Construction of a Ferroptosis-Related Gene Signature for Head and Neck Squamous Cell Carcinoma Prognosis Prediction

Qun Li, Yangli Jin, Zhisen Shen, Huigao Liu, Yi Shen, Zhenhua Wu

Department of Otorhinolaryngology Head and Neck Surgery, Ningbo Medical Center Lihuili Hospital, Ningbo, Zhejiang, People’s Republic of China; Department of Otorhinolaryngology Head and Neck Surgery, Ningbo Zhenhai Longsai Hospital, Ningbo, Zhejiang, People’s Republic of China

Background: Head and neck squamous cell carcinoma (HNSCC) is one of the most common malignant cancers, and few studies have demonstrated the value of ferroptosis-related genes in prognosis.

Methods: The original counts of RNA sequencing data and clinicopathological data were obtained from TCGA and GSE65858 datasets. Common ferroptosis-related genes related to prognosis were identified from the training set and were included in LASSO to determine the best prognosis. To evaluate the efficacy, time-dependent ROC and Kaplan–Meier (KM) survival analyses were applied. Moreover, univariate and multivariate Cox regression analyses were used to screen independent parameters of prognosis and build a nomogram. Eventually, possible biological pathways were proposed based on GSEA.

Results: Among 242 ferroptosis-related genes, we identified that the FLT3, IL6, Keap1, NQO1, SOCS1 and TRIB3 genes were significantly connected with HNSCC patient prognosis as a six-gene signature. After, the patients were divided into high- and low-risk groups based on the six-gene signature. The KM survival curves demonstrated that the high-risk group had worse OS (p < 0.0001) and higher AUC values (0.654, 0.735, and 0.679 for 1-, 3-, and 5-year survival, respectively) for the prognostic signature of the six genes compared with other genes, which were also validated in the GSE65858 dataset. Moreover, GSEA suggested that the epithelial mesenchymal transition pathway was abundant and that the mesenchymal status in the high-risk group was substantially higher than that in the low-risk group. Finally, the immune microenvironment and differences in the content of immune cell types were demonstrated.

Conclusion: We established a six-ferroptosis-related-gene model crossing clinical prognostic parameters that can predict HNSCC patient prognosis and provide a reliable prognostic evaluation tool to assist clinical treatment decisions.

Keywords: head and neck squamous cell carcinoma, ferroptosis, prognosis, TCGA

Introduction

Head and neck squamous cell carcinoma (HNSCC) is a common, pathological, and fatal malignancy, with approximately 380,000 patients dying from the disease each year. Epidemiological investigations have indicated that the most common pathogenic factors of HNSCC include smoking, drinking and human papillomavirus (HPV) infection. Although some progress has been made in HNSCC treatment, approximately 50% of patients will die. Great progress has been made in multidisciplinary therapy, covering monoclonal antibody therapy, adoptive T cell transplantation, and cytokine therapy; however, for decades, there has been a stable
overall survival rate. The high heterogeneity of potential risk factors, different tumor locations, and molecular abnormalities make prognosis prediction challenging. Therefore, novel prognostic models are urgently needed.

Successful avoidance of programmed cell death is a major feature of cancer. As a newly regulated pattern of cell death, ferroptosis depends on iron and reactive oxygen species. Molecular features different from other forms of regulatory cell death have been found in ferroptosis. Significantly, previous studies have demonstrated that tumor cells escape other forms of cell death, but their sensitivity to ferroptosis remains or is acquired. Therefore, studies of ferroptosis in cancer have provided approaches for cancer therapy. In addition to inducible factors of ferroptosis, many genes have been identified as regulatory factors or markers of ferroptosis. The inference that ferroptosis plays an important role in HNSCC has been verified in previous studies. Additionally, distinct ferroptosis regulation patterns have provided a new mechanistic view and insight for clinical therapy and prognosis in oral and laryngeal squamous cell carcinoma. Therefore, illustrating the value of ferroptosis-related genes in the prognosis of HNSCC is important.

