The Prognostic Value of the DNA Repair Gene Signature in Head and Neck Squamous Cell Carcinoma

Ruijie Ming[†], Enhao Wang[†], Jiahui Wei, Jinxiong Shen[*], Shimin Zong[*] and Hongjun Xiao[*]

Department of Otorhinolaryngology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Purpose: To construct a prognostic signature composed of DNA repair genes to effectively predict the prognosis of patients with head and neck squamous cell carcinoma (HNSCC).

Methods: After downloading the transcriptome and clinical data of HNSCC from the Cancer Genome Atlas (TCGA), 499 patients with HNSCC were equally divided into training and testing sets. In the training set, 13 DNA repair genes were screened using univariate proportional hazard (Cox) regression analysis and least absolute shrinkage and selection operator (LASSO) Cox regression analysis to construct a risk model, which was validated in the testing set.

Results: In the training and testing sets, there were significant differences in the clinical outcomes of patients in the high- and low-risk groups showed by Kaplan-Meier survival curves ($P < 0.001$). Univariate and multivariate Cox regression analyses showed that the risk score had independent prognostic predictive ability ($P < 0.001$). At the same time, the immune cell infiltration, immune score, immune-related gene expression, and tumor mutation burden (TMB) of patients with HNSCC were also different between the high- and low-risk groups ($P < 0.05$). Finally, we screened several chemotherapeutics for HNSCC, which showed significant differences in drug sensitivity between the high- and low-risk groups ($P < 0.05$).

Conclusion: This study constructed a 13-DNA-repair-gene signature for the prognosis of HNSCC, which could accurately and independently predict the clinical outcome of the patient. We then revealed the immune landscape, TMB, and sensitivity to chemotherapy drugs in different risk groups, which might be used to guide clinical treatment decisions.

Keywords: head and neck squamous cell carcinoma, DNA repair gene, prognostic signature, immune microenvironment, tumor mutation burden, drug sensitivity

INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) is a type of tumor that originates from the squamous epithelium of the head and neck areas, including the mucous membranes of the lips, tongue, pharynx, larynx, and others (1). HNSCC is currently one of the most common malignant tumors worldwide, with morbidity and mortality accounting for 3.6 and 3.4% of all malignant

tumors in 2020, respectively (2). HNSCC is highly malignant, and there are no specific prognostic-related biomarkers for clinical application. Therefore, prognostic-related biomarkers with clinical applicability are urgently required.

DNA damage and repair play important roles throughout the life of a cell (3). DNA damage affects the expression of a variety of genes, including proto-oncogenes and cancer suppressor genes. Changes in the activity of proto-oncogenes and cancer suppressor genes are crucial in tumorigenesis (4). Several DNA repair genes have been confirmed to play an important role in the development and prognosis of HNSCC (5–7). Hence, constructing a risk model composed of DNA repair genes may be useful for predicting the prognosis of patients with HNSCC.

In this study, we aimed to establish a prognostic prediction model for HNSCC based on DNA repair genes. We first equally divided all patients with HNSCC into training and testing sets. In the training set, we screened prognostic-related DNA repair genes using univariate proportional hazard (Cox) regression analysis and least absolute shrinkage and selection operator (LASSO) regression analysis to construct a risk model (8). All patients with HNSCC were classified into high- and low-risk groups according to the median value of the training set risk score. Subsequently, we verified the prognostic relevance and prognostic predictive ability of the risk model in the training and testing sets. We also analyzed the tumor-infiltrating immune cells, immune-related gene expression, tumor mutation burden, and drug sensitivity of patients with HNSCC in the high- and low-risk groups. The results showed that the risk model composed of DNA repair genes could effectively distinguish patients with different clinical outcomes and has independent predictive prognostic ability.

METHODS

Data Download

The transcriptome profiling (RNA-seq) data harmonized to fragments per kilobase million (FPKM), clinical information, and tumor mutations in patients with HNSCC were downloaded from the Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) in March 2021 (9). The pathologic stages were reconfirmed according to the seventh edition of the American Joint Committee on Cancer staging system (10). The gene transfer format (GTF) files were downloaded from Ensembl (http://asia.ensembl.org) for annotation (11). Immune-related genes were downloaded from the Tracking Tumor Immunophenotype (http://biocc.hrbmu.edu.cn/TIP/index.jsp) (12). The gene list, containing 569 DNA repair genes, was downloaded from Gene Set Enrichment Analysis (GSEA), "GO_DNA REPAIR" gene set (http://www.gsea-msigdb.org/gsea/msigdb/cards/GOBP_DNA_REPAIR.html) (13, 14). After annotation by the GTF files, 545 DNA repair genes were eventually used for subsequent analyses. GSE41613 (15), GSE27020 (16), GSE117973 (17), and GSE65858 (18) datasets with transcriptome and clinical data of patients with HNSCC

were downloaded from Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) for external validation.

Construction of Risk Model

To construct the risk model, we first combined the transcriptome data and clinical information of patients with HNSCC to obtain 499 samples with complete clinical information and transcriptome information, and then randomly divided them into a training set and a testing set on average. Subsequently, LASSO regression analysis was performed to further screen out 13 more representative DNA repair genes for use in constructing the risk model, and the correlation coefficients (*Coef*) and expression (*EXP*) of these 13 genes were obtained using the "glmnet" package in R (19). Finally, the risk score of each patient was calculated by the following formula: Risk Score = $\sum_{i=1}^{n} Exp_i \times Coef_i$, where *n* refers to the number of selected DNA repair genes, *Exp_i* indicates the expression levels of gene *i* in each HNSCC sample, and *Coef_i* is the correlation coefficient of gene *i*. Finally, we classified all HNSCC samples into high- and low-risk groups based on the median value of the risk score of the training set.

Validation of the Risk Model

We verified the risk model separately in the training and testing sets. To this end, we first performed principal component analysis (PCA) in the training and testing sets to evaluate the discrimination of the risk model for patients in the high- and low-risk groups. We then utilized heat maps to show the expression patterns of the DNA repair genes in the risk model in the training and testing sets. The Kaplan-Meier survival curve was used to distinguish the difference in the clinical outcome of patients in the high- and low-risk groups, and the significant difference *P*-value was calculated by the log-rank test. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve was used to evaluate the prognostic diagnostic accuracy of the risk score and clinical characteristics. Univariate and multivariate Cox regression analyses of risk score and clinical characteristics were used to evaluate the independent correlation between the risk score and prognosis of patients with HNSCC. We also performed the above verification in all patients with HNSCC. Then we divided all samples into multiple clinical subgroups based on clinical characteristics, and the Kaplan-Meier survival curve was performed in each subgroup to demonstrate the good prognostic ability of the risk score.

Evaluation of the Tumor Immune Microenvironment and Immune-Related Gene Expression

Before analyzing the immune-infiltration situation using the CIBERSORT algorithm, which contains 22 types of immune cells, we first standardized the gene expression data through the "CIBERSORT" package in R (20). The Wilcoxon test was used to compare the different infiltrations of the 22 immune cells in the high- and low-risk groups. The Pearson test was used to analyze the correlation between risk genes and tumor-infiltrating immune cells through Statistical Product and Service Solutions 25.0 (SPSS 25.0) (21). The ESTIMATE (Estimation of STromal

and Immune cells in MAlignant Tumors using Expression data) algorithm was used to evaluate the immune score, stromal cell content, and ESTIMATE score of each sample (22). We analyzed the expression of negative regulatory immune genes in the high- and low-risk groups using the Wilcoxon test. Finally, as the research on the role of immune checkpoint genes in various tumors is increasing, we analyzed the correlation between these genes and risk scores using the Spearman test and analyzed their differences in expression in the high- and low-risk patients using the Wilcoxon test.

Assessment of Tumor Mutation Burden

We displayed the 30 genes with the highest mutation rate in all HNSCC samples and calculated the tumor mutation burden (TMB) of all samples through the "maftools" package in R (23). We then divided the HNSCC samples into high- and low-TMB groups according to the best cut-off value of the TMB of each sample. The Kaplan-Meier survival curve showed the clinical outcome of the two groups of patients with HNSCC. By combining the TMB groups and the risk groups, we further evaluated the impact of the risk score and tumor mutation burden on the clinical outcome of patients with HNSCC and displayed them with survival curves.

Online Website Verification

We verified the influence of the expression of the 13 DNA repair genes on the Oncolnc website (http://www.oncolnc.org).

Analysis of Drug Sensitivity

To evaluate the model in the clinical treatment of HNSCC, we calculated the half-inhibitory concentration (IC50) of chemotherapeutic drugs for HNSCC. The difference in the IC50 between the high- and low-risk groups was compared by Wilcoxon signed-rank test using the "pRRophetic" package in R (24).

Statistical Analysis

The significance level of the P-value was set to <0.05. All statistical analyses were performed using R 4.0.4 (https://www.r-project.org/).

RESULTS

Development and Validation of the Prognostic Model

The flowchart of this research is shown in **Figure 1**. After merging the transcriptome data and the clinical data of patients with HNSCC downloaded from TCGA, we obtained 499 samples with complete information. We then divided all HNSCC samples into a training set (n = 251) and a testing set (n = 248). The basic clinical information of the two groups of patients is shown in **Table 1**. Subsequently, we screened out 82 prognostic-related genes among 545 DNA repair genes through univariate Cox regression analysis (**Table S1**, $P < 0.05$), of which 21 were risk genes (hazard ratio > 1). Subsequently, we screened out a further

13 representative DNA repair genes through LASSO Cox regression analysis, which were used to construct the risk model. The risk score was calculated based on the sum of the product of the expression (*Exp*) of all genes in the model and its correlation coefficient (*Coef*). The formula of the risk score was as follow: Risk Score = MORF4L2 * (0.0037) + COPS2 * (0.0063) + USP10 * (0.0255) + WAS * (–0.0123) + UVSSA * (–0.1324) + PRRX1 * (–0.0148) + ZBTB1 * (–0.0632) + DCLRE1C * (–0.0502) + MSH5 * (–0.3824) + DOT1L * (–0.1573) + ZBTB7A * (–0.00610) + POLR2C * (0.0085) + MORF4L1 * (0.0047). A negative correlation coefficient indicated that the gene was a protective factor in patients with HNSCC. In contrary, the gene with a positive correlation coefficient was a risk factor.

After calculating the risk scores of all patients with HNSCC, we divided the training set and testing set samples into high- and low-risk groups according to the median value of the training set risk score, as shown in **Figures 2A, G**. We found that in both the training and testing sets, the proportion of patients with HNSCC who died in the high-risk group was higher than that in the low-risk group (**Figures 2B, C, H, I**). The high- and low-risk groups were well distinguished (**Figures 2D, J**). Moreover, the DNA repair genes in the risk model showed the same expression pattern in the training and testing sets (**Figures 2E, K**). The Kaplan-Meier survival curve showed that the clinical outcomes of patients in the low-risk group were better than those in the high-risk group (**Figure 2L**), both in the training set ($P = 8.439e–09$; **Figure 2F**) and the testing set ($P = 1.161e-04$; **Figure 2L**).

