*CORRESPONDENCE
Yan-jun Wang,
yjwang@hust.edu.cn

†These authors have contributed equally
to this work

Nine-gene signature and nomogram for predicting survival in patients with head and neck squamous cell carcinoma

Fan Yang†, Liu-qing Zhou†, Hui-wen Yang† and Yan-jun Wang*

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

Background: Head and neck squamous cell carcinomas (HNSCCs) are derived from the mucosal linings of the upper aerodigestive tract, salivary glands, thyroid, oropharynx, larynx, and hypopharynx. The present study aimed to identify the novel genes and pathways underlying HNSCC. Despite the advances in HNSCC research, diagnosis, and treatment, its incidence continues to rise, and the mortality of advanced HNSCC is expected to increase by 50%. Therefore, there is an urgent need for effective biomarkers to predict HNSCC patients' prognosis and provide guidance to the personalized treatment.

Methods: Both HNSCC clinical and gene expression data were abstracted from The Cancer Genome Atlas (TCGA) database. Intersecting analysis was adopted between the gene expression matrix of HNSCC patients from TCGA database to extract TME-related genes. Differential gene expression analysis between HNSCC tissue samples and normal tissue samples was performed by R software. Then, HNSCC patients were categorized into clusters 1 and 2 via NMF. Next, TME-related prognosis genes ($p < 0.05$) were analyzed by univariate Cox regression analysis, LASSO Cox regression analysis, and multivariate Cox regression analysis. Finally, nine genes were selected to construct a prognostic risk model and a prognostic gene signature. We also established a nomogram using relevant clinical parameters and a risk score. The Kaplan−Meier curve, survival analysis, time-dependent receiver operating characteristic (ROC) analysis, decision curve analysis (DCA), and the concordance index (C-index) were carried out to assess the accuracy of the prognostic risk model and nomogram. Potential molecular mechanisms were revealed by gene set enrichment analysis (GSEA). Additionally, gene correlation analysis and immune cell correlation analysis were conducted for further enriching our results.

Results: A novel HNSCC prognostic model was established based on the nine genes (GTSE1, LRRN4CL, CRYAB, SHOX2, ASNS, KRT23, ANGPT2, HOXA9, and CARD11). The value of area under the ROC curves (AUCs) (0.769, 0.841, and 0.816) in TCGA whole set showed that the model effectively predicted the 1-, 3-, and 5-year overall survival (OS). Results of the Cox regression assessment confirmed the nine-gene signature as a reliable independent prognostic factor in HNSCC patients. The prognostic nomogram developed using multivariate

Cox regression analysis showed a superior C-index over other clinical signatures. Also, the calibration curve had a high level of concordance between estimated OS and the observed OS. This showed that its clinical net can precisely estimate the one-, three-, and five-year OS in HNSCC patients. The gene set enrichment analysis (GSEA) to some extent revealed the immune- and tumor-linked cascades.

Conclusion: In conclusion, the TME-related nine-gene signature and nomogram can effectively improve the estimation of prognosis in patients with HNSCC.

Background

Globally, head and neck squamous cell carcinoma (HNSCC) is the sixth commonest cancer (Yang et al., 2017). HNSCC is the most common pathological type of head and neck tumors, accounting for over 90% of the cases (Shield et al., 2017). Despite advances in early diagnosis and multi-disciplinary management of cancer, the 5-year overall survival (OS) of HNSCC remains at 50% (González-Arriagada et al., 2018). Results of various studies have shown that gene signatures can predict the prognosis of patients with HNSCC. Multiview data, generated through data integration, are receiving more attention from scholars (Ge et al., 2017). However, gene expression datasets are characterized by high dimensionality, small sample sizes, and sample imbalance. Non-negative matrix factorization (NMF) is frequently used in data analysis to reduce dimension. Its applications in the analysis of gene expression data include sample clustering and feature selection (Liu et al., 2018). The tumor microenvironment (TME) is mainly composed of tumor cells and tumor-invading immune cells admixed with the stromal component. The TME of HNSCC harbors transformed cells, immune cells, and stromal cellular elements (Puram et al., 2017). The TME can have adverse or beneficial consequences. The TME can promote tumor growth and progression through the immune cells which are affected easily (Kim and Bae, 2016). The immune invasive landscape of HNSCC has been elucidated in a previous study (Zhang et al., 2020). A separate study has also performed a systematic analysis of cellular interactions in TME of HNSCC (Huo et al., 2020). In the present study, gene expression profiles of HNSCC TME-related genes were obtained and utilized to create a robust prognostic model and establish a prognostic nomogram.

Materials and methods

Sample datasets

We used The Cancer Genome Atlas (TCGA) database (https://cancergenome.nih.gov/) to retrieve HNSCC gene expression data and clinical data. Finally, we downloaded 213 samples (199 HNSCC tissue samples and 14 normal tissue samples) of HNSCC gene expression data. TCGA HNSCC clinical data of 511 patients were also downloaded. The clinical characteristics, including survival status, survival outcomes, age, gender, grade, TNM classification, and stage, were all collected. The GSE16076 dataset with 270 samples (including gene expression data and clinical data) from the Gene Expression Omnibus (GEO) database was downloaded to externally validate the reliability of the nine-gene signature.

Identification of TME-related genes

First, intersecting analysis was adopted between the gene expression matrix of HNSCC patients from TCGA database to extract TME-related genes. Then, differential expression analysis was performed to filter differentially expressed TME-related genes between HNSCC tissue samples and normal tissue samples by using the "limma" R package. The volcano plot and heatmap were generated to visualize the distribution of the identified TME-related genes by "ggplot2" and "pheatmap" R packages, respectively.

Identification and evaluation of subgroups

After differential expression analysis, NMF clustering algorithm was applied to the TME-related genes to classify HNSCC patients using the "NMF" R package. An elementary classification of HNSCC patient subgroups was set from 2 to 10. The optimal value for consensus clustering was identified as 2 by the NMF rank survey. Then, the consensus heatmap was generated to view the distribution characteristics among C1 and C2 groups. The Kaplan–Meier curves were applied to explore the discrimination between C1 and C2 groups in HNSCC patients' OS using "survival" and "survminer" R packages. The statistical difference was evaluated by the log-rank test. In

addition, we also adopted the MCP-counter algorithm to assess the difference of tumor-infiltrating immune cells between different clusters (Becht et al., 2016).

Construction and validation of the prognostic risk model

Differentially expressed TME-related genes between the HNSCC tissue samples and normal tissue samples were abstracted from the previously downloaded TCGA HNSCC gene expression data. From this process, we removed six HNSCC tumor tissue samples. Then, we combined TCGA-HNSCC gene expression data along with clinical data using the "limma" R package. Patients with incomplete clinical data were excluded. Based on previous efforts, a list of 193 TME-related tumor samples with essential clinical information and an OS of >1 month was extracted. Next, total samples from TCGA database were split randomly into the training set (n = 137) and the testing set (n = 56). The training set was applied to determine the TME-related genes as prognostic biomarkers along with the signature. Univariate COX regression analysis, LASSO Cox regression analysis, and multivariate Cox regression analysis were conducted in this process using the "survival" and "glmnet" R packages. Ultimately, nine genes were recognized as prognostic biomarkers to construct the prognostic risk model. The risk score for each HNSCC patient was determined using the same predictive gene signature-based approach. TCGA testing set was used for internal validation for the nine-gene signature. The time-dependent receiver operating characteristic (ROC) analysis and Kaplan–Meier assessments were then used to test the prognostic gene signature using "timeROC" and "survival" R packages. Furthermore, the Wilcoxon signed-rank test was adopted to compare HNSCC tissue samples with normal tissue samples. Statistical significance was set at $p < 0.05$.

Construction and verification of a predictive nomogram

Nomograms are visual statistical tools that combine multiple clinical and pathological factors to estimate the survival outcomes of patients with cancer (Iasonos et al., 2008). The independent clinical parameters were merged into a nomogram to estimate 1-, 3-, and 5-year survival of patients with HNSCC. Cox regression analysis was conducted to analyze the retrieved clinical parameters from TCGA-HNSCC clinical data using the "survival" R package. Clinical parameters ($p < 0.05$) significantly related to OS were used to establish the predictive nomogram using "regplot" and "rms" R packages. The nomogram was then verified based on the discrimination along with calibration curves. The concordance index (C-index) of the nomogram was determined for each independent clinical parameter and risk score to evaluate the nomogram preceding others through a bootstrap approach with 1,000 resamples. Decision curve

analysis (DCA) was applied to determine the clinical net benefit of each clinical parameter and showed that the nomogram harbored the highest net benefit, followed by the risk model. Generally, the nomogram was significantly superior at predicting clinical outcomes (Vickers and Elkin, 2006).