In our study, to predict HNSCC prognosis, we built a gene-based model that integrated the clinical parameters of prognosis. Clinopathological and RNA-seq data were obtained from TCGA. Six genes (FLT3, IL6, Keap1, NQO1, SOCS1 and TRIB3) were screened out and found to be significantly related to HNSCC patient prognosis. Based on Cox regression analysis, we used independent parameters of prognosis, including risk score and M stage, to establish a nomogram. Compared to the risk score and M stage in the validation sets, the time-related receiver operating characteristic (ROC) curve showed that the nomogram was the best predictor of patient survival rate. Good consistency between the predicted and observed results of the nomogram was displayed in the calibration plot. Eventually, functional enrichment analysis of the six genes was performed, and possible biological pathways were proposed.

Materials and Methods
Dataset Sources and Processing
As of February 8, 2021, the RNA sequencing (RNA-seq) data and clinical information of 495 HNSCC patients were downloaded from TCGA (https://portal.gdc.cancer.gov/repository). To standardize the gene expression profile, the scaling method provided in the “limma” R software package was used. TCGA data are public; therefore, this study did not need to be approved by the local ethics committees. The research strictly followed TCGA data access strategy and release guidelines. Another website (http://www.zhounan.org/ferrdb/) was the source of 242 ferroptosis-related genes and their expression data.

Selection and Verification of Prognosis-Related Genes
The transcripts per million (TPM) method was used to standardize the original counts of RNA-seq data, and the subsequent survival analysis employed a log 2-based transformation (log2tpm).

After, based on the standardized expression data in the training set (n = 348), a set of polygenic signatures was built to predict HNSCC patient prognosis. In the univariate Cox proportional hazard regression analysis, the expression data of log2 TPM transformation and the relevant clinical data were applied to screen prognostic genes (hazard ratio (HR) ≠ 1, P < 0.05). A group of candidate genes was obtained by crossing the prognosis-related genes with ferroptosis-related genes. By applying the “glmnet” software package of R software and the least absolute shrinkage and selection operator (LASSO) regression method, a polygenic signature of OCGS was constructed to predict HNSCC prognosis.

Establishment and Estimation of a Multigene Prognostic Signature
A multivariate Cox proportional hazards regression model was used to calculate the regression coefficients of the six genes most related to prognosis. After, the expression level and coefficient of each gene were combined using the linear combination method to obtain the risk score formula as follows: Risk score = \sum_{i=1}^{6} (Expi \times Coei). Expi represents each prognostic gene expression level, N represents the number of genes, and Coei represents the regression coefficient. Based on the median risk score, the patients were partitioned into high- and low-risk groups in the training set. For KM survival analysis, a Log rank test was used to determine the difference in survival rate between the high-risk group and low-risk group for comparison. Comparison of the predictive ability of the risk score and a few clinical parameters (including age, sex, stage, smoking, T stage, N stage, M stage, grade, HPV and risk score) was based on univariate Cox proportional hazards regression.
analysis. The R package “timeROC” was used to draw the time-dependent ROC curve to evaluate the predictive accuracy of the clinical parameters. Additionally, multivariate Cox proportional hazard regression analysis was applied to identify whether the risk score can be an independent factor for HNSCC patient prognosis. The univariate Cox proportional hazards regression analysis included other clinical parameters with statistically significant differences (p < 0.05).

Validation of Multigene Prognostic Signature
To test the predictive ability and applicability of the polygenic prognostic signature in HNSCC, the testing set (n = 147), whole set (n = 495), and external validation set (n = 270) were used for internal and external validation. The calculation of the risk score for each patient was based on the coefficients of the above six genes in the validation set. After, consistent with the median risk score in the training set, the patients were partitioned into high- and low-risk groups. The characteristics of multiple genes related to prognosis were assessed using a Log rank test and ROC analysis.

Construction and Validation of the Prognostic Gene Nomogram
Based on the independent parameters of prognosis screened out by univariate and multivariate Cox proportional hazards regression analysis and time-dependent ROC curve analysis, two nomograms were constructed with the “rms” package of R software. The prediction capabilities of the two nomograms were evaluated through calibration curves and time-dependent ROC curves.