To verify the ability and independence of our model to predict the prognosis of patients with HNSCC, we conducted ROC curves and univariate and multivariate Cox regression analyses in the training and testing sets, respectively. The sensitivity and specificity of the risk score were assessed using the ROC curve. In the training set, the area under the curve (AUC) of the 1-, 3-, and 5-year ROC curves of the risk score were all >0.7 (**Figure 3A**). The risk score had the largest AUCs of the 3-year ROC curve, compared to the clinical traits of patients with HNSCC (**Figure 3B**). In the testing set, the AUCs of the 1-, 3-, and 5-year ROC curves were all >0.65 (**Figure 3E**), and the risk score also had the largest AUCs of the 3-year ROC curve (**Figure 3F**). The hazard ratio (HR) value of the risk score was the largest in the univariate and multivariate Cox regression analyses of the risk score and multiple clinical features, which showed that the risk score was an independent prognostic factor (**Figures 3C, D**). The independence of the risk score for predicting the prognosis of HNSCC was confirmed in the test set (**Figures 3G, H**). **Table S2** shows the univariate and multivariate Cox regression analyses of the training and testing sets. Overall, the risk score was an independent prognostic factor for HNSCC. We were unable to find an external validation dataset with transcripts of all risk genes. However, we still verified the predictive ability of other genes except MSH5 in patients with HNSCC in GSE41613. We found that despite the lack of MSH, patients with HNSCC in the low-risk group showed better clinical outcomes than those in the high-risk group in our model ($P < 0.05$), and the expression pattern of the remaining genes was consistent with the training

FIGURE 1 | Flowchart of this study.
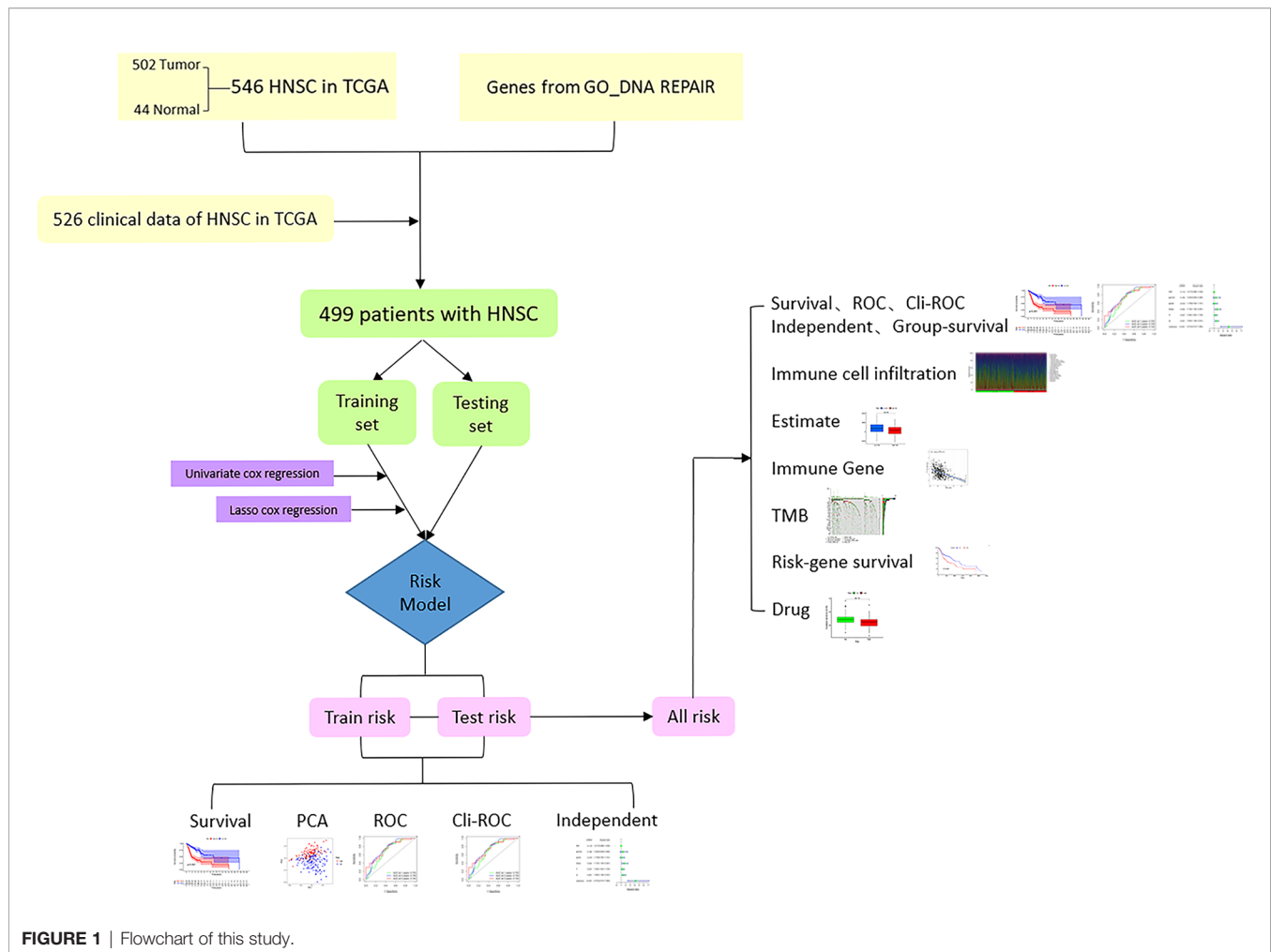
TABLE 1 | Basic clinical information of training set and testing set.

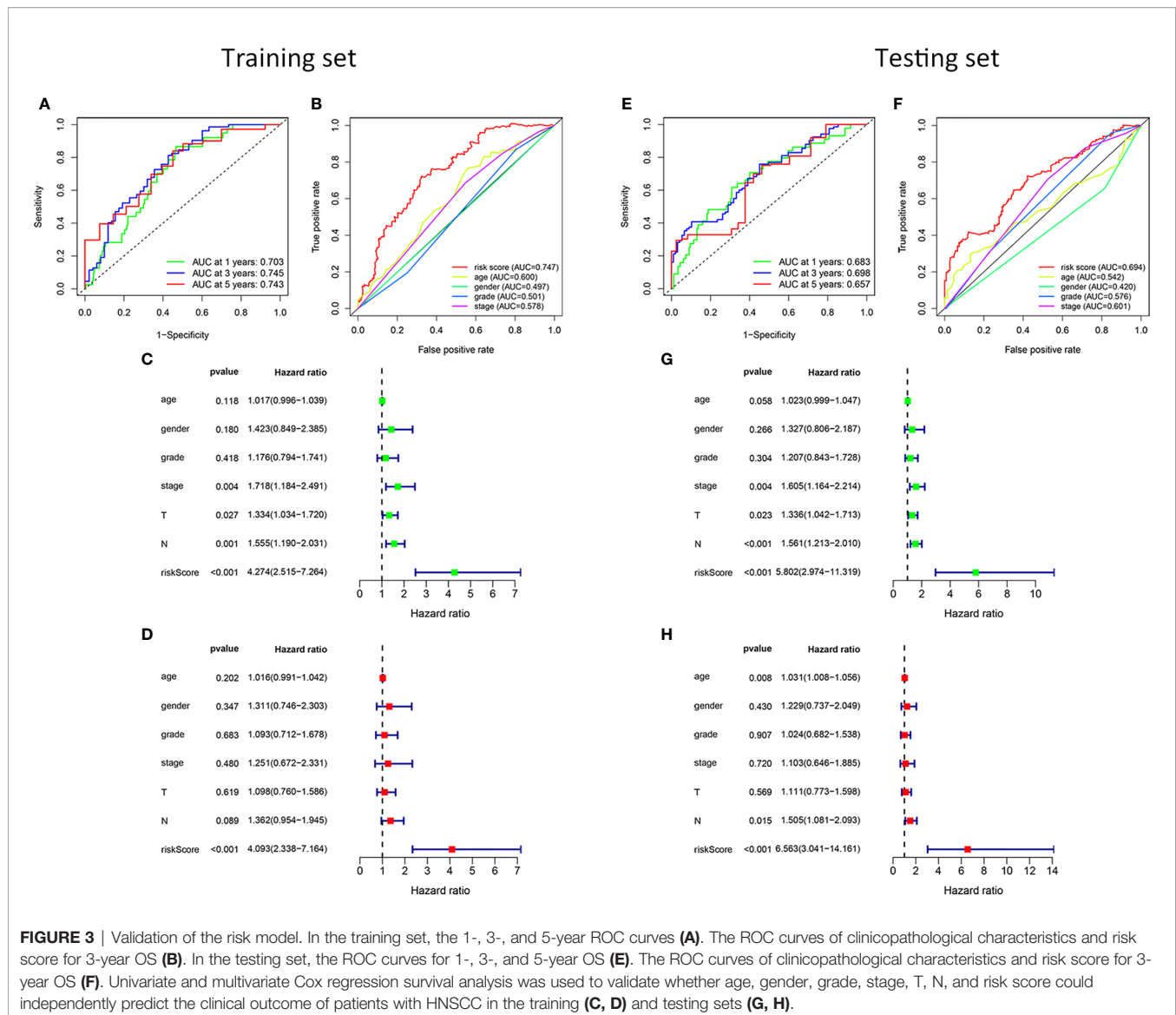| Characteristics	Training set (n = 251)	Testing set (n = 248)
Age		
<=65	162	155
>65	89	93
Gender		
Female	70	63
Male	181	185
Stage		
Stage I-III	83	86
Stage IV	130	127
Unknown	36	35
Tumor		
T 1-2	85	87
T 3-4	136	134
Unknown	30	27
Lymph node		
N 0	84	83
N 1-3	122	118
Unknown	45	47

and testing sets (**Figure S1**). And we still verified in GSE117973 without transcript of UVSSA, GSE27020 without transcripts of MSH5 and UVSSA, and GSE65858 without transcripts of

UVSSA and ZBTB1. The differences of prognosis of patients with HNSCC in the high- and low-risk groups were not significant ($P > 0.05$, **Figure S2**). For these external validation, we did a sensitivity analysis by using only 12 risk genes to recalculate the risk score. And we found that deleting every risk gene had little effect on the Kaplan-Meier survival curves (**Figures S3, S4**).