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) was performed using KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis. Furthermore, Gene Ontology (GO) functional annotation was used to identify the active genes in various groups of the predictive gene signature in the present study (Subramanian et al., 2005). The differentially expressed TME-related genes were stratified into the high- and low-risk groups based on the median risk score using the "clusterProfiler" R package. The low- and high-risk groups were then subjected to GSEA using the "enrichplot" R package. The statistical significance function was indicated by $p < 0.05$.

Differentiating performance of the prognostic risk model

The present study used survival analysis to assess the differential diagnostic potential of the risk score. Patients were stratified by stage, grade, age, and gender. Then, we conducted subgroup analysis to confirm the prognostic significance of the risk model in different subgroups.

External validation of the nine-gene signature

We carefully retrieved and selected three previously established prognostic risk models from PubMed and obtained the relevant genes contained in each model (Wang et al., 2020; Yao et al., 2020; Xiong et al., 2021). The included gene signatures were all validated as reliable prognostic models for forecasting the HNSCC patients' OS. Then, the Kaplan–Meier curve, time-independent ROC analysis, and the C-index analysis were used to confirm the advantages of the prognostic model established by us. In addition, we downloaded an independent dataset, GSM16076, to further validate the nine-gene signature. The obtained results showed that the prognostic model of the current study is significantly reliable.

Correlation analysis

Using the "corrplot" R package, immune cell correlation analysis was conducted for further enriching our results.

FIGURE 1
Flow chart of the present study.

FIGURE 2
(A) Heatmap to show TME-related differentially expressed genes between HNSCC and normal tissue samples. **(B)** Volcano plot displays 1,335 TME-related differentially expressed genes in TCGA HNSCC cohort. **(C)** Non-negative matrix factorization (NMF) clustering was conducted to classify HNSCC patients into 2 to 10 different subtypes, and relevant heatmaps were generated. **(D)** NMF rank survey with multiple parameters (including dispersion, cophenetic, residuals, evar, rss, and silhouette coefficients) to determine the optimal value for consensus clustering.

FIGURE 3
(A) Consensus map of the NMF clustering data of TCGA HNSCC cohort. Rank = 2 means that HNSCC patients were separated into two groups.
(B) Survival curve of OS (P=<0.001) in cluster 1 and 2. **(C)** Sankey plot to show the association between different subtypes and immune subtypes. **(D)** Differential analysis of tumor-infiltrating immune cells was conducted using MCP-counter.

Furthermore, we also conducted Spearman correlation analysis using "corrplot" R package to demonstrate the relationship between risk score and genes.
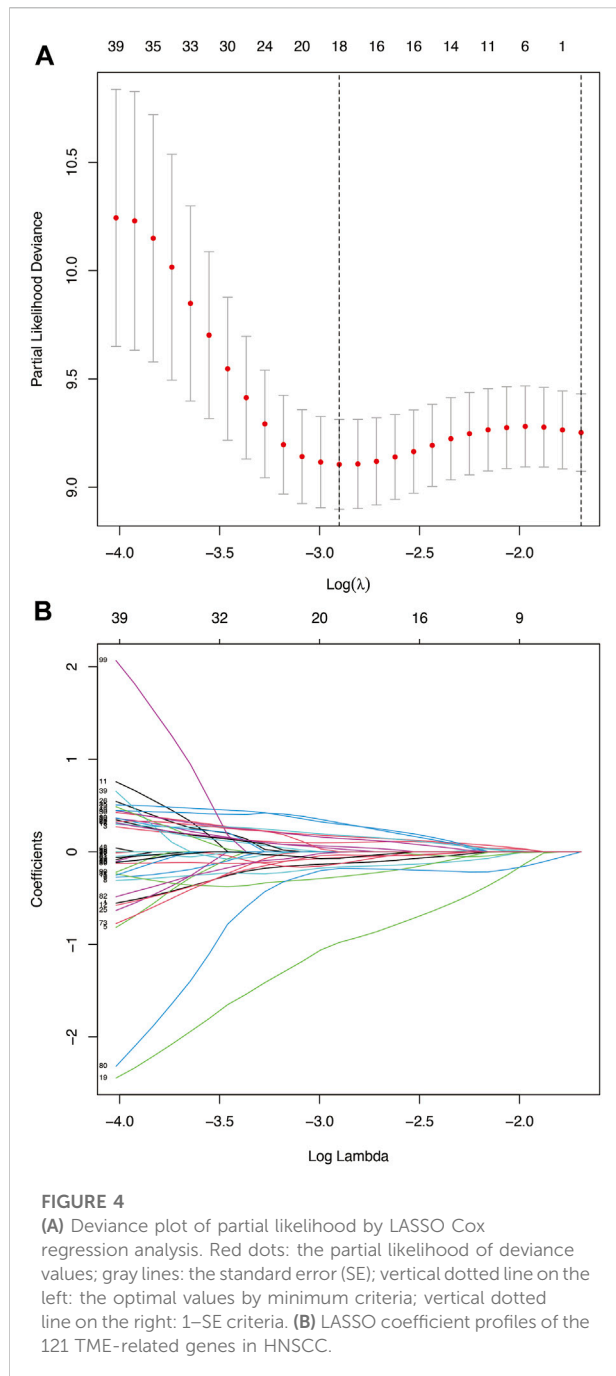
Results

Identification of TME-related genes

The flow chart in Figure 1 presents the analytical process of this study. Analysis of the data of 213 samples (199 HNSCC tumor tissue samples and 14 normal tissue samples) obtained from TCGA database was performed. Based on the gene expression levels in TME of the 213 TCGA HNSCC samples, the TME-related genes were selected ($p < 0.05$). After differential expression analysis, a total of 1,335 genes were identified as differentially expressed TME-related genes (Figure 2A). The volcano plot was applied to visualize the upregulated and downregulated genes (Figure 2B).

Identification and evaluation of subgroups

Then, we applied the NMF clustering algorithm method based on the 1,335 identified TME-related genes. Patients were first classified into 2 to 10 different subtypes, and relevant heatmaps were generated (Figure 2C). According to the parameters such as cophenetic, dispersion, silhouette, and sparseness in the NMF rank survey, the optimal number of the cluster was identified as 2 (Figure 2D). HNSCC patients were divided into two subgroups (cluster 1 and cluster 2) (Figure 3A). OS was compared between clusters 1 and 1 to further understand the clustering finding and its links with survival outcomes. According to this assessment, patients in cluster 2 had a superior OS to those in cluster 1 ($p < 0.001$, Figure 3B). The Sankey plot could also be applied to reveal the association between different immune subtypes and clusters (Figure 3C). In addition, the MCP-counter algorithm was used to evaluate the infiltration of the immune cells in cluster 1 and cluster 2. The results revealed that cluster 2 had a higher infiltration level of cytotoxic lymphocytes, T cells, B lineage, CD8+ T cells, NK

FIGURE 4
(A) Deviance plot of partial likelihood by LASSO Cox regression analysis. Red dots: the partial likelihood of deviance values; gray lines: the standard error (SE); vertical dotted line on the left: the optimal values by minimum criteria; vertical dotted line on the right: 1–SE criteria. **(B)** LASSO coefficient profiles of the 121 TME-related genes in HNSCC.

cells, myeloid dendritic cells, and monocytic lineages than cluster 1 (Figure 3D).

