimination method (Figure 2(c)). Of note, when we applied the 1-, 2-, and 3-year ROC curves to verify the predictive capacity of the DEirlnc-PSig the AUCs were almost all  $\sim 0.75$ , suggesting a satisfactory predictive performance (Figure 3(a)). Importantly, the

predictive capacity of the DEirlnc-PSig was superior to that of other clinical features (Figure 3(b)).

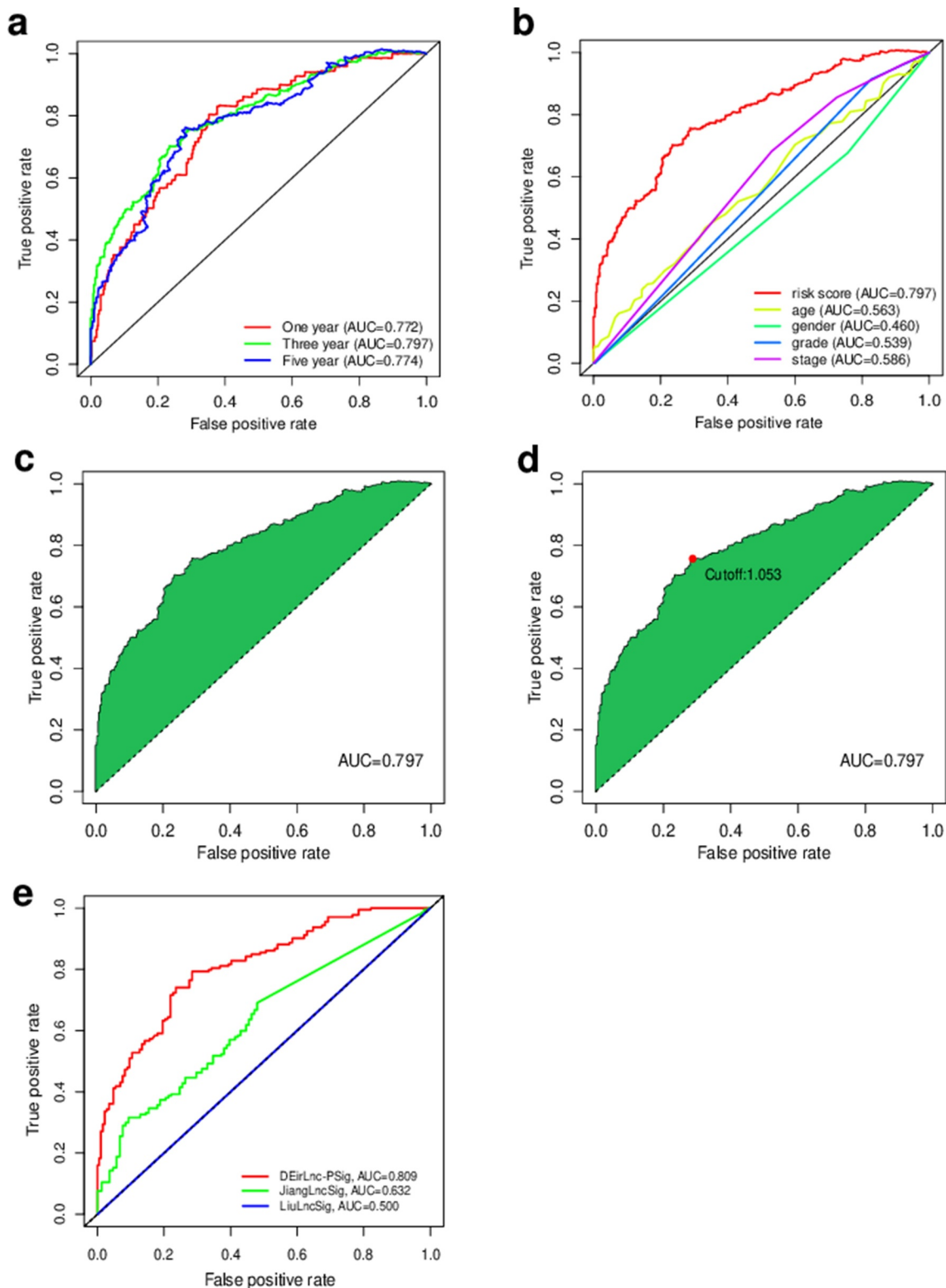
### **Clinical evaluation using the risk assessment model and analysis of the independence of the DEirlnc-PSig from other clinical factors**

We defined the maximum inflection point as the cutoff value in the 3-year ROC curve using the AUC values (Figure 3(c,d)). HNSCC patients (TCGA cohort) were then divided into the low-risk ( $n = 255$ ) and high-risk ( $n = 235$ ) groups. Interestingly, when we compared the survival in the two groups we found that the DEirlnc-PSig possessed a powerful ability to predict the differences in the prognosis of HNSCC. In fact, the OS was significantly better in the low-risk *versus* high-risk groups ( $P < 0.001$ , log-rank test; Figure 4(a-c)).