To further verify the accuracy of the model, we divided all samples into clinical subgroups based on different clinical traits, and we analyzed differences in the clinical outcomes of high- and low-risk samples in each clinical subgroup. Before clinical subgroup validation, we conducted a risk model validation for all samples. The Kaplan-Meier survival curve of all patients showed that the clinical outcomes of patients in the low-risk group were significantly better than those in the high-risk group ($P = 1.884e-11$; **Figure 4A**). The sensitivity and specificity of the risk scores of all HNSCC samples were assessed using the ROC curve. The AUCs of the ROC curves of risk score for 1-, 3-, and 5-year were all >0.65 (**Figure 4B**). The risk score had the largest AUCs of the ROC curves for 3-year compared to the clinical traits of patients with HNSCC (**Figure 4C**). PCA showed that patients in the high- and low-risk groups showed
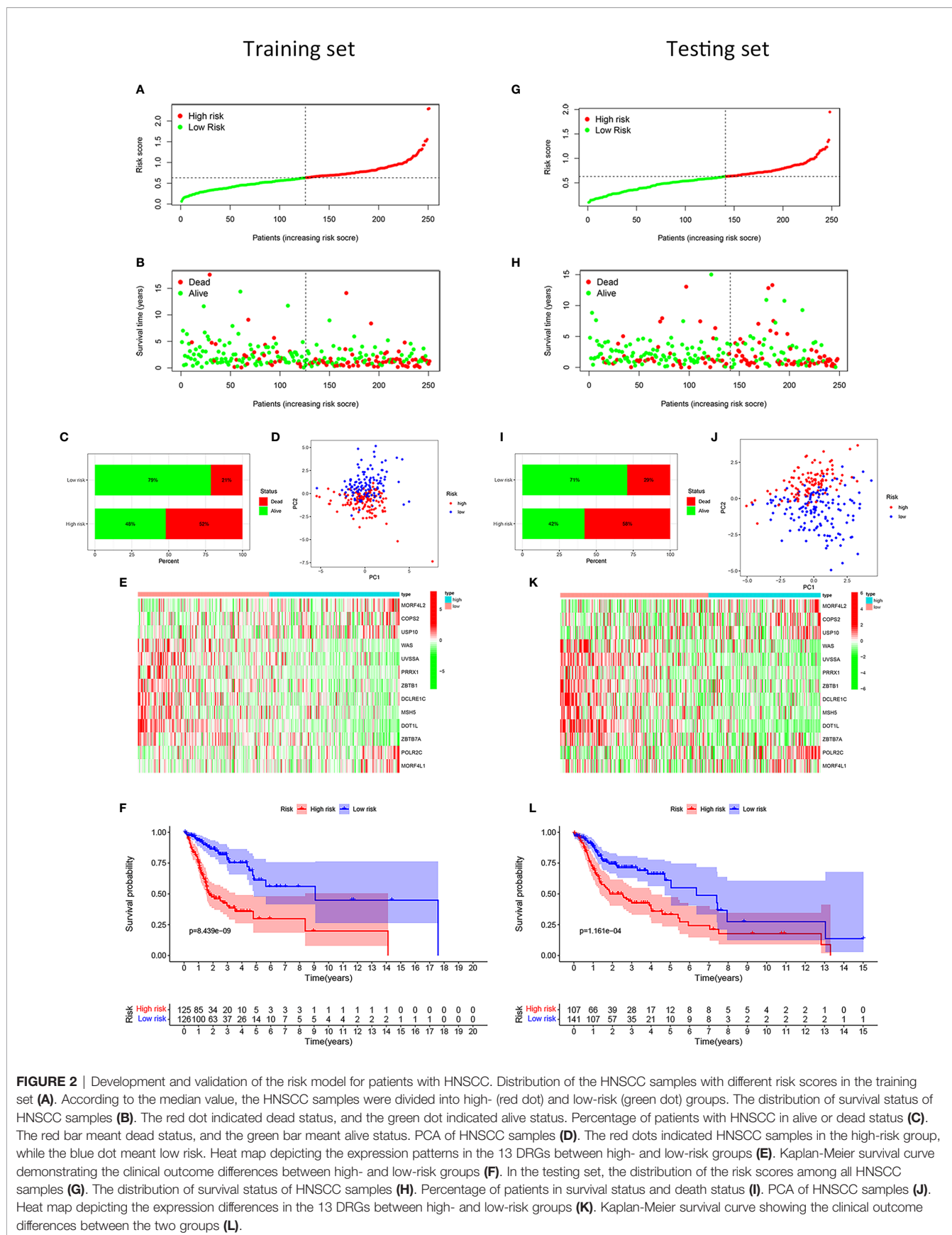
FIGURE 2 | Development and validation of the risk model for patients with HNSCC. Distribution of the HNSCC samples with different risk scores in the training set **(A)**. According to the median value, the HNSCC samples were divided into high- (red dot) and low-risk (green dot) groups. The distribution of survival status of HNSCC samples **(B)**. The red dot indicated dead status, and the green dot indicated alive status. Percentage of patients with HNSCC in alive or dead status **(C)**. The red bar meant dead status, and the green bar meant alive status. PCA of HNSCC samples **(D)**. The red dots indicated HNSCC samples in the high-risk group, while the blue dot meant low risk. Heat map depicting the expression patterns in the 13 DRGs between high- and low-risk groups **(E)**. Kaplan-Meier survival curve demonstrating the clinical outcome differences between high- and low-risk groups **(F)**. In the testing set, the distribution of the risk scores among all HNSCC samples **(G)**. The distribution of survival status of HNSCC samples **(H)**. Percentage of patients in survival status and death status **(I)**. PCA of HNSCC samples **(J)**. Heat map depicting the expression differences in the 13 DRGs between high- and low-risk groups **(K)**. Kaplan-Meier survival curve showing the clinical outcome differences between the two groups **(L)**.

FIGURE 3 | Validation of the risk model. In the training set, the 1-, 3-, and 5-year ROC curves **(A)**. The ROC curves of clinicopathological characteristics and risk score for 3-year OS **(B)**. In the testing set, the ROC curves for 1-, 3-, and 5-year OS **(E)**. The ROC curves of clinicopathological characteristics and risk score for 3-year OS **(F)**. Univariate and multivariate Cox regression survival analysis was used to validate whether age, gender, grade, stage, T, N, and risk score could independently predict the clinical outcome of patients with HNSCC in the training **(C, D)** and testing sets **(G, H)**.
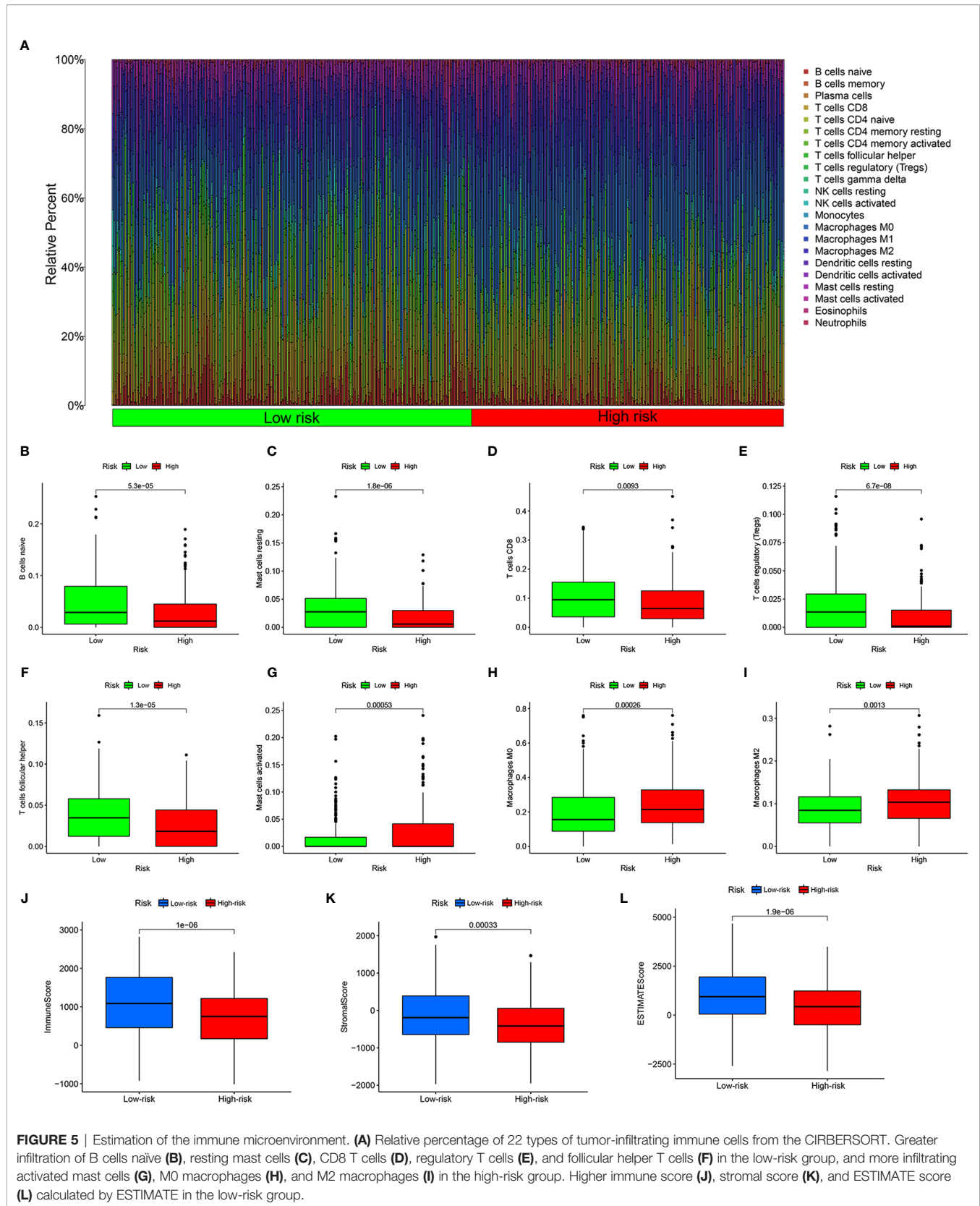
good discrimination (**Figure 4D**). The HR value of the risk score was the highest in the univariate and multivariate regression analyses of risk score and clinical characteristics (**Figures 4E, F**). Details of the univariate and multivariate Cox regression analyses of the training and testing sets are shown in **Table S3**. We divided all HNSCC samples into different clinical subgroups according to age, gender, stage, tumor (T), and lymph node (N) of patients with HNSCC. The clinical outcomes of patients in the low-risk group were significantly better than those in the high-risk group in all clinical subgroups, including those aged ≤65 years ($P < 0.001$, **Figure 4G**) and >65 years ($P < 0.001$, **Figure 4H**), male ($P < 0.001$, **Figure 4I**) and female ($P = 0.003$, **Figure 4J**), stage I-III ($P = 0.007$, **Figure 4K**), stage IV ($P < 0.001$, **Figure 4L**), T1-2 ($P < 0.001$, **Figure 4M**) and T3-4 ($P < 0.001$, **Figure 4N**), and N0 ($P = 0.011$, **Figure 4O**), and N1-3 ($P < 0.001$, **Figure 4P**).