Construction and internal validation of the nine-gene signature
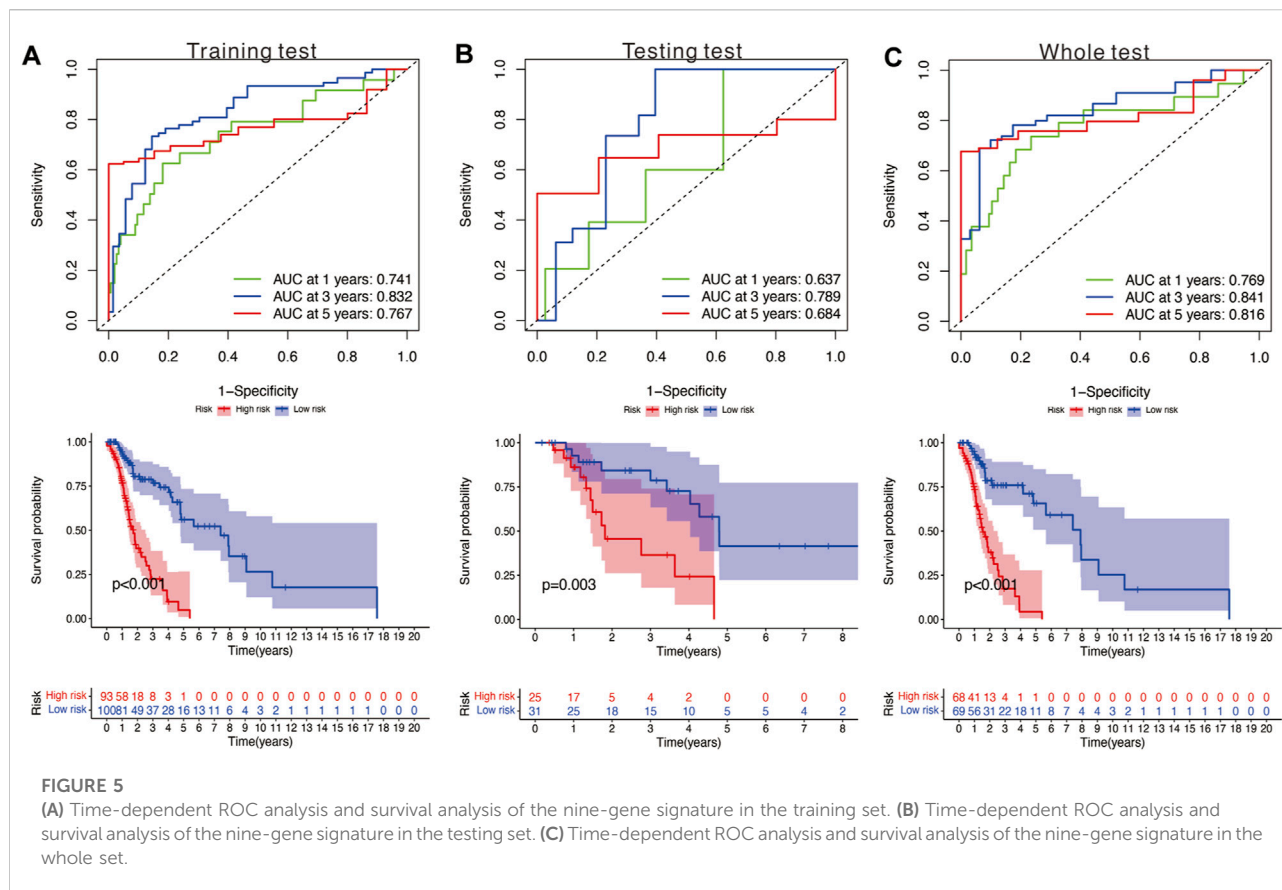
A total of 193 patients (including 1,335 TME-related genes) included in the survival analysis underwent a follow-up

exceeding one month and were randomly assigned into the training set (n = 137) and testing set (n = 56). Data from the training set were retrieved using the "survival" R package and subjected to univariate Cox regression analysis. Subsequently, 121 TME-related genes were selected based on a $p < 0.05$ cut-off value for LASSO Cox regression analysis (Figure 4A). Results of the LASSO Cox regression analysis detected 18 TME-related genes. Then, using multivariate Cox regression analysis, nine genes (GTSE1, LRRN4CL, CRYAB, SHOX2, ASNS, KRT23, ANGPT2, HOXA9, and CARD11) were determined to establish the prognostic risk model. The nine-gene signature as a prognostic marker was an independent prognostic factor tested by the multivariate Cox regression analysis. Significance was defined as $p < 0.05$. Next, nine genes and their corresponding regression coefficients were used to calculate risk scores for each sample using the following formula: risk score = sum of each gene's (regression coefficient × gene expression value). The optimal risk score cutoff was calculated from the median risk score using the "survminer" R package in the training set. Patients with risk scores above the median scores were assigned to the high-risk group, whereas those with scores below the median risk score were assigned to the low-risk group.

We utilized the time-dependent ROC analysis and Kaplan–Meier analysis to verify the prognostic value of the risk model. In the training set, the AUC (area under the ROC curve) values for predicting the 1-, 3-, and 5-year OS were 0.741, 0.832, and 0.767, respectively (Figure 5A). In the testing set, the AUC values for predicting the 1-, 3-, and 5-year OS were 0.637, 0.789, and 0.684, respectively (Figure 5B). In the whole set, the AUC values were 0.769, 0.841, and 0.816, respectively (Figure 5C). Moreover, HNSCC patients in the high-risk group showed significantly poorer OS than those in the low-risk group (training test: $p < 0.001$; testing set: $p = 0.003$; and whole set: $p < 0.001$).

Development and validation of a predictive nomogram

Seven independent prognostic factors (gender, N, age, T, stage, risk score, and grade) were used to construct a nomogram for predicting the 1-, 3-, and 5-year OS for patients with HNSCC (Figure 6A). The X-axis indicates the nomogram's estimated survival probability, while the Y-axis denotes the actual survival probability. The dotted line (45° diagonal line) between the calibration curves indicates full agreement between the actual probability and the observed probability. The calibration curves showed that the nomogram accurately predicted the 1-, 3-, and 5-year OS (Figure 6B). The estimated probability ROC curves were utilized to evaluate the sensitivity and specificity of the nine-gene signature and in predicting the survival. Among other prognostic factors (risk score, age, grade, gender, and stage), the constructed nomogram had the highest AUC value (0.981). The AUC values of the risk score, age, gender, grade, and stage were 0.728, 0.617, 0.425, 0.710, and

FIGURE 5
(A) Time-dependent ROC analysis and survival analysis of the nine-gene signature in the training set. **(B)** Time-dependent ROC analysis and survival analysis of the nine-gene signature in the testing set. **(C)** Time-dependent ROC analysis and survival analysis of the nine-gene signature in the whole set.

0.609, respectively (Figure 6C). Analysis of the nomogram's clinical net benefit using the DCA test revealed that the model combining other independent prognostic factors had the best clinical net benefit when compared with any individual feature (Figure 6D).

Gene set enrichment analysis

The results of the GSEA of the nine-gene signature (Figure 7) revealed that oncological signatures (cornification, epiderma cell differentiation, epidermis development, and keratinization) were enriched in the high-risk group, whereas immune-related cascades, consisting of immune response activation, B cell activation, adaptive immune response, adaptive immune response based on somatic recombination of immune receptors built, and antigen receptor-mediated signaling cascade, were enriched in the low-risk group.
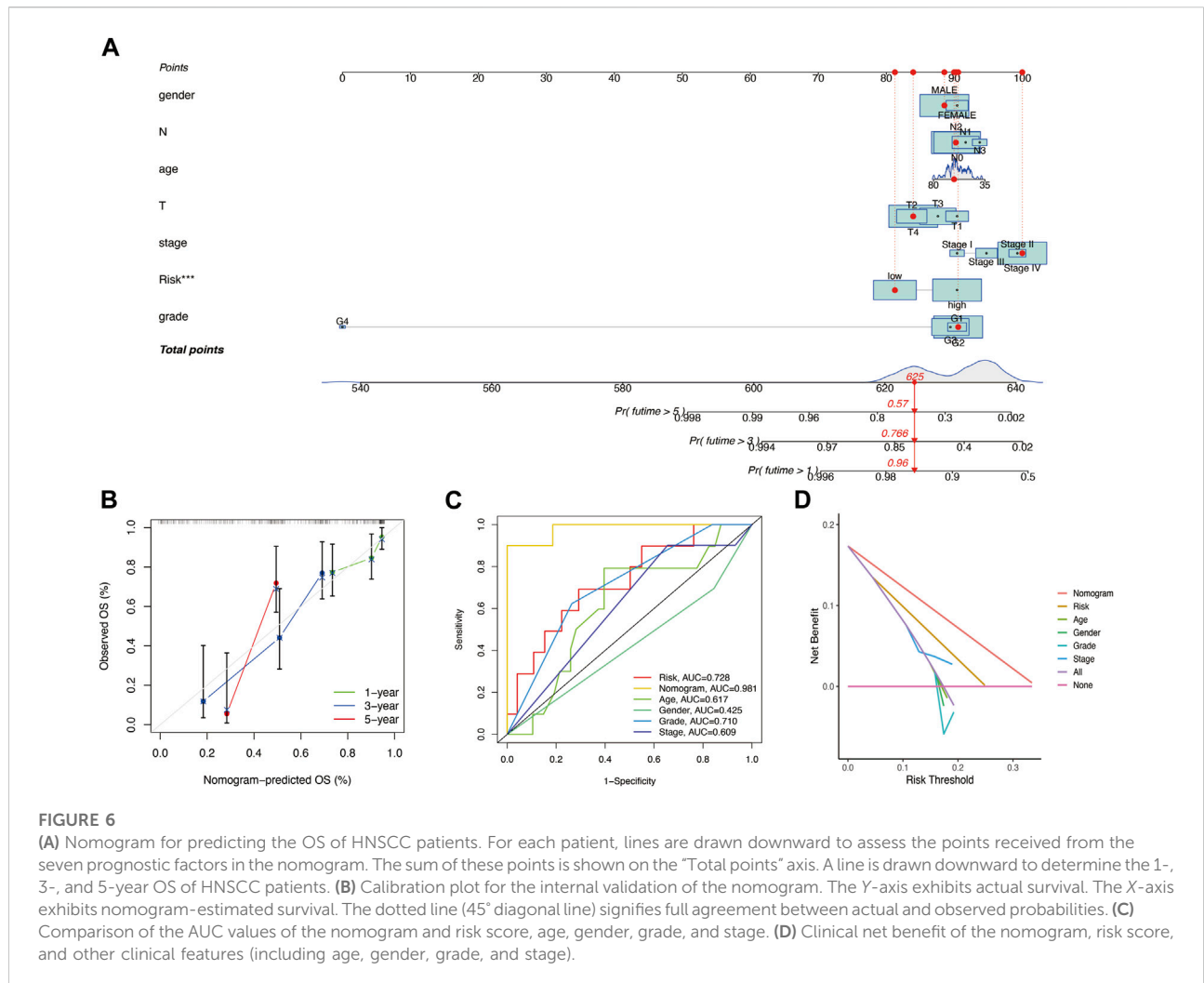
Differentiation of the prognostic performance of the signature