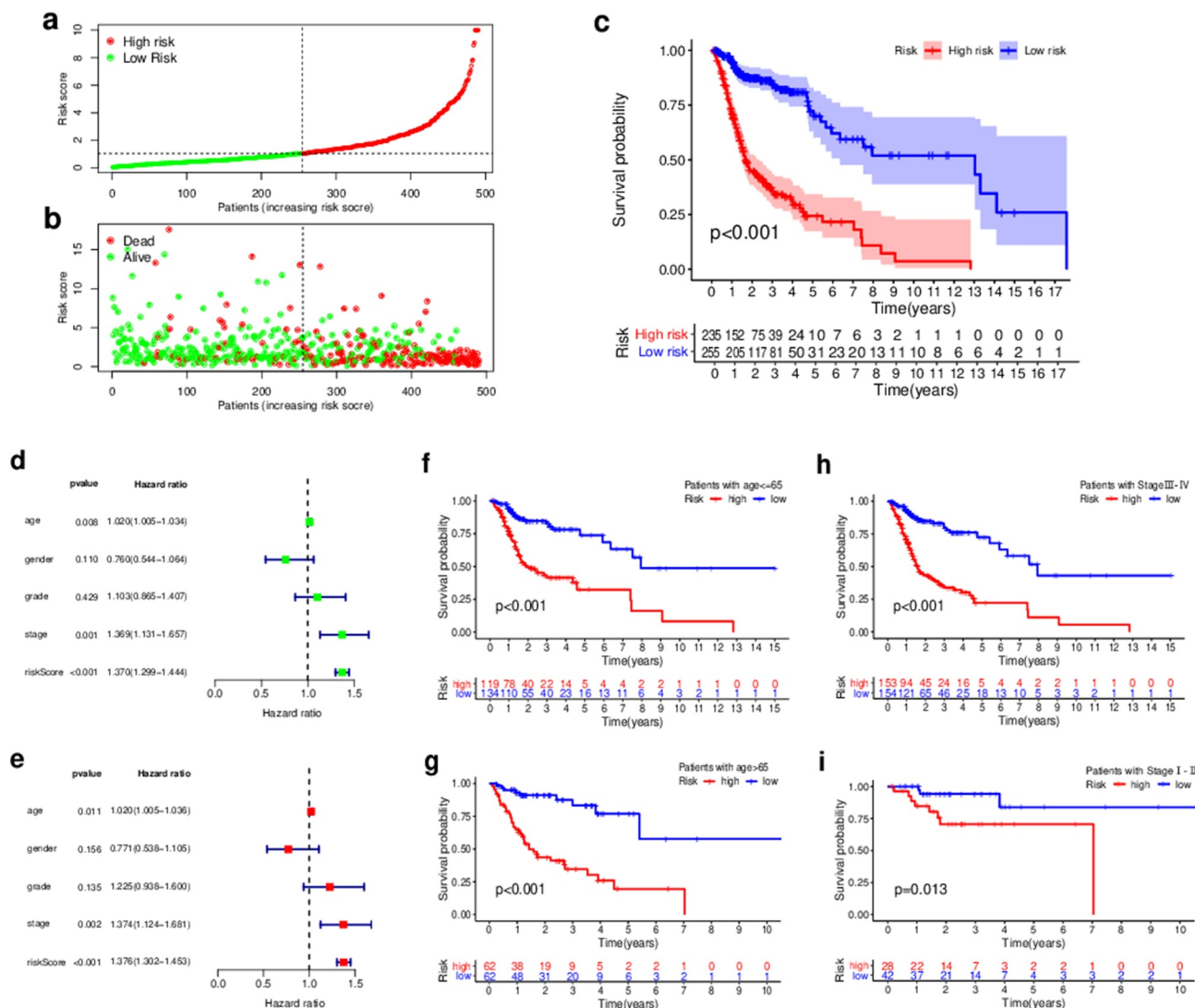
Next, to evaluate whether the prognostic value of the DEirlnc-PSig was independent of the common clinical features, univariate and multivariate



**Figure 2.** Diagrams of the lasso and multivariable Cox regression analyses. (a) Lasso regression analysis. (b) The penalty coefficient was used to minimize the mean square error of the models. (c) A forest graph of the 18 DE lncRNA pairs identified via multivariable Cox regression analysis.



**Figure 3.** Prediction of the prognosis accuracy of DEirLnc-PSig in patients with HNSCC. (a) The 1-, 2-, and 3-year receiver operating characteristic (ROC) curves of the DEirLnc-PSig. (b) A comparison of the 3-year ROC curves with other common clinical characteristics showed the superiority of the risk score. (c, d) The maximum inflection point obtained using the Akaike information criterion. (e) ROC analysis with respect to the 3-year overall survival for DEirLnc-PSig, JiangLncSig, and LiuLncSig.



**Figure 4.** Prediction of the survival of HNSCC patients and identification of the risk factors using the risk score model and univariate and multivariate Cox analyses, respectively. (a, b) Patients were scored and grouped into a low-risk group (green) and a high-risk group (red). Then, scatter diagrams of the survival rate of patients were plotted based on the risk scores from low to high, with green indicating live patients and red indicating deaths. (c) Kaplan–Meier survival curves. (d) Univariate analysis. (e) Multivariate analysis. (f, g) Kaplan–Meier curve analysis of the overall survival (OS) in the high- and low-risk groups, in the context of young and old patients. (h, i) Kaplan–Meier curve analysis of the OS in the high- and low-risk groups in the context of early-stage and late-stage patients. Statistical analysis was performed using the log-rank test.