Evaluation of the Immune Microenvironment and Expression of Immunoregulatory Genes

To reveal the differences in the immune microenvironment of high- and low-risk groups, including immune cell infiltration and expression of immunoregulatory and immune checkpoint genes, we first used the bioinformatics algorithm CIBERSORT to estimate 22 types of tumor-infiltrating immune cells in HNSCC. First, we found that among these 22 cell types, acquired immune-related immune cells infiltrated to a greater extent in HNSCC samples (**Figure 5A**). There were more naïve B cells ($P = 5.3e-05$, **Figure 5B**), resting mast cells ($P = 1.8e-06$, **Figure 5C**), T cells CD8 ($P = 0.0093$, **Figure 5D**), regulatory T cells (Tregs, $P = 6.7e-08$, **Figure 5E**), and follicular helper T cells ($P = 1.3e-05$, **Figure 5F**) in the low-risk group. In contrary, activated mast cells ($P = 0.00053$, **Figure 5G**), M0 macrophages ($P = 0.00026$,

Ming et al.

A Prognostic Signature for HNSCC

7
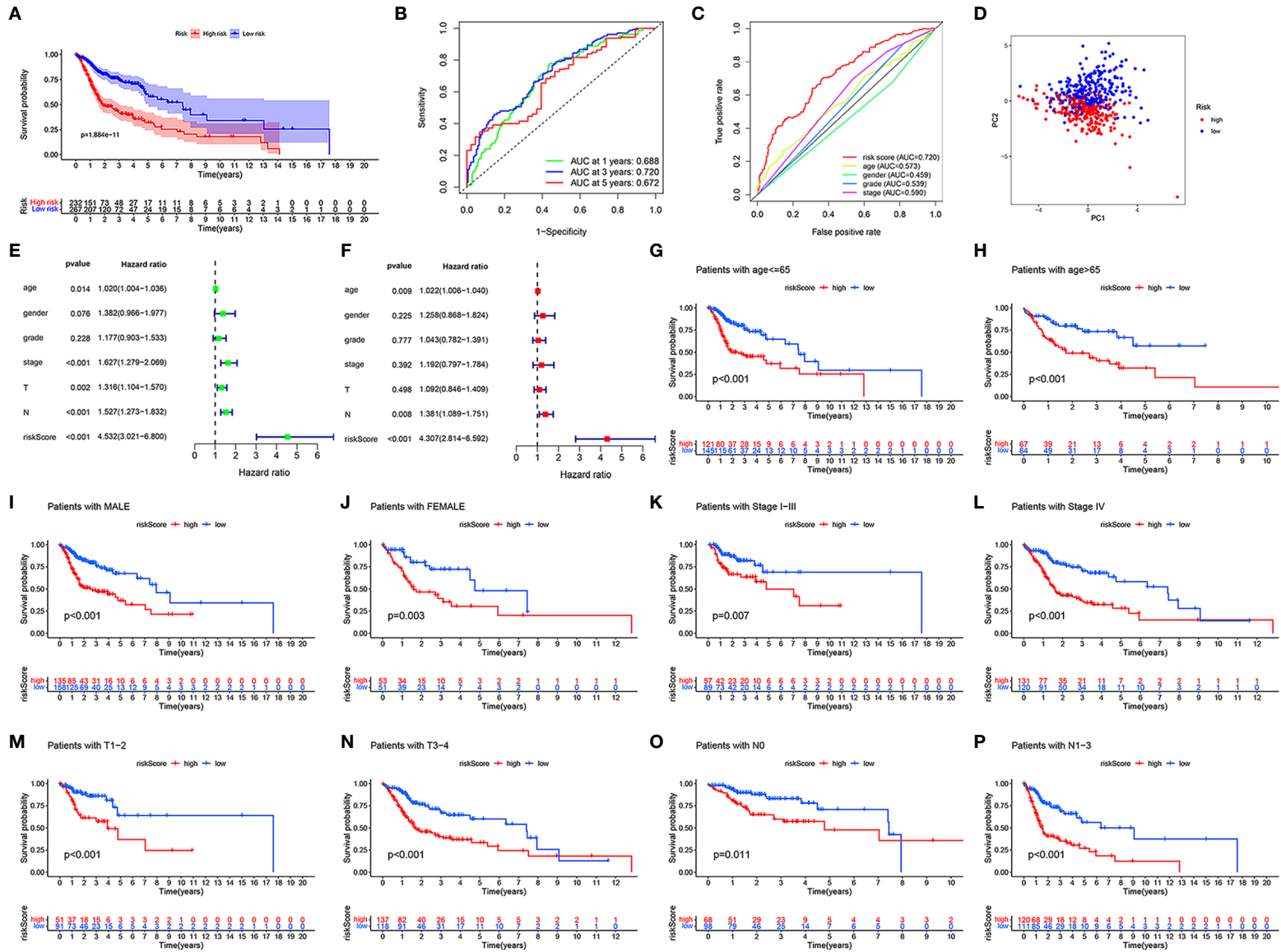
FIGURE 4 | Validation in different clinical traits subgroups. In all HNSCC samples, the Kaplan-Meier survival curve demonstrating the clinical outcome differences between the high- and low-risk groups **(A)**. The ROC curves for 1-, 3-, and 5-year OS **(B)**, ROC curves of clinicopathological characteristics and risk score **(C)** for 3-year OS. PCA of all HNSCC samples **(D)**. Univariate and multivariate Cox regression survival analysis validated whether age, gender, grade, stage, T, N, and risk score could independently predict the clinical outcomes of patients with HNSCC **(E, F)**. Kaplan-Meier curves showing the differences in prognosis between the high- and low- risk groups in different clinical subgroups, including ≤65 **(G)**, >65 **(H)**, male **(I)**, female **(J)**, stage I-III **(K)**, stage IV **(L)**, T1-2 **(M)**, T3-4 **(N)**, N0 **(O)**, and N1-3 **(P)**.

FIGURE 5 | Estimation of the immune microenvironment. **(A)** Relative percentage of 22 types of tumor-infiltrating immune cells from the CIRBERSORT. Greater infiltration of B cells naïve **(B)**, resting mast cells **(C)**, CD8 T cells **(D)**, regulatory T cells **(E)**, and follicular helper T cells **(F)** in the low-risk group, and more infiltrating activated mast cells **(G)**, M0 macrophages **(H)**, and M2 macrophages **(I)** in the high-risk group. Higher immune score **(J)**, stromal score **(K)**, and ESTIMATE score **(L)** calculated by ESTIMATE in the low-risk group.

Figure 5H), and M2 macrophages (*P* = 0.0013, **Figure 5I**) showed greater infiltration in the high-risk group. The HNSCC samples in the low-risk group had higher immune scores (*P* = 1e-06, **Figure 5J**), stromal scores (*P* = 0.00033, **Figure 5K**), and ESTIMATE scores (*P* = 1.9e-06, **Figure 5L**) evaluated by ESTIMATE than the high-risk group. In other words, the tumor purity of HNSCC was lower in the low-risk group. Naïve B cells were positively correlated with eight risk genes that had negative correlation coefficients and negatively correlated with MORF4L2, which had a positive correlation coefficient (*P* < 0.05). CD8+ T cells were positively correlated with five risk genes that had negative correlation coefficients and negatively correlated with eight risk genes that had negative correlation coefficients (*P* < 0.05). Tregs and follicular helper T cells were positively correlated with all risk genes that had negative correlation coefficients and negatively correlated with all risk genes that had positive correlation coefficients (*P* < 0.05). Monocytes and Macrophages M2 were negatively related to most risk genes (*P* < 0.05). Macrophages M0 were negatively correlated with some risk genes that had negative correlation coefficients and positively correlated with some risk genes that had positive correlation coefficients (*P* < 0.05). Details are shown in **Table S4**.

Next, we analyzed the relevant immune regulatory genes to further reveal the differences in the immune microenvironment of HNSCC in the high- and low-risk groups. Almost all negative immune regulatory genes in **Figure 6A** were highly expressed in the low-risk group, similar to CD4+ T cell and CD8+ T cell regulatory genes (**Figure 6B**). In addition, in recent years, immune checkpoint inhibitors have become increasingly common in the treatment of various tumors, including HNSCC. Therefore we investigated whether the risk model was related to immune checkpoint inhibitor-related biomarkers by Spearman correlation analysis, and we discovered that high risk scores were negatively correlated with the expression of CTLA4 (*R* = −0.34, *P* = 4.7e-15, **Figure 6C**), LAG3 (*R* = −0.28, *P* = 3e-10, **Figure 6D**), PD1 (*R* = −0.37, *P* < 2.2e-16, **Figure 6E**), PD-L1 (*R* = −0.16, *P* = 0.00051, **Figure 6F**), and TIM3 (*R* = −0.26, *P* = 7.4e-09, **Figure 6G**). A further Wilcoxon rank test also confirmed the expression pattern of CTLA4 (*P* = 4.8e-09, **Figure 6H**), LAG3 (*P* = 1.6e-06, **Figure 6I**), PD1 (*P* = 1.5e-11, **Figure 6J**), PD-L1 (*P* = 0.025, **Figure 6K**), and TIM3 (*P* = 6.2e-06, **Figure 6L**).

Assessment of Tumor Mutation Burden

To determine the tumor mutation burden (TMB), we first downloaded all the mutation data of HNSCC from TCGA and showed the top 30 mutation rate genes (**Figure 7A**). Subsequently, we identified the genes with the top 20 mutation rates in the high- and low-risk groups (**Figures 7B, C**). The tumor mutation rate of high-risk group samples was slightly higher than that of patients in the low-risk group, and the gene with the highest mutation rate in the high- and low-risk groups samples was TP53. According to the best cut-off point of TMB, all patients with HNSCC were divided into high- and low-TMB groups. The Kaplan-Meier survival curve showed that the clinical outcomes of patients with low TMB were significantly better