Stratification survival analysis was performed to assess the predictive performance of the nine-gene signature in multiple

HNSCC subtypes. Subgroup analysis, including stage (Figures 8A,B), grade (Figures 8C,D), age (Figures 8E,F), and gender (Figures 8G,H), was conducted to evaluate the different performances of the nine-gene signature. The results revealed that the high-risk group was associated with poorer OS and significantly differentiated the low-risk group into different subgroups (all $p < 0.05$).

External validation of the nine-gene signature

The results of time-independent ROC analysis and survival analysis indicated that the nine-gene signature had a robust and reliable performance in predicting OS. The AUC values of the nine-gene signature (TME signature) for predicting 1-, 3-, and 5-year OS were 0.741, 0.832, and 0.767, respectively (Figure 9A). Each previous gene signature was validated as a reliable prognostic model (Figures 9B–D). The results of the C-index analysis and RMS values demonstrated that the nine-gene signature was superior to other signatures in predicting the OS of HNSCC patients (Figures 10A,B). External validation of the nine-gene signature was performed using the GSE16076 dataset from the

FIGURE 6

(A) Nomogram for predicting the OS of HNSCC patients. For each patient, lines are drawn downward to assess the points received from the seven prognostic factors in the nomogram. The sum of these points is shown on the "Total points" axis. A line is drawn downward to determine the 1-, 3-, and 5-year OS of HNSCC patients. **(B)** Calibration plot for the internal validation of the nomogram. The *Y*-axis exhibits actual survival. The *X*-axis exhibits nomogram-estimated survival. The dotted line (45° diagonal line) signifies full agreement between actual and observed probabilities. **(C)** Comparison of the AUC values of the nomogram and risk score, age, gender, grade, and stage. **(D)** Clinical net benefit of the nomogram, risk score, and other clinical features (including age, gender, grade, and stage).

GEO database. The AUC values for predicting 1-, 3-, and 5-year OS were unknown, 0.558, and 0.623, respectively (Figure 10C). The validation of the GSE16076 dataset revealed that patients in the high-HNSCC risk group had a lower 5-year OS than those in the low-HNSCC risk group (*p* = 0.046), which was consistent with the findings from TCGA cohort (Figure 10D). Altogether, these results demonstrated that the nine-gene signature effectively predicted the survival of HNSCC patients.

Correlation analysis

The risk model was negatively correlated with CD8+ T cells, B lineage, monocytic lineage, myeloid dendritic cells, and fibroblasts (Figure 11A). In addition, the relationship between the risk model and these genes, such as PDCD1, CTLA4, POLE2, FEN1, MCM6, POLD3, MSH6, and MSH2, were negatively correlated (Figure 11B).
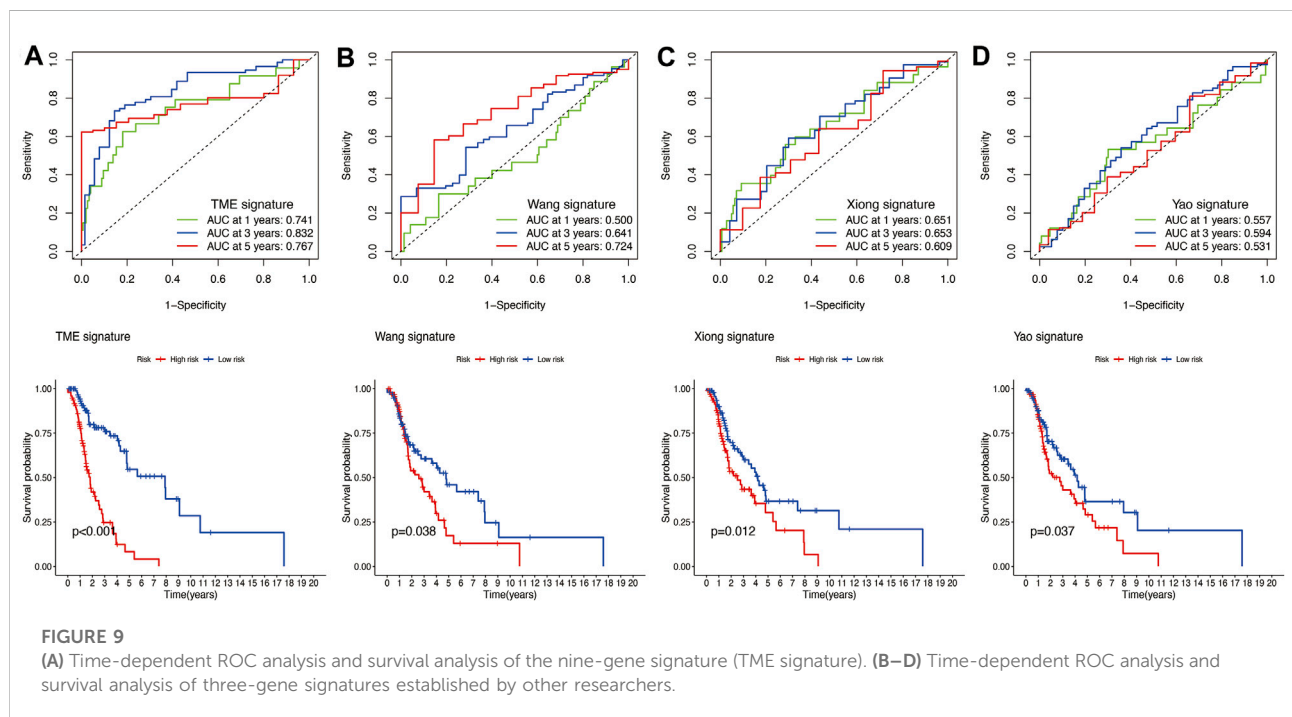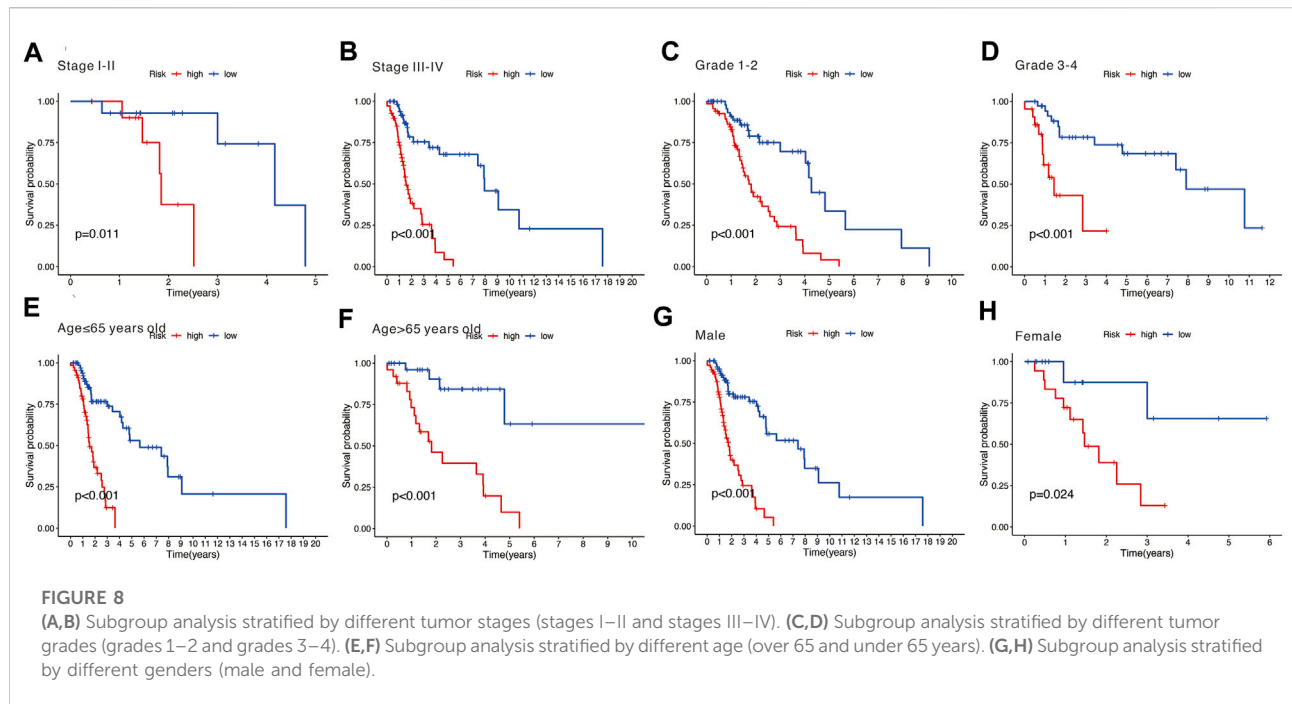
Discussion