Functional Enrichment Analysis of Genes
To determine potential biological pathways, the “ClusterProfiler” R package was used for gene set enrichment analysis (GSEA). According to the median expression of prognosis-related genes, 495 HNSCC samples in the entire set were partitioned into two groups. The hallmark gene sets (h.all.v7.1.symbols.gmt) were downloaded from the GSEA website (https://www.gseasigdb.org/gsea/index.jsp).

Computation of EMT Score
To compute the EMT score of each sample, we adopted an approach similar to that used in single-sample GSEA (ssGSEA). Published EMT gene signatures were analyzed using this method, yielding a single “EMT score” per signature for each sample.

Estimation of Immune Cell Type Content
To visualize the immune cell type content of each patient, the “CIBERSORT” R package was used to convert mRNA data into the infiltration fractions of nontumor cells in the tumor microenvironment. For each sample, the sum of all estimated immune cell type scores is equal to 1.

Statistical Analysis
Using the “sample” function of R software, tumor tissue samples were randomly partitioned into two groups. For Kaplan–Meier curves, a Log rank test and univariate Cox proportional hazard regression were used to determine the risk ratio (hazard ratio, HR) of the P value and the 95% confidence interval (CI). R software version 3.6.1 (The R Foundation for Statistical Computing, 2019) and R software packages were used for all the above analyses. All statistical tests are bidirectional without exception. When P < 0.05, the result was considered statistically significant.

Results
Identification of Prognostic Ferroptosis-Related DEGs in the TCGA Cohort
Figure 1 shows the study flowchart. In total, 495 HNSCC patients were enrolled in the TCGA-HNSCC cohort. Through univariate Cox regression analysis, we analyzed 242 ferroptosis-related genes and retained 10 prognostic genes related to ferroptosis (Figure 2A). To further identify the genes significantly connected with HNSCC patient prognosis, tenfold cross-validation of LASSO regression was used to obtain the best lambda value from the least partial likelihood deviation. The lambda value was associated with seven ferroptosis-related genes that were significantly correlated with OS (Figure 2B and C).

Establishment and Estimation of a Six-Gene Prognostic Signature
To determine whether each gene has prognostic significance in patients with HNSCC, multivariate Cox proportional hazard regression analysis was conducted using the six prognostic genes (Figure 2D). Then, consistent with their Cox coefficients, we constructed six gene-based risk scores: risk score = - 0.0966956 * exp (FLT3) + 0.0873983


\[ \text{exp}(\text{IL6}) - 0.4095449 \times \text{exp}(\text{Keap1}) + 0.1254306 \times \text{exp}(\text{NQO1}) - 0.1389582 \times \text{exp}(\text{SOCS1}) + 0.1288005 \times \text{exp}(\text{TRIB3}). \]

After, we calculated each patient’s risk score, where we used the “survminer” R software package to obtain the median cutoff point and partitioned the patients into low- and high-risk groups (Figure 3A). In the training group, the survival status of all patients is shown in Figure 3B, and a heatmap of the six prognostic genes is shown in Figure 3C. The KM survival curves demonstrated that the high-risk group had worse OS than the low-risk group (Figure 3D). In addition, in the time-dependent ROC analysis, higher AUC values were observed for the prognostic signature of the six genes compared with other genes (Figure 3E). The above results indicate that the polygenic model has better predictive ability for 1-, 3-, and 5-year OS.

**Internal and External Validation of the Six-Gene Prognostic Signature**

To validate the significant prognostic value of the six genes, the internal validation set (n = 147), the entire validation set (n = 495) and GSE65858 as the external validation set (n = 270) were used to evaluate the training set results. In accordance with the outcomes of the training set, the KM curves of the three test sets showed that the low-risk group had a better patient prognosis than the high-risk group (Figure 4A–C). As shown in Figure 4D–F, the AUCs for the internal validation set, overall validation set and external validation set for 1-, 3-, and 5-year OS were 0.64, 0.68, and 0.63; 0.65, 0.71, and 0.66; and 0.72, 0.63, and 0.57 in the time-dependent ROC analysis, respectively. In conclusion, in predicting OS in HNSCC patients, the prognostic signature of the six genes was good.