than those of patients with high TMB (*P* = 0.003, **Figure 7D**). To further evaluate the influence of TMB and risk score on the prognosis of patients with HNSCC, we combined the TMB group with the risk group and analyzed the clinical outcomes of different groups using the Kaplan-Meier survival curve. The results showed that patients with low risk and low TMB had the best clinical outcome, followed by patients with low risk and high tumor mutation load, and that patients with high risk and high tumor mutation load had the worst clinical outcome (*P* < 0.001, **Figure 7E**). Considering the high mutation rate of TP53, we analyzed the correlation between TP53 and the risk score and its expression in the high- and low-risk groups. As a result, we found that TP53 was negatively correlated with the risk score (*R* = −0.31, *P* = 3.9e-12, **Figure 7F**) and was highly expressed in the low-risk group (*P* = 2.5e-05, **Figure 7G**).
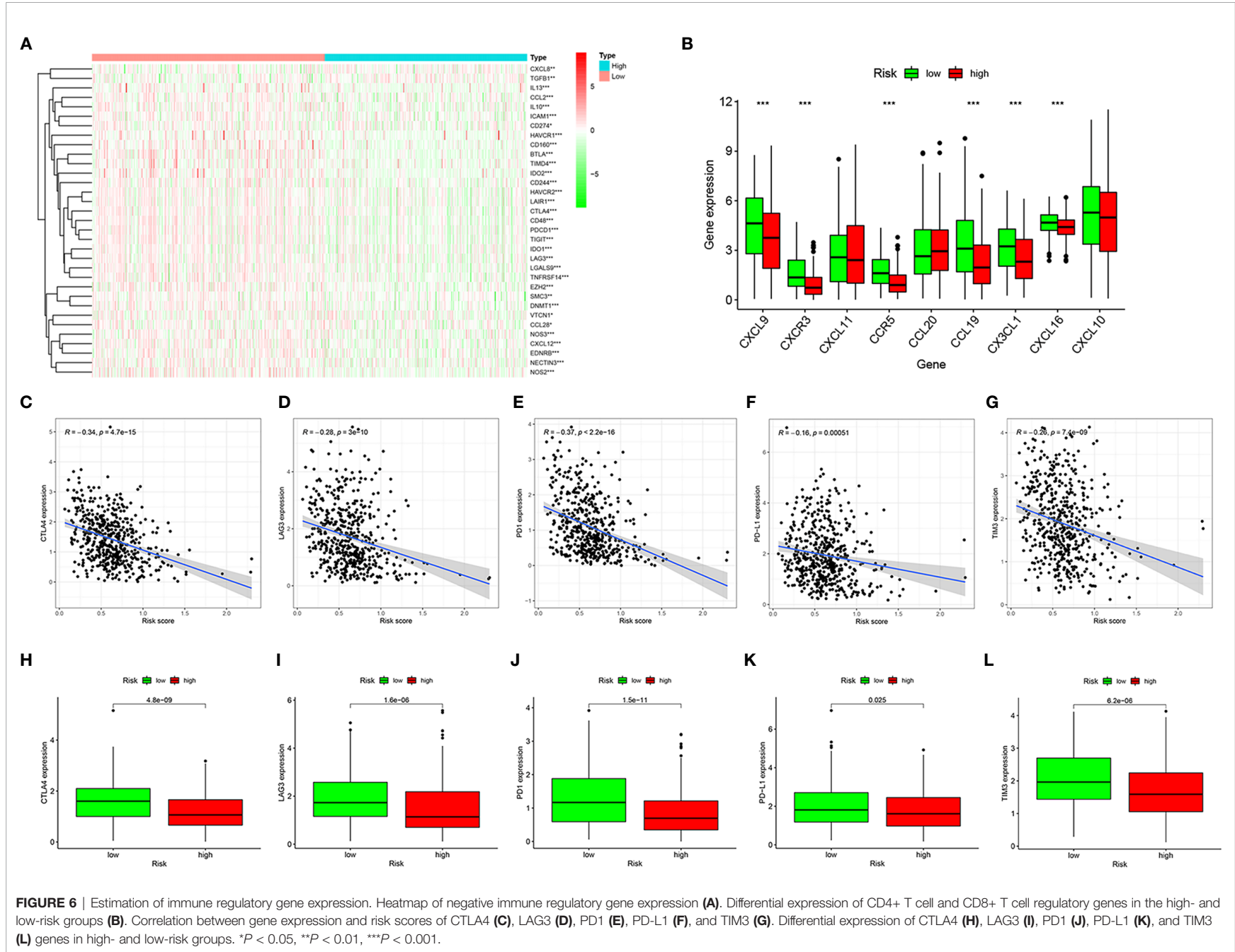
Validation of the Website Oncolnc

We searched on the Oncolnc (http://www.oncolnc.org/) to verify the impact of high- and low-risk DRGs in the model on the prognosis of HNSCC and found that high-risk DRGs were correlated with poor prognosis and low-risk DRGs were associated with favorable patient prognosis. There were significant p-values for COPS2 (*P* = 0.000031, **Figure 8A**), DCLRE1C (*P* = 0.0051, **Figure 8B**), DOT1L (*P* = 0.0261, **Figure 8C**), UVSSA (*P* = 0.00589, **Figure 8D**), MORF4L2 (*P* = 0.00254, **Figure 8E**), POLR2C (*P* = 0.000262, **Figure 8F**), WAS (*P* = 0.0146, **Figure 8G**), ZBTB1 (*P* = 0.0153, **Figure 8H**), and USP10 (*P* = 0.0376, **Figure 8I**), whereas MORF4L1 (*P* = 0.088, **Figure 8J**), PRRX1 (*P* = 0.144, **Figure 8K**), ZBTB7A (*P* = 0.205, **Figure 8L**), and MSH5 (*P* = 0.391, **Figure 8M**) were not significant. The risk genes with negative correlation coefficients were also protective factors in the Oncolnc database.
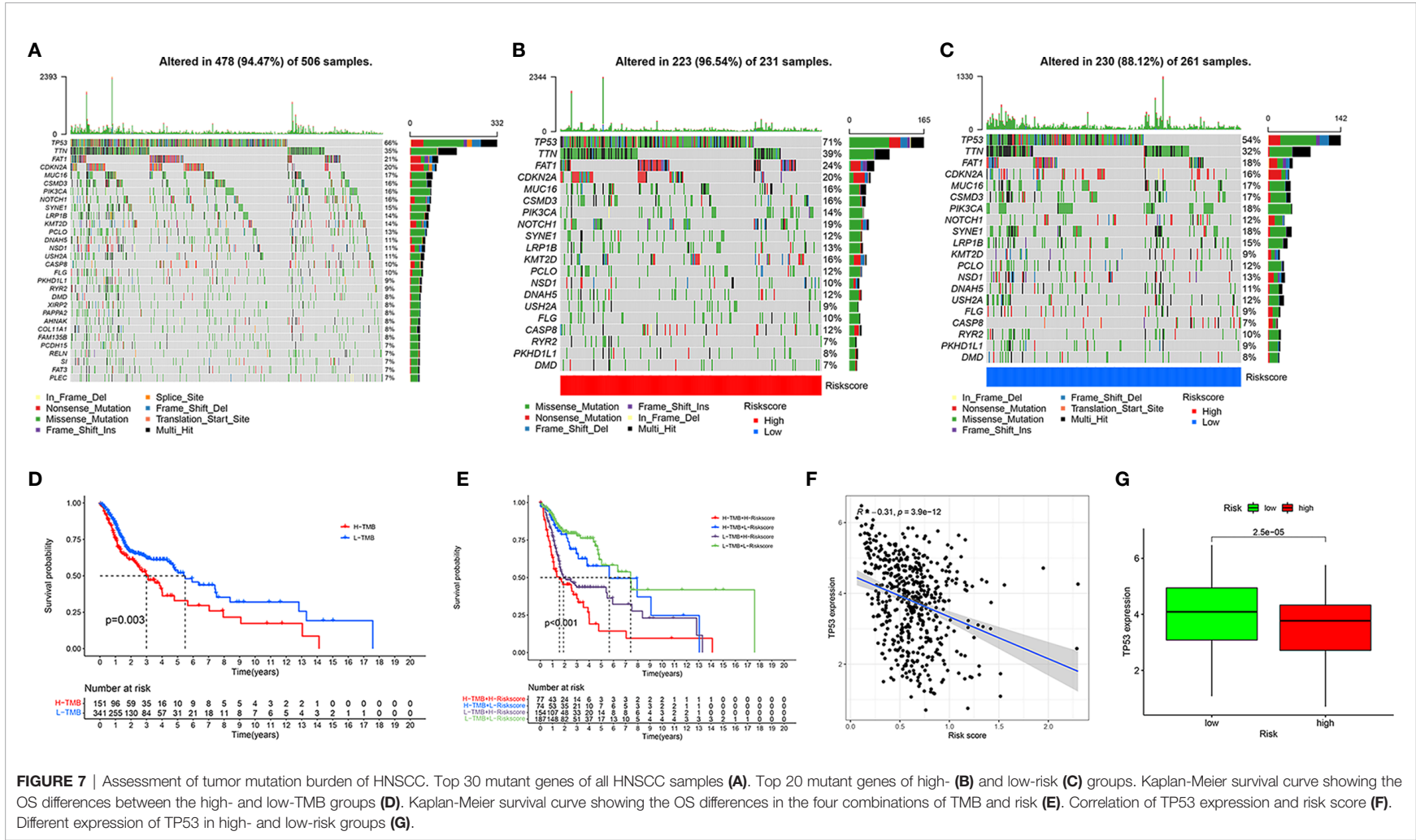
Analysis of Drug Sensitivity

To evaluate the possible clinical application of the risk model, we analyzed the sensitivity difference of chemotherapy drugs for HNSCC in the current stage of clinical trials between the high- and low-risk groups, with the drug sensitivity expressed by IC50. We showed that patients in the high-risk group were more sensitive to erlotinib (*P* = 8.3e-16, **Figure 9A**), gefitinib (*P* = 0.00056, **Figure 9B**), paclitaxel (*P* = 2.9e-05, **Figure 9C**), docetaxel (*P* = 2e-10, **Figure 9D**), and sorafenib (*P* = 2.7e-05, **Figure 9E**), whereas patients in low-risk group were more sensitive to methotrexate (*P* = 6e-07, **Figure 9F**), vinorelbine (*P* = 8.3e-05, **Figure 9G**), and rapamycin (*P* = 0.00015, **Figure 9H**), which indicated that the model could be used as a potential predictor of chemotherapy sensitivity.