HNSCC is the sixth most common cancer globally, with 890,000 new cases and 450,000 deaths recorded in 2018 (Bray et al., 2018; Ferlay et al., 2019). HNSCC is characterized by early metastasis, especially lymphatic metastasis, which results in poor prognosis (Evans et al., 2019). Several studies have investigated the ability of various factors, such as TNM staging, age, gender, grade, P53 mutations, and HPV negativity, to predict HNSCC prognosis. Clinically, HNSCC is difficult to treat because of its high heterogeneity. Thus, reliable prognostic tools are urgently needed to promote accurate prediction of HNSCC outcomes. Recent studies have indicated that signatures based on aberrantly expressed genes can predict cancer prognosis (Cohen et al., 2019).

In the present study, we developed a nine-gene signature consisting of GTSE1, LRRN4CL, CRYAB, SHOX2, ASNS, KRT23, ANGPT2, HOXA9, and CARD11 for predicting HNSCC prognosis. Internal validation based on TCGA database and external validation based on the GEO database
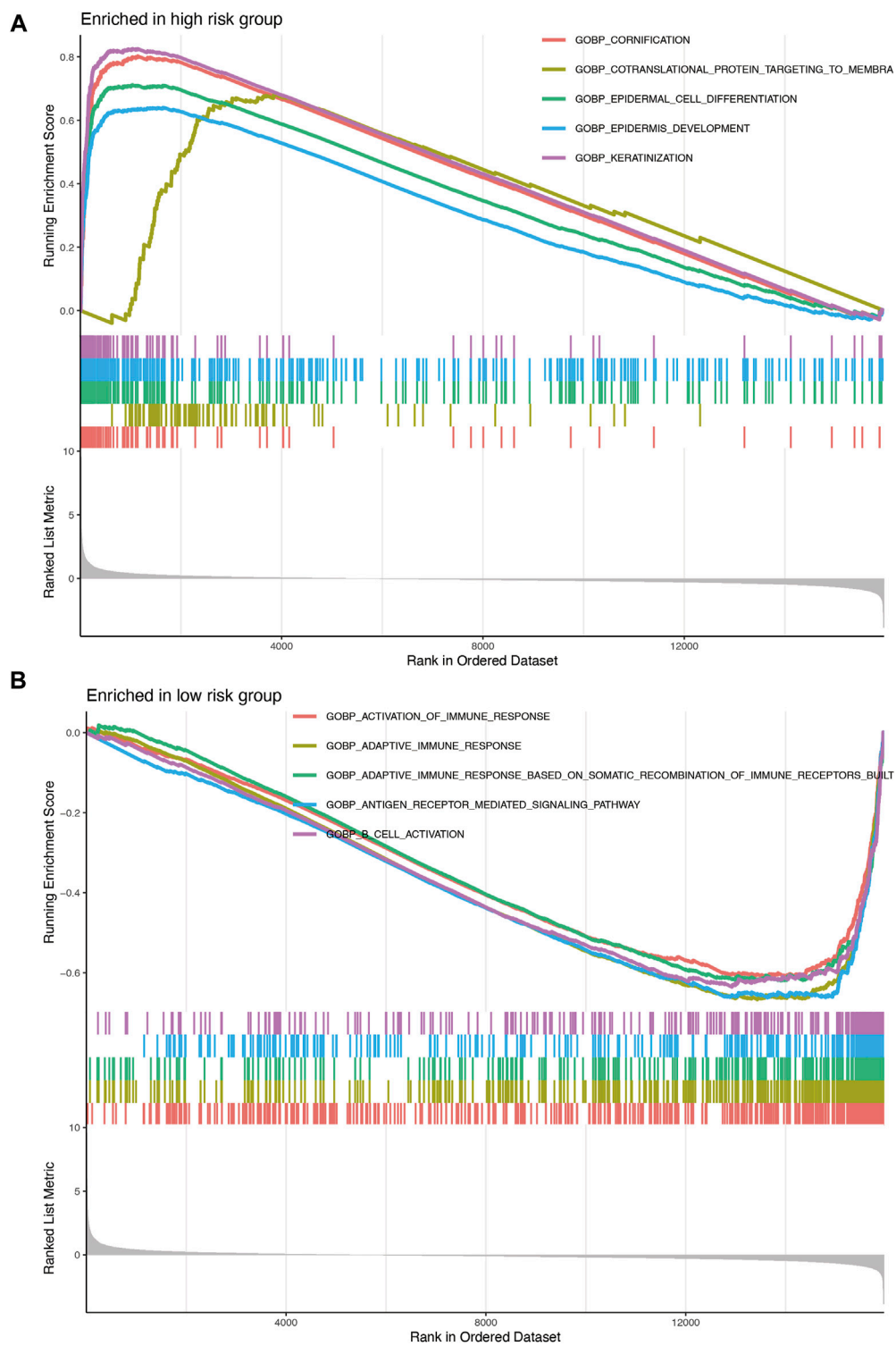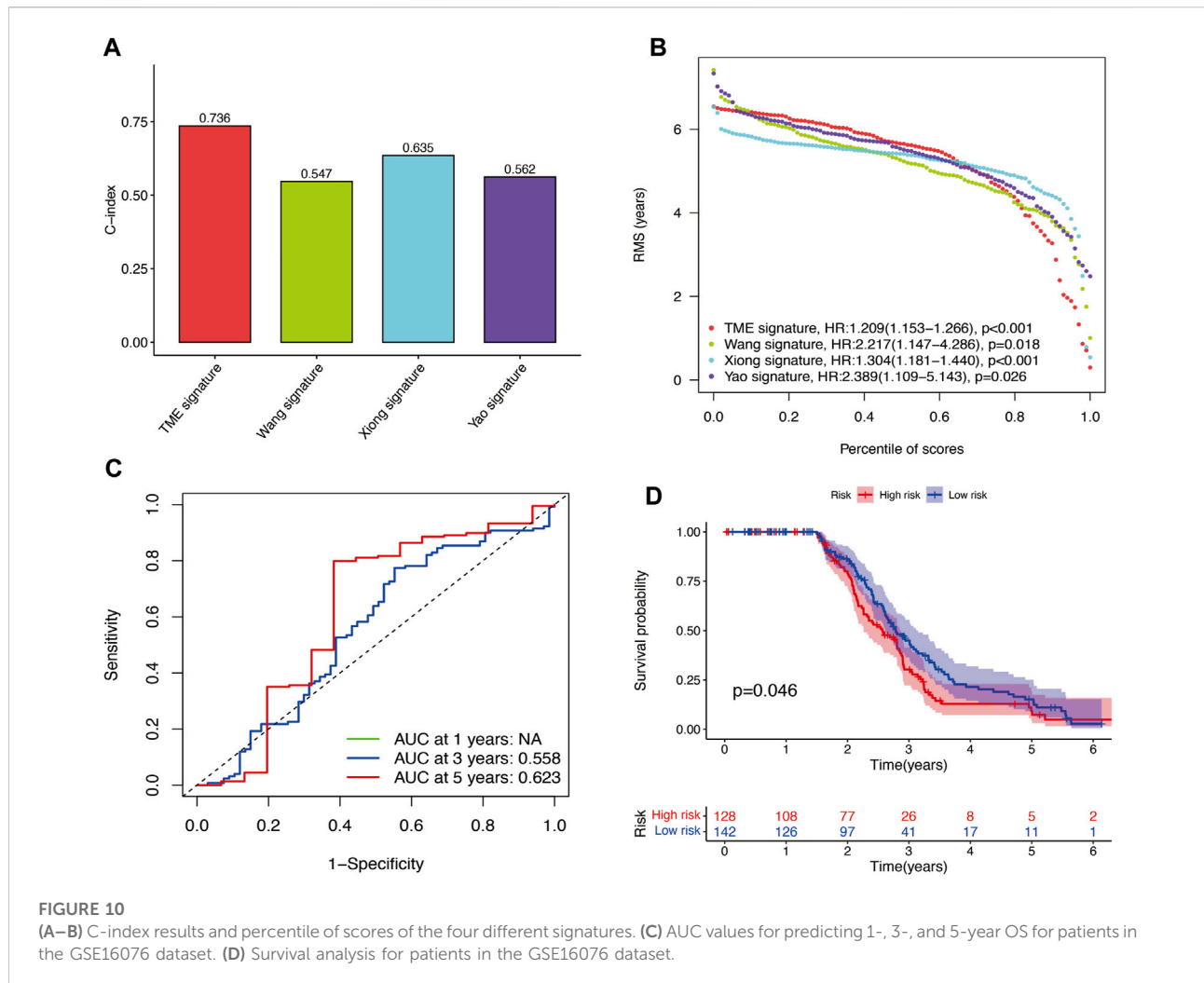
FIGURE 7
(A) Enrichment plots showing epidermis development (blue), epidermal cell differentiation (green), cornification (red), and keratinization (purple) gene enrichment in the high-risk group. **(B)** Enrichment plots showing the activation of immune response (red), adaptive immune response (orange), adaptive immune response based on somatic recombination of immune receptors built (green), B-cell activation (purple), and antigen receptor-mediated signaling pathway (blue) enrichment in the low-risk group.