Cox regression analyses were performed using age, gender, pathologic grade, and clinical stage data, and our DEirLnc-PSig-based prognostic risk score model (Figure 4(d,e)). Importantly, the risk score model showed statistical differences, suggesting it can be used as an independent prognostic predictor. Of note, there were two other clinical factors, age, and clinical stage, defined as significant in the multivariate analysis. Therefore, a stratification analysis was performed to determine whether the DEirLnc-PSig possessed a prognostic value independent of the age and clinical stage. Patients

(TCGA dataset) were stratified into the young ( $\leq 65$  years,  $n = 124$ ) and old ( $> 65$  years,  $n = 253$ ) groups, followed by the subsequent categorization into the high- or low-risk groups using the DEirLnc-PSig. Importantly, there was a still significant difference in the OS between the high- and low-risk groups in both the young ( $P < 0.001$ , log-rank test) and old ( $P < 0.001$ , log-rank test; Figure 4(f, g) patient groups. Additionally, all HNSCC patients were stratified into the early-(stages I and II,  $n = 70$ ) or late-stage (stages III and IV,  $n = 307$ ) groups, followed by the same



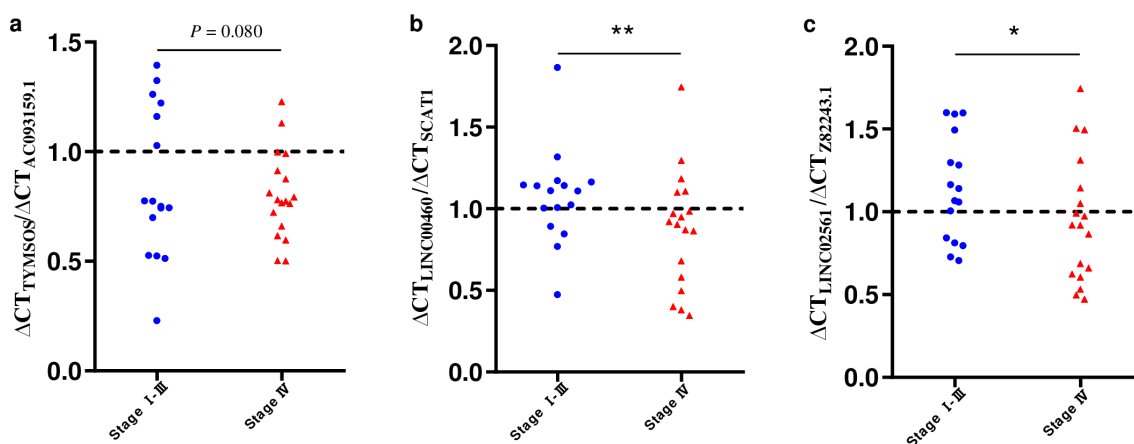
categorization into the high- or low-risk groups using the DEirLnc-PSig. Again, there was a significant difference in the OS between the high- and low-risk groups in both the early-stage and late-stage groups ( $P = 0.013$  and  $P < 0.001$ , respectively, log-rank test; Figure 4(h,i)). Therefore, altogether, these results indicate that the DEirLnc-PSig is an independent prognostic factor associated with the OS of HNSCC patients.

### Comparison of the performances of the DEirLnc-PSig and other existing lncRNA-based signatures in the prediction of survival

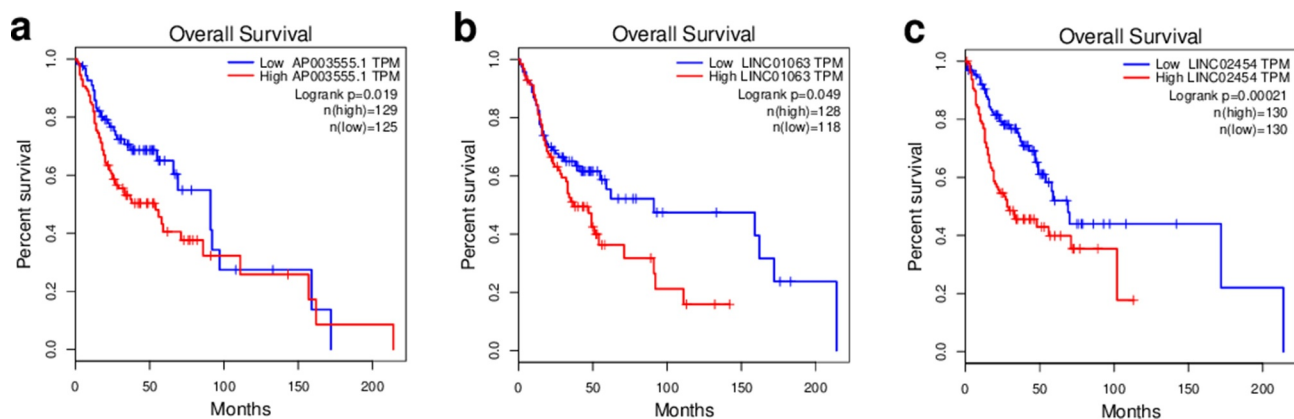
We further compared the predictive performance of the DEirLnc-PSig with that of two recently published lncRNA signatures: the 3-lncRNA signature derived proposed by the study by Jiang *et al* (hereafter referred to as JiangLncSig) [29] and the 5-lncRNA signature proposed by Liu *et al* (hereafter referred to as LiuLncSig) [30] using the same TCGA patient cohort. As shown in Figure 3(e), the AUC of the 3-year ROC curve for the DEirLnc-PSig (AUC = 0.809) is significantly higher than those in the context of the JiangLncSig (AUC = 0.632) and LiuLncSig (AUC = 0.5). These results clearly highlight that the DEirLnc-PSig has better prognostic performance (with respect to the OS) than the two recently published lncRNA signatures.