DISCUSSION

An increasing number of studies have shown that DNA damage and repair play important roles in malignant tumors, including HNSCC (25). DNA repair has been proven to be widely involved in the development, prognosis, and metastasis of HNSCC (26). Further studies on the expression profile of DNA repair genes in
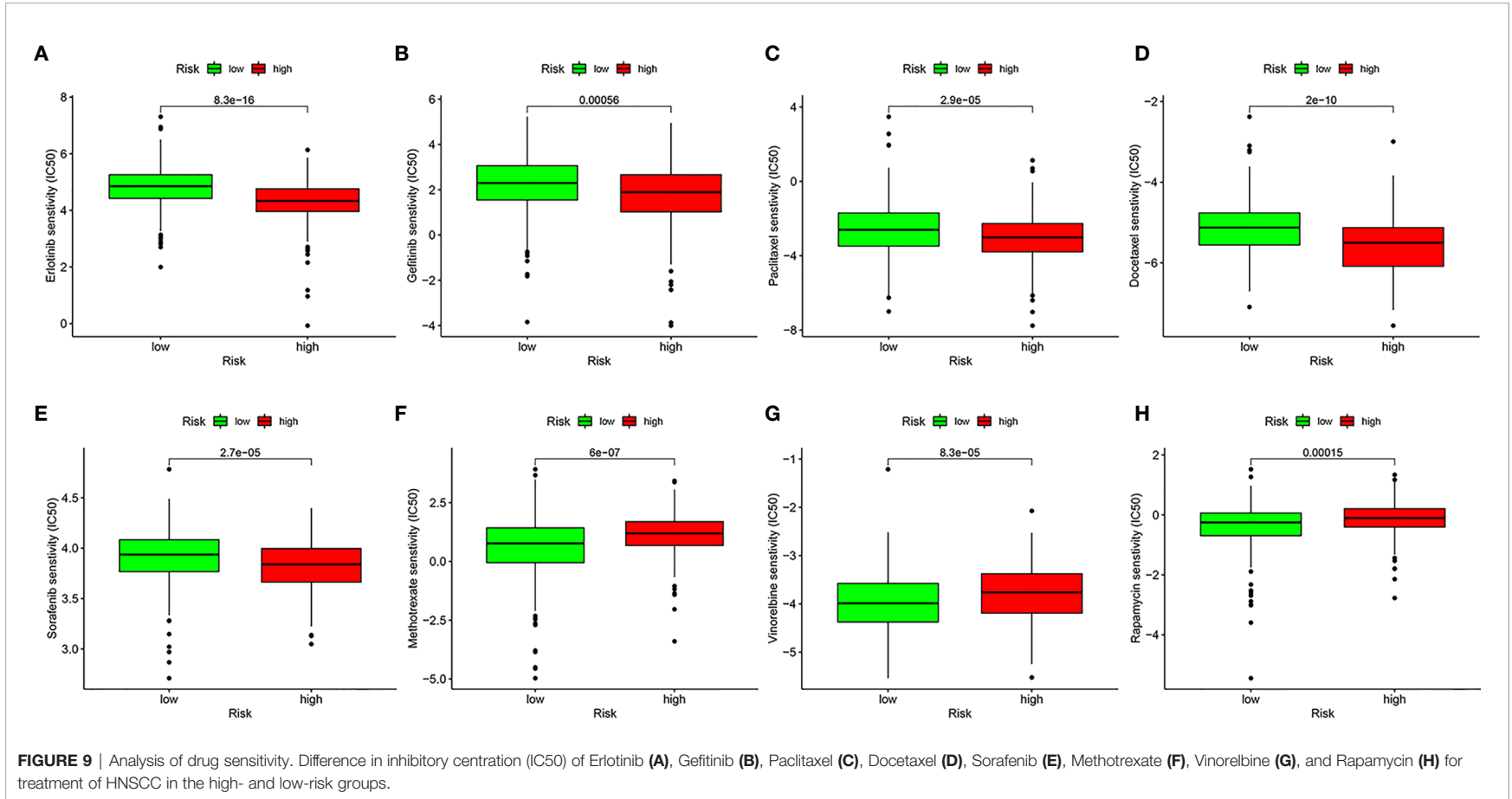
FIGURE 6 | Estimation of immune regulatory gene expression. Heatmap of negative immune regulatory gene expression **(A)**. Differential expression of CD4+ T cell and CD8+ T cell regulatory genes in the high- and low-risk groups **(B)**. Correlation between gene expression and risk scores of CTLA4 **(C)**, LAG3 **(D)**, PD1 **(E)**, PD-L1 **(F)**, and TIM3 **(G)**. Differential expression of CTLA4 **(H)**, LAG3 **(I)**, PD1 **(J)**, PD-L1 **(K)**, and TIM3 **(L)** genes in high- and low-risk groups. *P < 0.05, **P < 0.01, ***P < 0.001.

FIGURE 7 | Assessment of tumor mutation burden of HNSCC. Top 30 mutant genes of all HNSCC samples **(A)**. Top 20 mutant genes of high- **(B)** and low-risk **(C)** groups. Kaplan-Meier survival curve showing the OS differences between the high- and low-TMB groups **(D)**. Kaplan-Meier survival curve showing the OS differences in the four combinations of TMB and risk **(E)**. Correlation of TP53 expression and risk score **(F)**. Different expression of TP53 in high- and low-risk groups **(G)**.
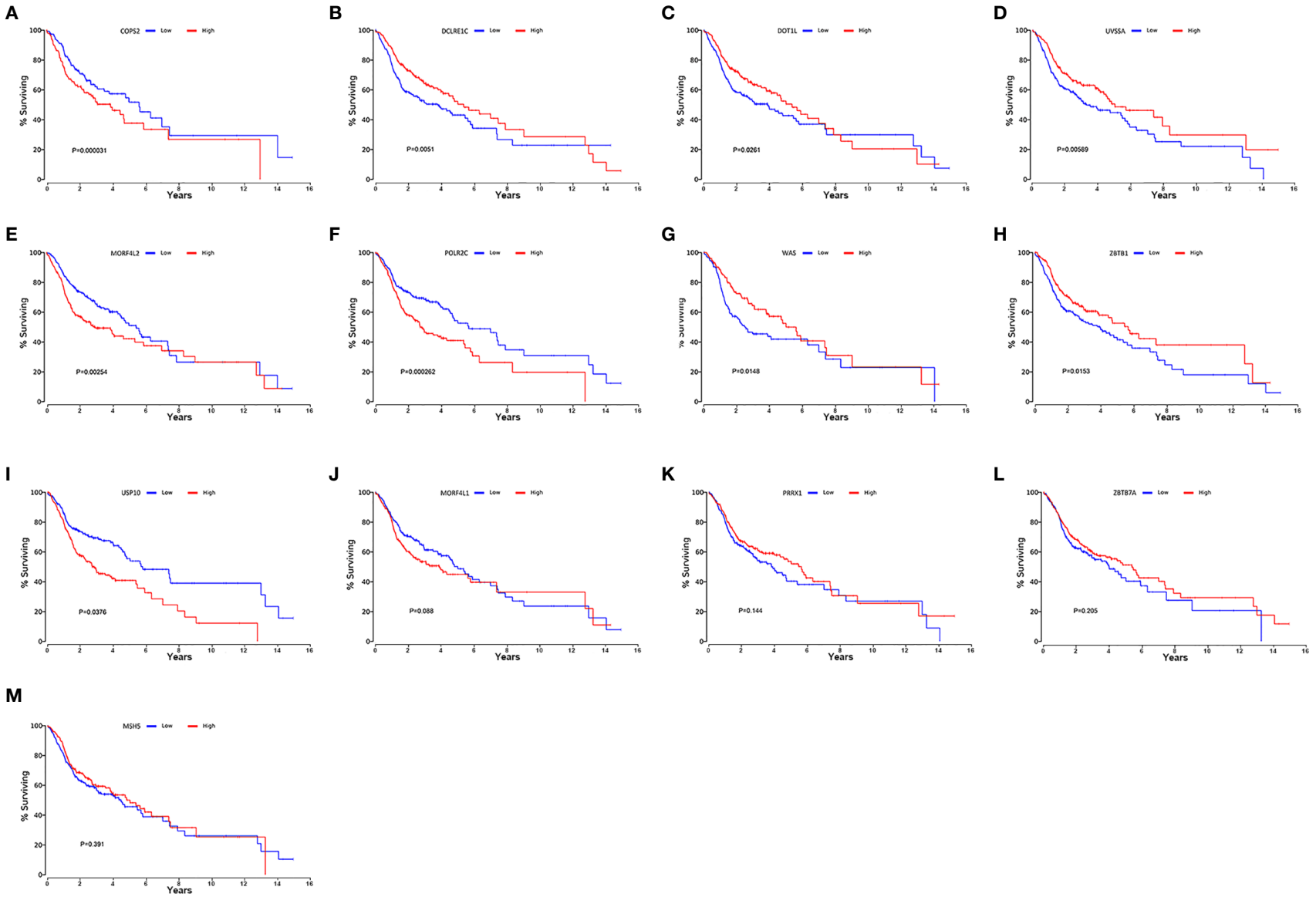
FIGURE 8 | Verification of online website Oncolnc. Kaplan-Meier survival curve from Oncolnc (http://www.oncolnc.org/) of COPS2 **(A)**, DCLRE1C **(B)**, DOT1L **(C)**, UVSSA **(D)**, MORF4L2 **(E)**, POLR2C **(F)**, WAS **(G)**, ZBTB1 **(H)** and USP10 **(I)**, MORF4L1 **(J)**, PRRX1 **(K)**, ZBTB7A **(L)**, and MSH5 **(M)** for HNSCC.

FIGURE 9 | Analysis of drug sensitivity. Difference in inhibitory centration (IC50) of Erlotinib **(A)**, Gefitinib **(B)**, Paclitaxel **(C)**, Docetaxel **(D)**, Sorafenib **(E)**, Methotrexate **(F)**, Vinorelbine **(G)**, and Rapamycin **(H)** for treatment of HNSCC in the high- and low-risk groups.

HNSCC specimens may provide new ideas to improve the clinical prognosis of patients.

A total of 545 DNA repair genes were obtained from the "GO_DNA REPAIR" gene set of the GSEA database for subsequent analysis. Through univariate and LASSO Cox regression analyses in the training set, we constructed a risk model that included 13 DNA repair genes. Patients in high-risk group had worse clinical outcomes than low-risk patients. The AUC of the ROC at 1-, 3-, and 5-year confirmed the good prediction performance of the risk score. In addition, prediction accuracy and independence were verified using univariate and multivariate Cox regression analyses. We also performed clinical subgroup validation in the internal dataset and further validated the model in the online database Oncolnc, which reflected good accuracy and repeatability of the risk model.

We illustrated the immune landscape of patients with HNSCC using CIBERSORT and ESTIMATE, including tumor-infiltrating immune cells, immune score, immune regulatory genes, and immune checkpoint genes, all of which are considered important in HNSCC (27). Comprehensive analysis revealed that the risk score was more negatively related to tumor-infiltrating cells such as naïve B cells, resting mast cells, CD8+ T cells, Tregs, and follicular helper T cells, and positively related to activated mast cells and macrophages. According to **Table S4**, the correlation between risk score and tumor-infiltrating immune cells was contributed by the influence of all risk genes on tumor-infiltrating immune cells. Tumor-infiltrating immune cells both correlated with eight gene transcripts that have a negative correlation coefficient and five gene transcripts having a positive correlation coefficient. In addition, patients in the low-risk group had higher immune scores, stromal scores, and ESTIMATE scores, which indicated that their tumor purity was lower.

In this study, some of the DRGs in the risk model have already been identified as having an important role in the immune system while others have not been well studied in the immune system at present. Decreasing the activity of DOT1L (DOT1 like histone lysine methyltransferase) through silencing or an inhibitor preferentially suppressed the production of interleukin 6 (IL-6) and interferon β (IFN-β) but not of tumor necrosis factor α (TNF-α) in macrophages triggered by Toll-like receptor (TLR) ligands or virus infection. DOT1L-mediated selective histone 3 lysine 79 (H3K79me2/3) modifications at the IL-6 and IFN-β1 promoters are required for the full activation of innate immune responses (28). DO1L plays an important role in regulating the differentiation and complete function of CD4+, CD8+T cells and B cells in the process of acquired immunity, while DO1L knockdown or mutation invalidates acquired immunity (29–32). ZBTB1 (zinc finger and BTB domain containing 1) prevents DNA damage in replicating immune progenitors, allowing the generation of B cells, T cells, and myeloid cells (33). In alveolar macrophages, antigen presentation was ZBTB7A (zinc finger and BTB domain containing 7A)-dependent where alveolar macrophages deficient in ZBTB7A failed to induce antibody production and T cell responses (34).