FIGURE 8

(A,B) Subgroup analysis stratified by different tumor stages (stages I–II and stages III–IV). **(C,D)** Subgroup analysis stratified by different tumor grades (grades 1–2 and grades 3–4). **(E,F)** Subgroup analysis stratified by different age (over 65 and under 65 years). **(G,H)** Subgroup analysis stratified by different genders (male and female).

FIGURE 9

(A) Time-dependent ROC analysis and survival analysis of the nine-gene signature (TME signature). **(B–D)** Time-dependent ROC analysis and survival analysis of three-gene signatures established by other researchers.

were conducted and showed that the prognostic model is robust and reliable.

Through the tumor-suppressive protein p53, GTSE1 has been implicated in the pathogenesis of several malignant tumors. Previous research has proved its negative regulation in malignant tumors (Lai et al., 2021). Moreover, GTSE1 has been linked to the development of chemoresistance in osteosarcoma (Xie et al., 2021) and hepatocellular carcinoma (Li et al., 2021). To date, the role of GTSE1 in HNSCC has not been fully clarified. LRRN4CL is a cytotoxicity-associated gene associated with CD8[+] T-cell infiltration.

FIGURE 10
(A–B) C-index results and percentile of scores of the four different signatures. **(C)** AUC values for predicting 1-, 3-, and 5-year OS for patients in the GSE16076 dataset. **(D)** Survival analysis for patients in the GSE16076 dataset.

Upregulation of LRRN4CL expression on cell surfaces correlates with increased pulmonary metastases in mice (van der Weyden et al., 2021). However, the mechanism of LRRN4CL upregulation in HNSCC was not clear. CRYAB, a member of the small heat shock protein family, can regulate several cellular processes, including apoptosis, inflammation, and oxidative stress (Zhang et al., 2019). Under hypoxic conditions, CRYAB upregulation was shown to improve the survival of HNSCC cells (van de Schootbrugge et al., 2014). The relationship between CRYAB expression and HNSCC survival outcome is indistinct (Boslooper et al., 2008). Methylated SHOX2 in circulating cell-free DNA correlates with the tumor stage and prognosis (Franzen et al., 2020). There is evidence that SHOX2 can predict the occurrence and prognosis of HNSCC (Bergheim et al., 2018). It has been reported that ASNS transforms aspartate and glutamine to asparagine and glutamate in an ATP-dependent manner, respectively. The activity of human ASNS is influenced by cellular stress (Lomelino et al., 2017). The high expression of ASNS promotes growth, metastasis, and chemoresistance in neoplastic cells. In contrast, the low expression of ASNS impaired metabolic function in specified cancer models (Chiu et al., 2019). ASNS is also an important predictor of HNSCC prognosis (Mai et al., 2021). Loss of KRT23 affects the cell cycle, DNA replication, recombination, and repair (Birkenkamp-Demtröder et al., 2013). Evidence from previous studies has indicated that KRT23 influences the proliferation, migration, and prognosis of cancer (Odena et al., 2016; Zhang et al., 2017; Chen et al., 2021). ANGPT2 is a ligand for TIE1–TIE2 signaling involved in the development and maintenance of blood and lymphatic vessels (Smeland et al., 2021). In this study, we found that ANGPT2 regulates physiological and pathological angiogenesis in HNSCC by modulating vasodilation, microvascular permeability, and vasoconstriction (Xu et al., 2019). ANGPT2-deficient mice show abnormalities in blood and lymphatic vasculature and deficits leukocyte mobilization to inflammation sites (Fiedler et al., 2006). ANGPT2 is associated with shorter OS in HNSCC (Arends et al., 2021). HOXA9, a homeotic transcription factor (Lambert et al., 2019), is upregulated in HNSCC tissues and cells. HOXA9 knockdown inhibits cell proliferation, migration,
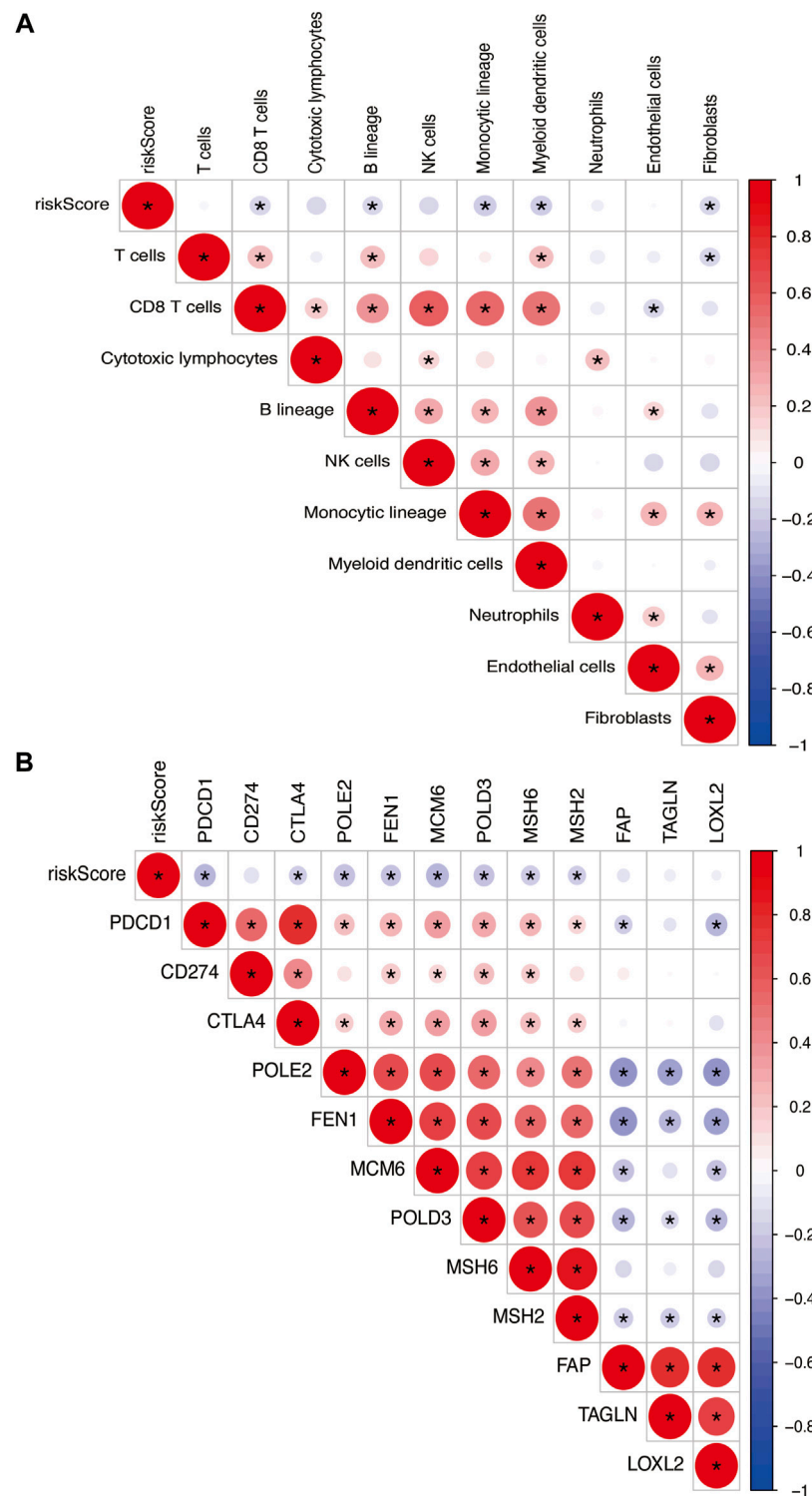
FIGURE 11
(A) Heatmap showing the correlation between risk score and immune cells. **(B)** Spearman correlation analysis of TME-related genes in PTC. The number on the right vertical axis represents the Spearman correlation coefficient between two genes. The black asterisk (*) inside the circle indicates $p < 0.05$. Red indicates positive correlation. Blue indicates negative correlation. The stronger the correlation, the larger the circle and the deeper the color.