**Construction and Validation of the Gene-Based Nomogram**

After univariate and multivariate Cox proportional hazard regression analysis, the six-gene prognostic signature and M stage could be regarded as independent variables related to the prognosis of OS in the training set (Figure 5A and B). Moreover, the time-dependent ROC curve analysis indicated that risk score and age were better predictive factors than others (Figure 5C). To build a more reliable clinical prediction method, two composite nomograms using the risk score and M stage or the risk score and RNA expression data and clinical data from TCGA

![Figure 1 Study flowchart.](https://doi.org/10.2147/IJGM.S343233)
The survival prediction in the calibration plot showed that the prediction results of the six-gene prognostic nomogram were consistent with the actual results, and the composite nomogram from risk score and M stage was closer to the actual result (Figure 5E–G). Moreover, time-dependent ROC curve analysis indicated that the composite nomogram based on the risk score and M stage had a high predictive ability (Figure 5H), and the nomogram based on the risk score and M stage is shown in Figure 5I. Furthermore, predicting 1-, 3-, and 5-year OS might be the best function for the six-gene prognostic nomogram, as the AUC value for 1-, 3-, and 5-year OS in the nomogram was higher than that of the risk score (Figure 6A–J). We used the testing set (n = 107), whole set (n = 495) and GSE65858 (n = 164) to validate the above findings.

**Gene Set Enrichment Analysis**

To determine the potential biological processes of the six prognostic genes, GSEA was conducted. The results showed that all of the high-risk samples were rich in interferon alpha reaction, angiogenesis, interferon gel reaction and epithelial mesenchymal transition. Samples with high-risk genes in GSE65858 were enriched in angiogenesis, epithelial mesenchymal transition, myogenesis and TNFA signal transduction through NF-κB (Figure 7A and B). The epithelial mesenchymal transition (EMT) pathway was abundant in the two sets (Figure 7C and D).
To verify the effects of the six ferroptosis-related gene signatures on EMT, we further performed ssGSEA using gene expression signatures associated with mesenchymal status. The EMT signal indicated that the mesenchymal status in the high-risk group was substantially higher than that in the low-risk group (Figure 7E).

**Comparison of the Immune Microenvironment of Patients with HNSCC Between the High- and Low-Risk Groups**

IL6 was one of six ferroptosis-related genes. As a pleiotropic cytokine, IL6-activated signal transduction and transduction 3 (STAT3) signaling has an important impact on tumor infiltrating immune cells in the tumor immune microenvironment. Therefore, we compared the immune microenvironment of HNSCC patients in the high- and low-risk groups. Figure 8A shows that there were significant differences in the content of immune cell types. After univariate (Figure 8B) and multivariate (Figure 8C) Cox regression analyses, activated memory CD4 T cells, regulatory T cells and resting mast cells were considered to be positive independent prognostic factors.