### Validation of the gene signature using qRT-PCR

Next, using qRT-PCR, we evaluated the expression of the gene pairs comprising the top 5 coefficients in the context of the DEirLnc-PSig, in 34 HNSCC tissues. According to tumor tissue sources, the 34 HNSCC tissues were divided into two groups: stage I-III and stage IV. The qRT-PCR results showed that the ratio less than 1 of  $\Delta Ct_{AC093159.1}/\Delta Ct_{TYMSOS}$ ,  $\Delta Ct_{SCAT1}/\Delta Ct_{LINC00460}$  and  $\Delta Ct_{Z82243.1}/\Delta Ct_{LINC02561}$  (indicating the expression of the former lncRNA was higher than that of the latter), stage I-III group were significantly more than those in the stage IV group ( $P = 0.08$ ,  $P < 0.01$  and  $P < 0.05$ , respectively, chi-square test). Of note, since the Ct value is inversely proportional to the gene expression levels, in order to facilitate display, as shown in Figure 5(a-c), we used the reciprocal of the paired gene Ct value to represent the expression levels. On the other hand, for risk factors, such as AP003555.1/P3H2-AS1, LINC01063/HOXC-AS1 and LINC02454/HS1BP3-IT1, there were no statistical differences, as per the PCR results. Still, in order to further verify the accuracy of the model, gene expression profiling interaction analysis (GEPIA) (<http://gepia.cancer-pku.cn/>) was used performed in the context of survival [31]. Importantly, and as expected, a lower expression of P3H2-AS1, HOXC-AS1, and HS1BP3-IT1 was associated with a longer OS (Figure 6(a-c);  $P = 0.019$ ,  $P < 0.05$  and  $P < 0.001$ , respectively).



**Figure 5.** qRT-PCR analysis of the two-lncRNA combinations in HNSCC tissues at different developmental stages. (a)  $\Delta Ct_{AC093159.1}/\Delta Ct_{TYMSOS}$ . (b)  $\Delta Ct_{SCAT1}/\Delta Ct_{LINC00460}$ . (c)  $\Delta Ct_{Z82243.1}/\Delta Ct_{LINC02561}$ . \*\* $P < 0.01$ , \* $P < 0.05$ .



**Figure 6.** Overall survival analysis of HNSCC patients in the context of the expression of different risk factors. The GEPIA website was used. (a) *P3H2-AS1*, (b) *HOXC-AS1*, and (c) *HS1BP3-IT1*.

### Estimation of tumor-infiltrating immune cells and the expression of immunosuppression biomarkers using the risk assessment model