invasion, and chemoresistance (Zhou et al., 2019; Sun et al., 2020). CARD11, a key member of the protein arginine methyltransferase (PRMT) family, encodes an adaptor protein that expresses dominant-negative (Guo et al., 2020). CARD11 gain-of-function variants may cause various immunodeficiencies (Meitlis et al., 2020). Mounting evidence indicates that CARM1 is associated with carcinoma metastasis (Sharma et al., 2017; Cai et al., 2019; Hu et al., 2020). CARM1-silencing in oral squamous cell carcinoma cells effectively suppresses tumor invasion (Lyu et al., 2022).

To our knowledge, the use of nine-gene signature for predicting the prognosis of HNSCC has not been reported before. The risk score for each of the nine prognostic genes was based on the gene expression profile but not somatic mutations or methylation status. This signature can be applied in many clinical centers because it is affordable and eliminates the requirement for whole-genome sequencing. Although our results relied on an open-access online TCGA database, without further validation of conventional prediction methods, we chose the GSM16076 dataset to validate our results. The results of the time-independent ROC analysis and survival analysis indicated that the nine-gene signature is reliable and stable. Moreover, the nomogram showed better performance than conventional clinical parameters, especially in predicting short-term (1-year or 3-year) survival. Thus, our signature can be used to guide clinical decisions regarding HNSCC treatment.

Conclusion

In conclusion, we established and verified a TME-related nine-gene signature and nomogram, which might have the potential to act as new therapeutic targets for HNSCC patients.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials; further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

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

Publisher's note

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

References

Arends, R., Guo, X., Baverel, P. G., Gonzalez-Garcia, I., Xie, J., Morsli, N., et al. (2021). Association of circulating protein biomarkers with clinical outcomes of durvalumab in head and neck squamous cell carcinoma. *Oncoimmunology* 10 (1), 1898104. doi:10.1080/2162402X.2021.1898104

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Erratum to: Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17 (1), 249. doi:10.1186/s13059-016-1113-y

Bergheim, J., Semaan, A., Gevensleben, H., Groening, S., Knoblich, A., Dietrich, J., et al. (2018). Potential of quantitative SEPT9 and SHOX2 methylation in plasmatic circulating cell-free DNA as auxiliary staging parameter in colorectal cancer: A prospective observational cohort study. *Br. J. Cancer* 118 (9), 1217–1228. doi:10.1038/s41416-018-0035-8

Birkenkamp-Demtröder, K., Hahn, S. A., Mansilla, F., Thorsen, K., Maghnouj, A., Christensen, R., et al. (2013). Keratin23 (KRT23) knockdown decreases proliferation and affects the DNA damage response of colon cancer cells. *PLoS One* 8 (9), e73593. doi:10.1371/journal.pone.0073593

Boslooper, K., Lam, A. K. Y., Gao, J., Weinstein, S., and Johnson, N. (2008). The clinicopathological roles of alpha-B-crystallin and p53 expression in patients with head and neck squamous cell carcinoma. *Pathology* 40 (5), 500–504. doi:10.1080/00313020802198010

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* 68 (6), 394–424. doi:10.3322/caac.21492

Cai, X. C., Zhang, T., Kim, E-J., Jiang, M., Wang, K., Wang, J., et al. (2019). A chemical probe of CARM1 alters epigenetic plasticity against breast cancer cell invasion. *Elife* 8, e47110. doi:10.7554/eLife.47110

Chen, H., Li, L., Qin, P., Xiong, H., Chen, R., Zhang, M., et al. (2021). A 4-gene signature predicts prognosis of uterine serous carcinoma. *BMC Cancer* 21 (1), 154. doi:10.1186/s12885-021-07834-4

Chiu, M., Taurino, G., Bianchi, M. G., Kilberg, M. S., and Bussolati, O. (2019). Asparagine synthetase in cancer: Beyond acute lymphoblastic leukemia. *Front. Oncol.* 9, 1480. doi:10.3389/fonc.2019.01480

Cohen, E. E. W., Bell, R. B., Bifulco, C. B., Burtness, B., Gillison, M. L., Harrington, K. J., et al. (2019). The Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of squamous cell carcinoma of the head and neck (HNSCC). *J. Immunother. Cancer* 7 (1), 184. doi:10.1186/s40425-019-0662-5

Evans, M., Baddour, H. M., Jr, Magliocca, K. R., Muller, S., Nannapaneni, S., Chen, A. Y., et al. (2019). Prognostic implications of peritumoral vasculature in head and neck cancer. *Cancer Med.* 8 (1), 147–154. doi:10.1002/cam4.1910

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Pineros, M., et al. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer* 144 (8), 1941–1953. doi:10.1002/ijc.31937

Fiedler, U., Reiss, Y., Scharpfenecker, M., Grunow, V., Koidl, S., Thurston, G., et al. (2006). Angiopoietin-2 sensitizes endothelial cells to TNF-alpha and has a crucial role in the induction of inflammation. *Nat. Med.* 12 (2), 235–239. doi:10.1038/nm1351

Franzen, A., Bootz, F., and Dietrich, D. (2020). Prognostic and predictive methylation biomarkers in HNSCC : Epigenomic diagnostics for head and neck squamous cell carcinoma (HNSCC). *Hno* 68 (12), 911–915. doi:10.1007/s00106-020-00902-4

Ge, S. G., Xia, J., Sha, W., and Zheng, C. H. (2017). Cancer subtype discovery based on integrative model of multigenomic data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 14 (5), 1115–1121. doi:10.1109/TCBB.2016.2621769

González-Arriagada, W. A., Lozano-Burgos, C., Zuniga-Moreta, R., Gonzalez-Diaz, P., and Coletta, R. D. (2018). Clinicopathological significance of chemokine receptor (CCR1, CCR3, CCR4, CCR5, CCR7 and CXCR4) expression in head and neck squamous cell carcinomas. *J. Oral Pathol. Med.* 47 (8), 755–763. doi:10.1111/jop.12736

Guo, N., Nulahou, A., Bu, Q., Liu, M., Wang, Y., Zhao, Y., et al. (2020). MiR-542-5p regulates the progression of diabetic retinopathy by targeting CARM1. *Acta Biochim. Pol.* 67 (3), 373–378. doi:10.18388/abp.2020_5228

Hu, B., Li, X., Chen, L., and Liu, Z. (2020). High expression of CARM1 inhibits lung cancer progression by targeting TP53 by regulating CTNNB1. *Lung* 198 (2), 415–422. doi:10.1007/s00408-020-00324-7