CD8+ T cell infiltration indicates better prognosis of patients with HNSCC (35). Because of the negative correlation between the risk score and tumor-infiltrating cells, we investigated the differential expression of negative immune regulatory genes, CD4+ T cell and CD8+ T cell regulatory genes in different groups. The results showed that almost all of these genes were highly expressed in the low-risk group, potentially due to increased infiltration of immune cells in the low-risk group samples. Subsequently, the correlation between the risk score and the expression of five immune checkpoint genes, CTLA4, LAG3, PD1, PD-L1, and TIM3, indicated that the expression of immune checkpoint genes was negatively correlated with the risk score and was highly expressed in the low-risk group, suggesting that immune checkpoint inhibitors may be beneficial to patients with HNSCC with low risk scores.

In recent years, there has been an increasing number of studies on the TMB of various tumors, including HNSCC, not only in the context of its use as a biomarker, but also in the treatment of immune checkpoint inhibitors (36). In our study, TMB was positively correlated with risk score and poorer clinical outcomes. Because TP53 showed the highest mutation rate, we compared its expression in different groups and found that it was negatively correlated with the risk score and highly expressed in the low-risk group. Our model suggested that patients with HNSCC with high risk scores were more sensitive to biological inhibitors such as erlotinib, gefitinib, and sorafenib, instead of chemotherapeutics like methotrexate. These analyses of drug sensitivity were based on "pRRophetic" package in R (20). Although the authenticity of the difference in drug sensitivity of these drugs among patients with HNSCC in different risk groups needs to be verified by further clinical trials, this model based on DNA repair genes provides the possibility for guiding clinical drug use. We speculated that the effect of immunotherapy on HNSCC would be better than that of traditional chemotherapy.

In this study, some of the DRGs in the process of modeling that have already been identified play an important role in the malignant phenotypes of various cancer types. DOT1L is involved in tumorigenesis and tumor metabolism or metastasis of ovarian cancer (37, 38), prostate cancer (39, 40), leukemia (41, 42), neuroblastoma (43), colorectal cancer (44), and breast cancer (45). PRRX1 (paired related homeobox 1), a homeodomain transcriptional factor, has been demonstrated to be important in pancreatic cancer, especially in the regulation of epithelial-to-mesenchymal transition (EMT) in pancreatic cancer (46–49). Moreover, UPS10 (ubiquitin-specific peptidase 10), a deubiquitinase, promotes proliferation of hepatocellular carcinoma by deubiquitinating and stabilizing YAP/TAZ, and suppresses lung tumorigenesis by deubiquitinating and stabilizing KLF4 (50, 51). ZBTB7A (zinc finger and BTB domain containing 7A) acts as a tumor suppressor through transcriptional repression in several carcinomas (52–54). Moreover, its mutation or downregulation promotes cancer progression (55, 56). Furthermore, its homologous gene, ZBTB1, participates in regulating the treatment effectiveness and resistance to chemotherapy (57, 58). At present, other DRGs in the model have not been studied in depth in tumors.

In general, the prognosis model constructed based on the DNA repair gene transcripts and clinical information of patients

with HNSCC in TCGA can well predict the prognosis of patients with HNSCC in the high- and low-risk groups. And this model systematically elaborated the molecular characteristics and immune microenvironment of HNSCC. The internal verification established based on the TCGA database also proved the stability of the model and provided reference value for prediction of the clinical outcomes of patients with HNSCC. In addition, the significant differences of multiple immune checkpoint genes between the high- and low-risk groups point out possible directions for the immunotherapy of patients with HNSCC.

However, we recognized that there were limitations to this study. On the one hand, the HNSCC samples involved in this study were not sufficient, and the DNA repair gene transcripts and clinical information of multiple GEO databases were incomplete, which hindered our external verification. On the other hand, the immaturity of the biobank of our institution was not enough to verify. Nevertheless, we still successfully completed external verification with the remaining genes in GSE41613 without MSH5 transcript, which further confirmed the availability and stability of the prognostic model. However, there were no significant differences in the Kaplan-Meier survival curves validated in the GSE117973 (without UVSSA), GSE27020 (without UVSSA and MSH5), and GSE65858 (without ZBTB1 and UVSSA). We assumed that the lack of a relatively important gene would reduce the predictive ability of the model, which might be the reason for the failure of the verification in GSE27020, GSE117973, and GSE65858.

CONCLUSION

In conclusion, this study constructed a 13-DRG signature for the prognosis of HNSCC, which could accurately and independently predict the clinical outcome of the patient. We then revealed the immune landscape, TMB, and sensitivity to chemotherapy drugs in different risk groups, which might be used to guide clinical treatment decisions.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/

http://asia.ensembl.org

http://biocc.hrbmu.edu.cn/TIP/index.jsp http://www.gsea-msigdb.org/gsea/msigdb/cards/GOBP_DNA_REPAIR.html

http://www.oncolnc.org/

https://www.r-project.org/

https://www.ncbi.nlm.nih.gov/geo/.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.710694/full#supplementary-material

REFERENCES

1. Puram SV, Rocco JW. Molecular Aspects of Head and Neck Cancer Therapy. *Hematol Oncol Clin North Am* (2015) 29(6):971–92. doi: 10.1016/j.hoc.2015.07.003
2. 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: A Cancer J Clin* (2018) 68 (6):394–424. doi: 10.3322/caac.21492
3. Lans H, Hoeijmakers JHJ, Vermeulen W, Marteijn JA. The DNA Damage Response to Transcription Stress. *Nat Rev Mol Cell Biol* (2019) 20(12):766–84. doi: 10.1038/s41580-019-0169-4

4. Stover EH, Konstantinopoulos PA, Matulonis UA, Swisher EM. Biomarkers of Response and Resistance to DNA Repair Targeted Therapies. *Clin Cancer Res* (2016) 22(23):5651–60. doi: 10.1158/1078-0432.Ccr-16-0247
5. Borchiellini D, Etienne-Grimaldi MC, Bensadoun RJ, Benezery K, Dassonville O, Poissonnet G, et al. Candidate Apoptotic and DNA Repair Gene Approach Confirms Involvement of ERCC1, ERCC5, TP53 and MDM2 in Radiation-Induced Toxicity in Head and Neck Cancer. *Oral Oncol* (2017) 67:70–6. doi: 10.1016/j.oraloncology.2017.02.003
6. Fan CY, Liu KL, Huang HY, Barnes EL, Swalsky PA, Bakker A, et al. Frequent Allelic Imbalance and Loss of Protein Expression of the DNA Repair Gene Hogg1 in Head and Neck Squamous Cell Carcinoma. *Lab Invest* (2001) 81 (10):1429–38. doi: 10.1038/labinvest.3780356

7. Lima LM, de Souza LR, da Silva TF, Pereira CS, Guimarães AL, de Paula AM, et al. DNA Repair Gene Excision Repair Cross Complementing-Group 1 (ERCC1) in Head and Neck Squamous Cell Carcinoma: Analysis of Methylation and Polymorphism (G19007A), Protein Expression and Association With Epidemiological and Clinicopathological Factors. *Histopathology* (2012) 60(3):489–96. doi: 10.1111/j.1365-2559.2011.04062.x

8. Tibshirani R. The Lasso Method for Variable Selection in the Cox Model. *Stat Med* (1997) 16(4):385–95. doi: 10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

9. Mounir M, Lucchetta M, Silva TC, Olsen C, Bontempi G, Chen X, et al. New Functionalities in the TCGAbiolinks Package for the Study and Integration of Cancer Data From GDC and GTEx. *PLoS Comput Biol* (2019) 15(3):e1006701. doi: 10.1371/journal.pcbi.1006701

10. Edge SB, Compton CC. The American Joint Committee on Cancer: The 7th Edition of the AJCC Cancer Staging Manual and the Future of TNM. *Ann Surg Oncol* (2010) 17(6):1471–4. doi: 10.1245/s10434-010-0985-4

11. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, et al. Ensembl 2021. *Nucleic Acids Res* (2021) 49(D1):D884–d891. doi: 10.1093/nar/gkaa942

12. Xu L, Deng C, Pang B, Zhang X, Liu W, Liao G, et al. TIP: A Web Server for Resolving Tumor Immunophenotype Profiling. *Cancer Res* (2018) 78(23):6575–80. doi: 10.1158/0008-5472.Can-18-0689

13. Blake JA, Dolan M, Drabkin H, Hill DP, Li N, Sitnikov D, et al. Gene Ontology Annotations and Resources. *Nucleic Acids Res* (2013) 49(Database issue):D530–5. doi: 10.1093/nar/gks1050

14. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway Enrichment Analysis and Visualization of Omics Data Using G:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc* (2019) 14(2):482–517. doi: 10.1038/s41596-018-0103-9

15. Lohavanichbutr P, Méndez E, Holsinger FC, Rue TC, Zhang Y, Houck J, et al. A 13-Gene Signature Prognostic of HPV-Negative OSCC: Discovery and External Validation. *Clin Cancer Res* (2013) 19(5):1197–203. doi: 10.1158/1078-0432.Ccr-12-2647

16. Fountzilas E, Kotoula V, Angouridakis N, Karasmanis I, Wirtz RM, Eleftheraki AG, et al. Identification and Validation of a Multigene Predictor of Recurrence in Primary Laryngeal Cancer. *PLoS One* (2013) 8(8):e70429. doi: 10.1371/journal.pone.0070429

17. Mock A, Plath M, Moratin J, Tapken MJ, Jäger D, Krauss J, et al. EGFR and PI3K Pathway Activities Might Guide Drug Repurposing in HPV-Negative Head and Neck Cancers. *Front Oncol* (2021) 11:678966. doi: 10.3389/fonc.2021.678966

18. Wichmann G, Rosolowski M, Krohn K, Kreuz M, Boehm A, Reiche A, et al. The Role of HPV RNA Transcription, Immune Response-Related Gene Expression and Disruptive TP53 Mutations in Diagnostic and Prognostic Profiling of Head and Neck Cancer. *Int J Cancer* (2015) 137(12):2846–57. doi: 10.1002/ijc.29649

19. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* (2010) 33(1):1–22. doi: 10.18637/jss.v033.i01

20. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust Enumeration of Cell Subsets From Tissue Expression Profiles. *Nat Methods* (2015) 12(5):453–7. doi: 10.1038/nmeth.3337

21. Weaver B, Wuensch KL. SPSS and SAS Programs for Comparing Pearson Correlations and OLS Regression Coefficients. *Behav Res Methods* (2013) 45(3):880–95. doi: 10.3758/s13428-012-0289-7

22. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring Tumour Purity and Stromal and Immune Cell Admixture From Expression Data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612

23. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res* (2018) 28(11):1747–56. doi: 10.1101/gr.239244.118