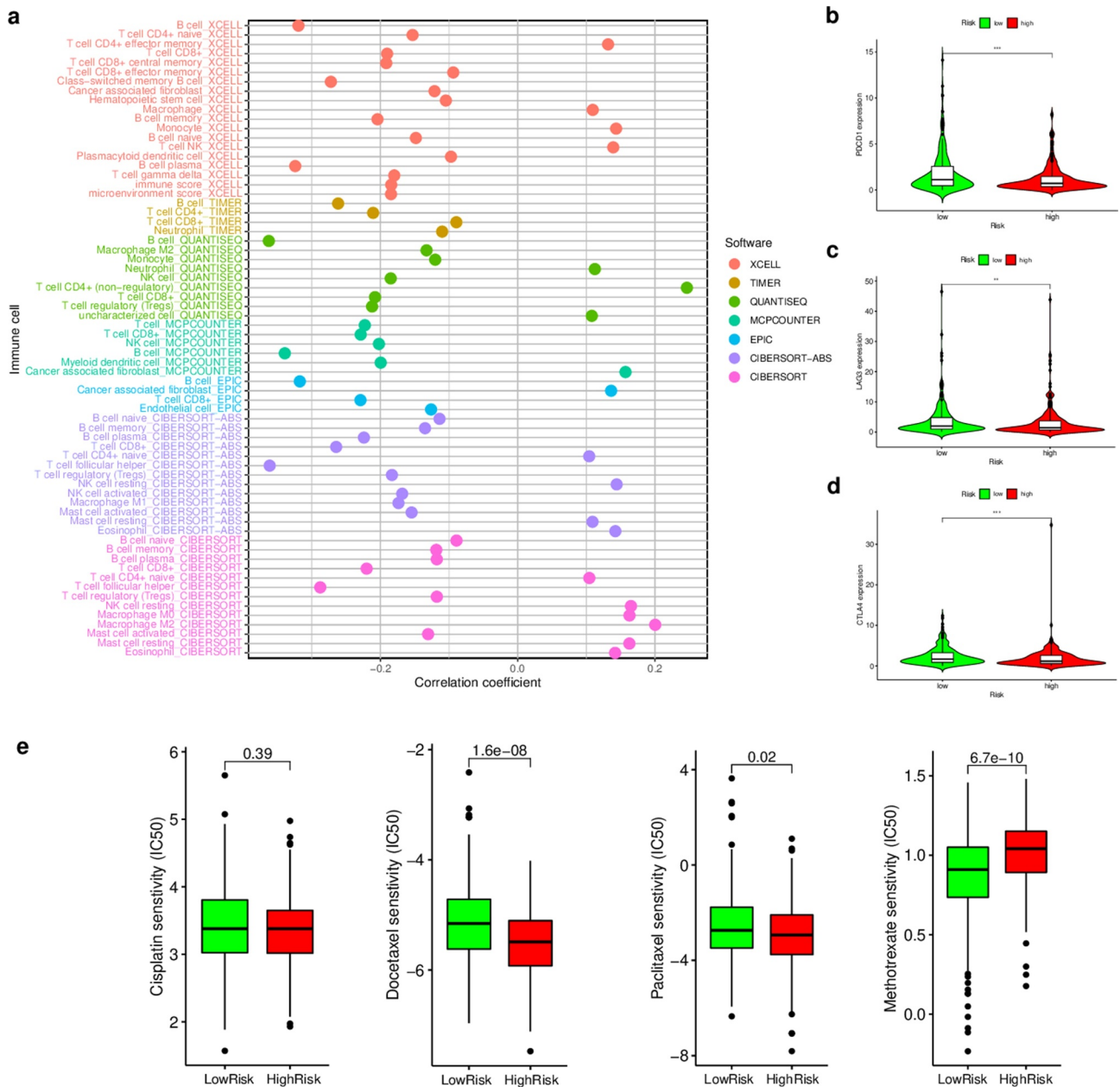
Since our model was based on lincRNAs, we further investigated the potential relationship with the tumor immune microenvironment. Correlation analysis in the context of immune cells infiltration revealed that the high-risk group was positively associated with CD4 + T cells, cancer-associated fibroblasts, monocytes, and macrophages, and negatively associated with CD8 + T cells and regulatory T cells, as revealed by the Wilcoxon signed-rank test (Table S2 and Figure S2). In addition, the analysis using XCELL revealed that the immune scores in the low-risk group were significantly higher than those in the high-risk group ( $P = 0.0038$ ); these results suggest that the tumors in the low-risk group had more immune cell infiltration than those in the high-risk group. A detailed Spearman's correlation analysis was further conducted, and the resulting diagram exhibited a lollipop shape, as shown in Figure 7(a). Since immunotherapy with ICIs has dramatically changed the HNSCC treatment panorama, improving the survival of HNSCC patients, next, we investigated whether the risk assessment model was related to the expression of ICI-related biomarkers. Curiously, we discovered that the high-risk scores were positively correlated with a low expression of *PDCD1* ( $P < 0.001$ , Figure 7(b)), *LAG3* ( $P < 0.01$ , Figure 7(c)), and *CTLA4* ( $P < 0.001$ , Figure 7(d)).

### Correlation between the risk assessment model and chemosensitivity

As chemotherapy is commonly used to treat HNSCC, we also attempted to identify associations between the risk scores and the efficacy of different chemotherapy agents in the treatment of HNSCC using the HNSCC dataset (TCGA project). Importantly, we found that a high-risk score was associated with a lower  $IC_{50}$  value for chemotherapeutic drugs, such as docetaxel ( $P < 0.001$ ) and paclitaxel ( $P = 0.02$ ), but with a higher  $IC_{50}$  value for methotrexate ( $P < 0.001$ ) (Figure 7(e)). Altogether, these results indicate that our model also acts as a potential predictor of chemosensitivity.

## Discussion

In recent years, many studies have highlighted lincRNAs as important factors for the prognosis of HNSCC patients [32]. In fact, many prognostic lincRNA expression signatures have been applied to both the diagnosis and prognosis prediction in the context of HNSCC. However, most signatures are based on the quantification of the expression levels. Here, we attempted to construct a reasonable model based on two-lincRNA combinations, instead of on their specific expression levels. Only pairs with high or low expression were detected (*versus* the examination of the specific expression levels of each lincRNA), aiming to increase the diagnosis/prognosis



**Figure 7.** Estimation of tumor-infiltrating cells, immunosuppression molecules, and the IC<sub>50</sub> values of chemotherapeutics using the risk assessment model. (a) Lollipop diagram of the correlation between the risk scores and immune cell infiltration. (b–d) Violin plots of the expression of genes related to Immune checkpoint inhibitors in the low- and high-risk groups. (e) Boxplots highlighting the difference in the IC<sub>50</sub> values of cisplatin, docetaxel, methotrexate, and paclitaxel between the high- and low-risk groups.

accuracy in the context of cancer, thinking on clinical practicability. Our findings showed that this novel model has a good predictive performance, better than those of two previously published expression-based lncRNA signatures. Of note, the risk score derived from the DEirLncPSig is reliable and independently predicts the prognosis in the context of HNSCC.