Huo, M., Zhang, Y., Chen, Z., Zhang, S., Bao, Y., and Li, T. (2020). Tumor microenvironment characterization in head and neck cancer identifies prognostic and immunotherapeutically relevant gene signatures. *Sci. Rep.* 10 (1), 11163. doi:10.1038/s41598-020-68074-3

Iasonos, A., Schrag, D., Raj, G. V., and Panageas, K. S. (2008). How to build and interpret a nomogram for cancer prognosis. *J. Clin. Oncol.* 26 (8), 1364–1370. doi:10.1200/JCO.2007.12.9791

Kim, J., and Bae, J. S. (2016). Tumor-associated macrophages and neutrophils in tumor microenvironment. *Mediat. Inflamm.* 2016, 6058147. doi:10.1155/2016/6058147

Lai, W., Zhu, W., Li, X., Han, Y., Wang, Y., Leng, Q., et al. (2021). GTSE1 promotes prostate cancer cell proliferation via the SP1/FOXM1 signaling pathway. *Lab. Invest.* 101 (5), 554–563. doi:10.1038/s41374-020-00510-4

Lambert, M., Alioui, M., Jambon, S., Depauw, S., Van Seuningen, I., and David-Cordonnier, M. H. (2019). Direct and indirect targeting of HOXA9 transcription factor in acute myeloid leukemia. *Cancers (Basel)* 11 (6), E837. doi:10.3390/cancers11060837

Li, S. S., Chen, D. M., Chen, L. B., Wei, H., Wang, J. L., Xiao, J., et al. (2021). GTSE1 promotes SNAIL1 degradation by facilitating its nuclear export in hepatocellular carcin oma cells. *Mol. Med. Rep.* 23 (6), 454. doi:10.3892/mmr.2021.12093

Liu, J. X., Wang, D., Gao, Y. L., Zheng, C. H., Xu, Y., and Yu, J. (2018). Regularized non-negative matrix factorization for identifying differentially expressed genes and clustering samples: A survey. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 15 (3), 974–987. doi:10.1109/TCBB.2017.2665557

Lomelino, C. L., Andring, J. T., McKenna, R., and Kilberg, M. S. (2017). Asparagine synthetase: Function, structure, and role in disease. *J. Biol. Chem.* 292 (49), 19952–19958. doi:10.1074/jbc.R117.819060

Lyu, X. Y., Shui, Y. S., Wang, L., Jiang, Q. S., Meng, L. X., Zhan, H. Y., et al. (2022). WDR5 promotes the tumorigenesis of oral squamous cell carcinoma via CARM1/β-catenin axis. *Odontology* 110 (1), 138–147. doi:10.1007/s10266-021-00649-6

Mai, Z., Chen, H., Huang, M., Zhao, X., and Cui, L. (2021). A robust metabolic enzyme-based prognostic signature for head and neck squamous cell carcinoma. *Front. Oncol.* 11, 770241. doi:10.3389/fonc.2021.770241

Meitlis, I., Allenspach, E. J., Bauman, B. M., Phan, I. Q., Dabbah, G., Schmitt, E. G., et al. (2020). Multiplexed functional assessment of genetic variants in CARD11. *Am. J. Hum. Genet.* 107 (6), 1029–1043. doi:10.1016/j.ajhg.2020.10.015

Odena, G., Chen, J., Lozano, J. J., Altamirano, J., Rodrigo-Torres, D., Affo, S., et al. (2016). LPS-TLR4 pathway mediates ductular cell expansion in alcoholic hepatitis. *Sci. Rep.* 6, 35610. doi:10.1038/srep35610

Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., et al. (2017). Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 171 (7), 1611–1624.e24. doi:10.1016/j.cell.2017.10.044

Sharma, P., Bhattacharyya, D. K., and Kalita, J. (2017). Disease biomarker identification from gene network modules for metastasized breast cancer. *Sci. Rep.* 7 (1), 1072. doi:10.1038/s41598-017-00996-x

Shield, K. D., Ferlay, J., Jemal, A., Sankaranarayanan, R., Chaturvedi, A. K., Bray, F., et al. (2017). The global incidence of lip, oral cavity, and pharyngeal cancers by subsite in 2012. *Ca. Cancer J. Clin.* 67 (1), 51–64. doi:10.3322/caac.21384

Smeland, M. F., Brouillard, P., Prescott, T., Boon, L. M., Hvingel, B., Nordbakken, C. V., et al. (2021). Biallelic ANGPT2 loss-of-function causes severe early-onset non-immune hydrops fetalis. *J. Med. Genet.* 2021, 108179. doi:10.1136/jmedgenet-2021-108179

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Sun, Q., Zhang, S. Y., Zhao, J. F., Han, X. G., Wang, H. B., and Sun, M. L. (2020). HIF-1α or HOTTIP/CTCF promotes head and neck squamous cell carcinoma progression and drug resistance by targeting HOXA9. *Mol. Ther. Nucleic Acids* 20, 164–175. doi:10.1016/j.omtn.2019.12.045

van de Schootbrugge, C., Schults, E. M. J., Bussink, J., Span, P. N., Grenman, R., Pruijn, G. J. M., et al. (2014). Effect of hypoxia on the expression of αB-crystallin in head and neck squamous cell carcinoma. *BMC Cancer* 14, 252. doi:10.1186/1471-2407-14-252

van der Weyden, L., Harle, V., Turner, G., Offord, V., Iyer, V., Droop, A., et al. (2021). CRISPR activation screen in mice identifies novel membrane proteins enhancing pulmonary metastatic colonisation. *Commun. Biol.* 4 (1), 395. doi:10.1038/s42003-021-01912-w

Vickers, A. J., and Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Med. Decis. Mak.* 26 (6), 565–574. doi:10.1177/0272989X06295361

Wang, J., Zhu, K., Ren, L., Li, H., Lin, S., Qing, X., et al. (2020). Quality of life and related risk factors after breast reconstruction in breast cancer patients. *Gland. Surg.* 12 (1), 767–774. doi:10.21037/gs-20-532

Xie, C., Xiang, W., Shen, H., and Shen, J. (2021). GTSE1 is possibly involved in the DNA damage repair and cisplatin resistance in osteosarcoma. *J. Orthop. Surg. Res.* 16 (1), 713. doi:10.1186/s13018-021-02859-8

Xiong, Y., Si, Y., Feng, Y., Zhuo, S., Cui, B., and Zhang, Z. (2021). Prognostic value of lipid metabolism-related genes in head and neck squamous cell carcinoma. *Immun. Inflamm. Dis.* 9 (1), 196–209. doi:10.1002/iid3.379

Xu, H., Chen, X., and Huang, Z. (2019). Identification of ESM1 overexpressed in head and neck squamous cell carcinoma. *Cancer Cell Int.* 19, 118. doi:10.1186/s12935-019-0833-y

Yang, B., Chen, Z., Huang, Y., Han, G., and Li, W. (2017). Identification of potential biomarkers and analysis of prognostic values in head and neck squamous cell carcinoma by bioinformatics analysis. *Onco. Targets. Ther.* 10, 2315–2321. doi:10.2147/OTT.S135514

Yao, Y., Yan, Z., Lian, S., Wei, L., Zhou, C., Feng, D., et al. (2020). Prognostic value of novel immune-related genomic biomarkers identified in head and neck squamous cell carcinoma. *J. Immunother. Cancer* 8 (2), e000444. doi:10.1136/jitc-2019-000444

Zhang, J., Liu, J., Wu, J., Li, W., Chen, Z., and Yang, L. (2019). Progression of the role of CRYAB in signaling pathways and cancers. *Onco. Targets. Ther.* 12, 4129–4139. doi:10.2147/OTT.S201799

Zhang, N., Zhang, R., Zou, K., Yu, W., Guo, W., Gao, Y., et al. (2017). Keratin 23 promotes telomerase reverse transcriptase expression and human colorectal cancer growth. *Cell Death Dis.* 8 (7), e2961. doi:10.1038/cddis.2017.339

Zhang, X., Shi, M., Chen, T., and Zhang, B. (2020). Characterization of the immune cell infiltration landscape in head and neck squamous cell carcinoma to aid immunotherapy. *Mol. Ther. Nucleic Acids* 22, 298–309. doi:10.1016/j.omtn.2020.08.030

Zhou, C., Li, J., Li, Q., Liu, H., Ye, D., Wu, Z., et al. (2019). The clinical significance of HOXA9 promoter hypermethylation in head and neck squamous cell carcinoma. *J. Clin. Lab. Anal.* 33 (5), e22873. doi:10.1002/jcla.22873