24. Geeleher P, Cox N, Huang RS. Prrophetic: An R Package for Prediction of Clinical Chemotherapeutic Response From Tumor Gene Expression Levels. *PLoS One* (2014) 9(9):e107468. doi: 10.1371/journal.pone.0107468

25. Nickoloff JA, Jones D, Lee SH, Williamson EA, Hromas R. Drugging the Cancers Addicted to DNA Repair. *J Natl Cancer Inst* (2017) 109(11):djx059. doi: 10.1093/jnci/djx059

26. Moutafi M, Economopoulou P, Rimm D, Psyrri A. PARP Inhibitors in Head and Neck Cancer: Molecular Mechanisms, Preclinical and Clinical Data. *Oral Oncol* (2021) 117:105292. doi: 10.1016/j.oraloncology.2021.105292

27. Cillo AR, Kürten CHL, Tabib T, Qi Z, Onkar S, Wang T, et al. Immune Landscape of Viral- and Carcinogen-Driven Head and Neck Cancer. *Immunity* (2020) 52(1):183–99.e9. doi: 10.1016/j.immuni.2019.11.014

28. Chen X, Liu X, Zhang Y, Huai W, Zhou Q, Xu S, et al. Methyltransferase Dot1l Preferentially Promotes Innate IL-6 and IFN-β Production by Mediating H3K79me2/3 Methylation in Macrophages. *Cell Mol Immunol* (2020) 17(1):76–84. doi: 10.1038/s41423-018-0170-4

29. Scheer S, Runting J, Bramhall M, Russ B, Zaini A, Ellemor J, et al. The Methyltransferase DOT1L Controls Activation and Lineage Integrity in CD4(+) T Cells During Infection and Inflammation. *Cell Rep* (2020) 33(11):108505. doi: 10.1016/j.celrep.2020.108505

30. Kwesi-Maliepaard EM, Aslam MA, Alemdehy MF, van den Brand T, McLean C, Vlaming H, et al. The Histone Methyltransferase DOT1L Prevents Antigen-Independent Differentiation and Safeguards Epigenetic Identity of CD8(+) T Cells. *Proc Natl Acad Sci USA* (2020) 117(34):20706–16. doi: 10.1073/pnas.1920372117

31. Kealy L, Di Pietro A, Hailes L, Scheer S, Dalit L, Groom JR, et al. The Histone Methyltransferase DOT1L Is Essential for Humoral Immune Responses. *Cell Rep* (2020) 33(11):108504. doi: 10.1016/j.celrep.2020.108504

32. Aslam MA, Alemdehy MF, Kwesi-Maliepaard EM, Muhaimin FI, Caganova M, Pardieck IN, et al. Histone Methyltransferase DOT1L Controls State-Specific Identity During B Cell Differentiation. *EMBO Rep* (2021) 22(2):e51184. doi: 10.15252/embr.202051184

33. Cao X, Lu Y, Zhang X, Kovalovsky D. Zbtb1 Safeguards Genome Integrity and Prevents P53-Mediated Apoptosis in Proliferating Lymphoid Progenitors. *J Immunol* (2016) 197(4):1199–211. doi: 10.4049/jimmunol.1600013

34. Nayak DK, Zhou F, Xu M, Huang J, Tsuji M, Yu J, et al. Zbtb7a Induction in Alveolar Macrophages Is Implicated in Anti-HLA-Mediated Lung Allograft Rejection. *Sci Transl Med* (2017) 9(398):eaal1243. doi: 10.1126/scitranslmed.aal1243

35. Mandal R, Şenbabaoğlu Y, Desrichard A, Havel JJ, Dalin MG, Riaz N, et al. The Head and Neck Cancer Immune Landscape and its Immunotherapeutic Implications. *JCI Insight* (2016) 1(17):e89829. doi: 10.1172/jci.insight.89829

36. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of Tumor Mutation Burden as an Immunotherapy Biomarker: Utility for the Oncology Clinic. *Ann Oncol* (2019) 30(1):44–56. doi: 10.1093/annonc/mdy495

37. Zhang X, Liu D, Li M, Cao C, Wan D, Xi B, et al. Prognostic and Therapeutic Value of Disruptor of Telomeric Silencing-1-Like (DOT1L) Expression in Patients With Ovarian Cancer. *J Hematol Oncol* (2017) 10(1):29. doi: 10.1186/s13045-017-0400-8

38. Salvati A, Gigantino V, Nassa G, Giurato G, Alexandrova E, Rizzo F, et al. The Histone Methyltransferase DOT1L Is a Functional Component of Estrogen Receptor Alpha Signaling in Ovarian Cancer Cells. *Cancers (Basel)* (2019) 11(11):1720. doi: 10.3390/cancers11111720

39. Vatapalli R, Sagar V, Rodriguez Y, Zhao JC, Unno K, Pamarthy S, et al. Histone Methyltransferase DOT1L Coordinates AR and MYC Stability in Prostate Cancer. *Nat Commun* (2020) 11(1):4153. doi: 10.1038/s41467-020-18013-7

40. Thomas T. DOT1L in Prostate Cancer. *Nat Rev Urol* (2020) 17(10):544. doi: 10.1038/s41585-020-0374-0

41. Daigle SR, Olhava EJ, Therkelsen CA, Majer CR, Sneeringer CJ, Song J, et al. Selective Killing of Mixed Lineage Leukemia Cells by a Potent Small-Molecule DOT1L Inhibitor. *Cancer Cell* (2011) 20(1):53–65. doi: 10.1016/j.ccr.2011.06.009

42. Okada Y, Feng Q, Lin Y, Jiang Q, Li Y, Coffield VM, et al. Hdot1l Links Histone Methylation to Leukemogenesis. *Cell* (2005) 121(2):167–78. doi: 10.1016/j.cell.2005.02.020

43. Wong M, Tee AEL, Milazzo G, Bell JL, Poulos RC, Atmadibrata B, et al. The Histone Methyltransferase DOT1L Promotes Neuroblastoma by Regulating Gene Transcription. *Cancer Res* (2017) 77(9):2522–33. doi: 10.1158/0008-5472.Can-16-1663

44. Yang L, Lei Q, Li L, Yang J, Dong Z, Cui H. Silencing or Inhibition of H3K79 Methyltransferase DOT1L Induces Cell Cycle Arrest by Epigenetically Modulating C-Myc Expression in Colorectal Cancer. *Clin Epigenet* (2019) 11(1):199. doi: 10.1186/s13148-019-0778-y

45. Byun WS, Kim WK, Han HJ, Chung HJ, Jang K, Kim HS, et al. Targeting Histone Methyltransferase DOT1L by a Novel Psammaplin A Analog Inhibits Growth and Metastasis of Triple-Negative Breast Cancer. *Mol Ther Oncolytics* (2019) 15:140–52. doi: 10.1016/j.omto.2019.09.005

46. Takano S, Reichert M, Bakir B, Das KK, Nishida T, Miyazaki M, et al. Prrx1 Isoform Switching Regulates Pancreatic Cancer Invasion and Metastatic Colonization. *Genes Dev* (2016) 30(2):233–47. doi: 10.1101/gad.263327.115

47. Reichert M, Takano S, von Burstin J, Kim SB, Lee JS, Ihida-Stansbury K, et al. The Prrx1 Homeodomain Transcription Factor Plays a Central Role in Pancreatic Regeneration and Carcinogenesis. *Genes Dev* (2013) 27(3):288–300. doi: 10.1101/gad.204453.112

48. Marchand B, Pitarresi JR, Reichert M, Suzuki K, Laczkó D, Rustgi AK. PRRX1 Isoforms Cooperate With FOXM1 to Regulate the DNA Damage Response in Pancreatic Cancer Cells. *Oncogene* (2019) 38(22):4325–39. doi: 10.1038/s41388-019-0725-6

49. Feldmann K, Maurer C, Peschke K, Teller S, Schuck K, Steiger K, et al. Mesenchymal Plasticity Regulated by Prrx1 Drives Aggressive Pancreatic Cancer Biology. *Gastroenterology* (2021) 160(1):346–61.e24. doi: 10.1053/j.gastro.2020.09.010

50. Wang X, Xia S, Li H, Wang X, Li C, Chao Y, et al. The Deubiquitinase USP10 Regulates KLF4 Stability and Suppresses Lung Tumorigenesis. *Cell Death Differ* (2020) 27(6):1747–64. doi: 10.1038/s41418-019-0458-7

51. Zhu H, Yan F, Yuan T, Qian M, Zhou T, Dai X, et al. USP10 Promotes Proliferation of Hepatocellular Carcinoma by Deubiquitinating and Stabilizing YAP/TAZ. *Cancer Res* (2020) 80(11):2204–16. doi: 10.1158/0008-5472.Can-19-2388

52. Shi DB, Wang YW, Xing AY, Gao JW, Zhang H, Guo XY, et al. C/Ebpα-Induced miR-100 Expression Suppresses Tumor Metastasis and Growth by Targeting ZBTB7A in Gastric Cancer. *Cancer Lett* (2015) 369(2):376–85. doi: 10.1016/j.canlet.2015.08.029

53. Liu XS, Haines JE, Mehanna EK, Genet MD, Ben-Sahra I, Asara JM, et al. ZBTB7A Acts as a Tumor Suppressor Through the Transcriptional Repression of Glycolysis. *Genes Dev* (2014) 28(17):1917–28. doi: 10.1101/gad.245910.114

54. Han D, Chen S, Han W, Gao S, Owiredu JN, Li M, et al. ZBTB7A Mediates the Transcriptional Repression Activity of the Androgen Receptor in Prostate Cancer. *Cancer Res* (2019) 79(20):5260–71. doi: 10.1158/0008-5472.Can-19-0815

55. Liu XS, Liu Z, Gerarduzzi C, Choi DE, Ganapathy S, Pandolfi PP, et al. Somatic Human ZBTB7A Zinc Finger Mutations Promote Cancer Progression. *Oncogene* (2016) 35(23):3071–8. doi: 10.1038/onc.2015.371

56. Alam H, Li N, Dhar SS, Wu SJ, Lv J, Chen K, et al. Hp1γ Promotes Lung Adenocarcinoma by Downregulating the Transcription-Repressive Regulators NCOR2 and ZBTB7A. *Cancer Res* (2018) 78(14):3834–48. doi: 10.1158/0008-5472.Can-17-3571

57. Zhang P, Yang Y, Qian K, Li L, Zhang C, Fu X, et al. A Novel Tumor Suppressor ZBTB1 Regulates Tamoxifen Resistance and Aerobic Glycolysis Through Suppressing HER2 Expression in Breast Cancer. *J Biol Chem* (2020) 295(41):14140–52. doi: 10.1074/jbc.RA119.010759

58. Williams RT, Guarecuco R, Gates LA, Barrows D, Passarelli MC, Carey B, et al. ZBTB1 Regulates Asparagine Synthesis and Leukemia Cell Response to L-Asparaginase. *Cell Metab* (2020) 31(4):852–61.e6. doi: 10.1016/j.cmet.2020.03.008