Some of the lincRNAs previously associated with the malignant phenotypes of various cancers, including MIAT [33], TYMSOS [34], LINC01063 [35], HOXC-AS1 [36], LINC02454 [37], SCAT1 [38], and LINC01322 [39] were identified in the process of modeling in this study. Furthermore, previous studies have also shown that LINC00460 [40], C5orf66-AS1 [41], and TMPO-AS1 [42], also

identified in our study, are associated with the prognosis and treatment outcomes of HNSCC. In fact, Jiang et al. [40] revealed that LINC00460 promotes epithelial-mesenchymal-transition in HNSCC via the facilitation of the transfer of peroxiredoxin-1 into the nucleus. Additionally, Lu et al. [41] first reported that C5orf66-AS1 is able to prevent the progression of oral squamous cell carcinoma, inhibiting tumor cell growth and metastasis via the regulation of CYC1 expression. Last but not least, Xing et al. [42] identified the novel TMPO-AS1/miR-320a/SOX4 pathway associated with nasopharyngeal carcinoma progression; interestingly, these authors suggested that TMPO-AS1 may be a potential therapeutic target for nasopharyngeal carcinoma. However, some lncRNAs were identified for the first time in the model established in the present study, suggesting that the proposed model could identify novel biomarkers for further research. Importantly, we validated these biomarkers using qRT-PCR and found that the gene expression rate of protective factors such as AC093159.1/ TYMSOS, SCAT1 / LINC00460 and Z82243.1/LINC02561 was significantly higher in early versus late tumor stages. Additionally, although no obvious differences in the expression of risk factors were found in the context of different tumor stages, survival analysis showed that such risk factors were closely related to survival. Therefore, overall, these results further validate the model in this study.

It should be noted that HNSCC is a type of immunosuppressive disease [43]. Despite the effectiveness of targeted therapies and immunotherapy agents in the context of HNSCC, not all patients respond to such treatments and the development of resistance is almost universal after a period of time. In fact, in clinical practice, ICI therapy does not appear to improve the survival of HNSCC patients. Instead, precision medicine approaches that seek to individualize therapy through the use of predictive biomarkers and stratification strategies have been the key to improve the therapeutic outcomes of HNSCC patients [32,43]. Hence, it is essential to develop sensitive predictive biomarkers, essential for the deeper understanding of the heterogeneity and complexity of the tumor immune microenvironment; only through this approach can we finally

improve the efficacy of tumor immunotherapy. For instance, an ideal signature would not only predict the prognosis of cancer patients, but also reflect the characteristics of the tumor and its immune microenvironment, both directly associated with disease progression and treatment response. Therefore, here we established a prediction model using lncRNAs and evaluated the relationships between the risk scores drawn from the DEirLnc-PSig and immune checkpoints, the abundance of tumor-infiltrating immune cells, and response to chemotherapy.

Tumor-infiltrating immune cells, part of the tumor immune microenvironment, are recognized as highly important for prognosis prediction; in fact, the responsiveness to immune checkpoint blockade may be related to the features of the tumor immune microenvironment. Therefore, here, to explore the relationship between risk scores and tumor-infiltrating immune cells, we used seven commonly acceptable methods to estimate the abundance of infiltrating immune cells, including XCELL, TIMER, QUANTISEQ, MCPcounter, EPIC, CIBERSORT-abs, and CIBERSORT. Multiple studies have indicated that the greater infiltration of CD8 + T cells and high immune scores are correlated with better response to standard chemotherapy and immune checkpoint blockade therapy in cancer patients [44,45]. Our analyses revealed that the high-risk group, as per the DEirLnc-PSig was negatively associated with CD8 + T cell infiltration and immune scores. Correspondingly, the high-risk group was also negatively correlated with immune checkpoint-related biomarkers, such as *PDCD1*, *LAG3*, and *CTLA4*. Therefore, our results suggest that patients in the high-risk group identified by our model might not benefit from immunotherapy. However, this conclusion needs to be verified through further studies, especially in the context of clinical trials. On the contrary, the estimated IC<sub>50</sub> of cisplatin, paclitaxel, doxorubicin, and methotrexate indicated that patients in the high-risk group could be more sensitive to chemotherapy than those in the low-risk group. Thus, altogether, our results suggest that this model may be particularly helpful to identify patients at a high clinical risk, the ones that will benefit from chemotherapy *in lieu* of immunotherapy. However, further benchwork and clinical studies are needed for further validation.

## Conclusion

In summary, our study reveals a comprehensive landscape of the tumor immune microenvironment in HNSCC and provides a promising prognostic tool in the context of HNSCC. For instance, patients in the high-risk group, as per the DEirLnc-PSig, negatively associated with CD8 + T cell infiltration and immune scores, have a worse prognosis. On the contrary, the low-risk group whose immune score was significantly higher than that of the high-risk group, was associated with a better prognosis. Therefore, overall, these details suggest that patients with low-risk scores may benefit from ICI therapy, while those with high-risk scores should receive chemotherapy rather than ICI therapy. However, prospective studies are needed to verify the prognosis accuracy of the established model and test its clinical utility.

## Research Highlights

- A DEirLncRNA-pair signature was identified to predict the OS of HNSCC patients.
- Different tumour-infiltrating immune cells are enriched in different risk groups.
- This tool may help distinguish HNSCC patients that could benefit from immunotherapy.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the National Key R&D Program of China [grant number 2020YFC1316903]; the 5010 Clinical Research Program of Sun Yat-sen University [grant number 2017004]; and the National Natural Science Foundation of China [grant number 81972528].

## Data availability statement

The data that support the findings of this study are openly available in TCGA repository at <https://portal.gdc.cancer.gov/>.

## Author contributions

(I) Conception and design: ZMC, LC, WBL; (II) Collection and assembly of data: LC, ZWC; (III) Data analysis and interpretation: LC, ZWC; (IV) Validation: LC, KXL, ZMC; (V) Manuscript writing: All authors; (VI) Final approval of the manuscript: All authors.

## References

- [1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018 Nov;68(6):394–424.
- [2] Katiyar SK. Emerging phytochemicals for the prevention and treatment of head and neck cancer. *Molecules.* 2016 Nov 24;21(12):1610.
- [3] Fleming JC, Woo J, Moutasim K, et al. HPV, tumour metabolism and novel target identification in head and neck squamous cell carcinoma. *Br J Cancer.* 2019 Feb;120(3):356–367.
- [4] Nelson HH, Pawlita M, Michaud DS, et al. Immune response to HPV16 E6 and E7 proteins and patient outcomes in head and neck cancer. *JAMA Oncol.* 2016;3(2):178–185.
- [5] Posner M, Vermorken JB. Induction therapy in the modern era of combined-modality therapy for locally advanced head and neck cancer. *Semin Oncol.* 2008 Jun;35(3):221–228.
- [6] Romano E, Romero P. The therapeutic promise of disrupting the PD-1/PD-L1 immune checkpoint in cancer: unleashing the CD8 T cell mediated anti-tumor activity results in significant, unprecedented clinical efficacy in various solid tumors. *J Immunother Cancer.* 2015;3(1):15.
- [7] Wang SM, Jiang B, Deng Y, et al. Clinical significance of MLH1/MSH2 for stage II/III sporadic colorectal cancer. *World J Gastrointest Oncol.* 2019 Nov 15;11(11):1065–1080.
- [8] Schepisi G, Brighi N, Cursano MC, et al. Inflammatory biomarkers as predictors of response to immunotherapy in urological tumors. *J Oncol.* 2019 Sep 19;2019:7317964.
- [9] Liu Y, Zhou S, Du Y, et al. Efficacy and safety of programmed death 1 inhibitors in patients with advanced non-small cell lung cancer: a meta-analysis. *Cancer Manag Res.* 2019 May 21;11:4619–4630.
- [10] Kopp F, Mendell JT. Functional classification and experimental dissection of long noncoding RNAs. *Cell.* 2018 Jan 25;172(3):393–407.
- [11] Schmitt AM, Chang HY. Long noncoding RNAs: at the intersection of cancer and chromatin biology. *Cold Spring Harb Perspect Med.* 2017 Jul 5;7(7):a026492.
- [12] Mao R, Chen Y, Xiong L, et al. Identification of a nomogram based on an 8-lncRNA signature as a novel diagnostic biomarker for head and neck

- squamous cell carcinoma. *Aging* (Albany NY). 2020 Oct 22;12(20):20778–20800.
- [13] Bao S, Zhao H, Yuan J, et al. Computational identification of mutator-derived lncRNA signatures of genome instability for improving the clinical outcome of cancers: a case study in breast cancer. *Brief Bioinform.* 2020 Sep 25;21(5):1742–1755.
- [14] Martinez LM, Robila V, Clark NM, et al. Regulatory T cells control the switch from in situ to invasive breast cancer. *Front Immunol.* 2019;10:1942.
- [15] Peng D, Wang L, Li H, et al. An immune infiltration signature to predict the overall survival of patients with colon cancer. *IUBMB Life.* 2019 Nov;71(11):1760–1770.
- [16] Li AL, Zhu YM, Gao LQ, et al. Exploration of the immune-related signatures and immune infiltration analysis in Melanoma. *Anal Cell Pathol (Amst).* 2021 Jan 16;2021:4743971.
- [17] Atianand MK, Caffrey DR, Fitzgerald KA. Immunobiology of long noncoding RNAs. *Annu Rev Immunol.* 2017 Apr 26;35(1):177–198.
- [18] Chen YG, Satpathy AT, Chang HY. Gene regulation in the immune system by long noncoding RNAs. *Nat Immunol.* 2017 Aug 22;18(9):962–972.
- [19] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33(1):1–22.
- [20] Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics.* 2000 Jun;56(2):337–344.
- [21] Aran D. Cell-type enrichment analysis of bulk transcriptomes using xCell. *Methods Mol Biol.* 2020;2120:263–276.
- [22] Li T, Fan J, Wang B, et al. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 2017 Nov 1;77(21):e108–e110.
- [23] Plattner C, Finotello F, Rieder D. Deconvoluting tumor-infiltrating immune cells from RNA-seq data using quanTIseq. *Methods Enzymol.* 2020;636:261–285.
- [24] Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016 Oct 20;17(1):218. Erratum in: *Genome Biol.* 2016 Dec 1;17(1):249.
- [25] Racle J, Gfeller D. EPIC: a tool to estimate the proportions of different cell types from bulk gene expression data. *Methods Mol Biol.* 2020;2120:233–248.
- [26] Sturm G, Finotello F, Petitprez F, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics.* 2019 Jul 15;35(14):i436–i445. PMID: 31510660; PMCID: PMC6612828.
- [27] Chen B, Khodadoust MS, Liu CL, et al. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol.* 2018;1711:243–259.
- [28] Adelstein D, Gillison ML, Pfister DG, et al. NCCN guidelines insights: head and neck cancers, version 2.2017. *J Natl Compr Canc Netw.* 2017 Jun;15(6):761–770.
- [29] Jiang H, Ma B, Xu W, et al. A Novel Three-lncRNA Signature Predicts the Overall Survival of HNSCC Patients. *Ann Surg Oncol.* 2021 Jun;28(6):3396–3406.
- [30] Liu G, Zheng J, Zhuang L, et al. A prognostic 5-lncRNA expression signature for head and neck squamous cell carcinoma. *Sci Rep.* 2018 Oct 15;8(1):15250.
- [31] Tang Z, Li C, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017 Jul 3;45(W1):W98–W102.
- [32] Wang Y, Wang S, Ren Y, et al. The role of lncRNA crosstalk in leading cancer metastasis of head and neck squamous cell carcinoma. *Front Oncol.* 2020 Oct 2;10:561833.
- [33] Peng L, Chen Y, Ou Q, et al. lncRNA MIAT correlates with immune infiltrates and drug reactions in hepatocellular carcinoma. *Int Immunopharmacol.* 2020 Dec;89(Pt A):107071.
- [34] Gu Y, Wan C, Zhou G, et al. TYMSOS drives the proliferation, migration and invasion of gastric cancer cells by regulating ZNF703 via sponging miR-4739. *Cell Biol Int.* 2021 Apr 13.
- [35] Zhou C, An N, Cao C, et al. lncRNA HOXC-AS1 promotes gastric cancer via binding eIF4AIII by activating Wnt/ $\beta$ -catenin signaling. *J Gene Med.* 2020 Sep;22(9):e3202.
- [36] Dong Y, Li X, Lin Z, et al. HOXC-AS1-MYC regulatory loop contributes to the growth and metastasis in gastric cancer. *J Exp Clin Cancer Res.* 2019 Dec 23;38(1):502.
- [37] Cai WY, Chen X, Chen LP, et al. Role of differentially expressed genes and long non-coding RNAs in papillary thyroid carcinoma diagnosis, progression, and prognosis. *J Cell Biochem.* 2018 Nov;119(10):8249–8259.
- [38] Zhang C, Zhang Z, Zhang G, et al. A three-lncRNA signature of pretreatment biopsies predicts pathological response and outcome in esophageal squamous cell carcinoma with neoadjuvant chemoradiotherapy. *Clin Transl Med.* 2020 Aug;10(4):e156.
- [39] Qing L, Gu P, Liu M, et al. Six-lncRNA signature as a novel prognostic biomarker for bladder cancer. *Onco Targets Ther.* 2020 Dec 7;13:12521–12538.
- [40] Jiang Y, Cao W, Wu K, et al. lncRNA LINC00460 promotes EMT in head and neck squamous cell carcinoma by facilitating peroxiredoxin-1 into the nucleus. *J Exp Clin Cancer Res.* 2019 Aug 20;38(1):365.

- [41] Lu T, Liu H, You G. Long non-coding RNA C5orf66-AS1 prevents oral squamous cell carcinoma through inhibiting cell growth and metastasis. *Int J Mol Med.* 2018 Dec;42(6):3291–3299.
- [42] Lee TW, Lai A, Harms JK, et al. Patient-derived xenograft and organoid models for precision medicine targeting of the tumour microenvironment in head and neck cancer. *Cancers (Basel).* 2020 Dec 12;12(12):3743.
- [43] Xing B, Qiao XF, Qiu YH, et al. TMPO-AS1 regulates the aggressiveness-associated traits of nasopharyngeal carcinoma cells through sponging miR-320a. *Cancer Manag Res.* 2021 Jan 15;13:415–425.
- [44] Chen Y, Chen H, Mao B, et al. Transcriptional Characterization of the tumor immune microenvironment and its prognostic value for locally advanced lung adenocarcinoma in a Chinese population. *Cancer Manag Res.* 2019 Oct 31;11:9165–9173.
- [45] Ge PL, Li SF, Wang WW, et al. Prognostic values of immune scores and immune microenvironment-related genes for hepatocellular carcinoma. *Aging (Albany NY).* 2020 Mar 25;12(6):5479–5499.