**Discussion**

Previous studies have shown that certain genes may regulate drug-induced ferroptosis in HNSCC, but their association with OS in HNSCC patients is still unclear. In our study, the expression of 242 ferroptosis-related genes...
genes in HNSCC and their relationship with OS were systematically studied. For the first time, by integrating six ferroptosis-related genes, a new prognostic model was built and validated in an external cohort. The prognostic model proposed in this study consists of six ferroptosis-related genes (FLT3, IL6, Keap1, NQO1, SOCS1 and TRIB3). As a receptor tyrosine kinase, activation of FLT3 mediates hematopoietic progenitor cell survival, proliferation and differentiation. Valuable studies have confirmed that FLT3 inhibitors are effective protective agents that prevent lipid peroxidation from resisting glutamate toxicity, and glutamate toxicity is related to ferroptosis. As an inflammatory cytokine, IL6 is related to various biological processes, including immune disorders and cancer. IL6 exposure induces cartilage cell ferroptosis by inducing oxidative stress and interfering with iron homeostasis. KEAP1, a component of the Cullin 3-based E3 ubiquitin ligase complex, regulates the stability and accumulation of Nrf2. Activation of the p62-keap1-nrf2 pathway can prevent ferroptosis in hepatocellular carcinoma cells. NQO1, a double electron reductase, is responsible for the detoxification of quinones and the biological activity of some quinones. NQO1 is overexpressed in many tumors, closely related to various carcinogenic processes and considered a cancer-specific therapeutic target. Knockdown of NQO1 increased the inhibitory effect of ferroptosis inducers on the growth of hepatoma cells. SOCS1 is regarded as a negative feedback regulator of cytokine and growth factor receptor signaling. Loss of its expression may affect the classical checkpoint pathway in various tumors. Exogenous SOCS1 regulates the expression of p53 target genes and cell sensitivity to ferroptosis. This effect was related to the ability of SOCS1 to downregulate SLC7A11 expression and reduce glutathione levels. TRIB3, a multifunctional protein, has recently been demonstrated to play a key regulatory role in many signaling pathways and is involved in various cellular processes. As reported, TRIB3 can directly bind to AKT and block its activation by regulating the phosphorylation status at Ser473 and Thr308. The role of TRIB3 in inhibiting Akt in ferroptosis is still unclear. In addition, some antitumor drugs, such as tetrahydrocannabinol and salinomycin, can induce TRIB3 expression. In summary, the six genes in the prognostic model are related to ferroptosis and cancer to different degrees. In our study, FLT3, KEAP1, and SOCS1 were associated with a good
prognosis in HNSCC, while the remaining three genes (IL6, NQO1, and TRIB3) were related to a poor prognosis. Whether these genes are involved in HNSCC patient prognosis by affecting the process of ferroptosis remains to be clarified because there is little research on these genes.
He et al also reported the ferroptosis-related gene signature in prognosis in HNSCC during our paper submission. He et al selected differentially expressed genes based on the comparison between HNSCC and adjacent nontumorous tissues and obtained prognosis-related genes through univariate Cox regression analysis. In our study, first, in univariate Cox regression analysis, 242 ferroptosis-related genes were analyzed, and 10 predictive ferroptosis-related genes remaining in the training set were used in LASSO regression analysis. Then, the best lambda value was obtained from the least partial likelihood deviation by using the tenfold cross-validation LASSO regression method, which was related to six ferroptosis-related genes that have a significant association with OS. Moreover, there are large differences during cancer initiation and development. Some genes will have little difference in the cancer initiation stage, while they could have an important influence on cancer development. Therefore, we think that screening based on the comparison between tumors and adjacent nontumors will lose some important genes that have important functions during cancer development. Furthermore, we used univariate Cox regression analysis, LASSO regression analysis, and multivariate Cox regression analysis for the screening and identification of prognostic models, which would be more reasonable and accurate. As a result, more reliable AUC values for 1-, 3-, and 5-year survival were obtained in our data compared to He et al’s paper data (AUC-1 year: 0.740 vs 0.717, −3 years: 0.618 vs 0.574, −5 years: 0.688 vs 0.575). Of note is that one common gene, KEAP1, which mediates the proteasomal degradation of the key ferroptosis regulator Nrf2, was reported in both papers, suggesting feasibility in workflow and approaches in both paper concepts.

The poor prognosis association of IL6, NQO1, and TRIB3 was further identified based on the Human Pathology Atlas project (HPA) (www.proteinatlas.org) to
analyze protein expression with immunohistochemistry staining between normal and cancer tissues and the overall survival analysis from the Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku.cn/) (Supplemental Figure 1), which indicated the tendency of increased expression level and worse overall survival.

Moreover, our GSEA indicated that the six-gene prognostic signature corresponding to biological functions was highly related to angiogenesis, epithelial-mesenchymal transition, and cytokine-related pathways and strongly linked to poor prognosis. The quantified EMT scores indicate that the high-risk group with poor prognosis is characterized by high mesenchymal levels. In addition, the significant differences in immune cell infiltration and univariate and multivariate Cox regression analysis suggest that the poor prognosis of the high-risk group is characterized by a decrease in memory CD4+ T cells, regulatory T cells and resting mast cells.

Nevertheless, first, our prediction model was built and validated with retrospective data from a public database. Second, it is inevitable that only one marker is considered to establish a prognostic model because many significant prognostic genes in HNSCC may be excluded. Third, in the TCGA database, there were significant differences in the protein levels of the six genes between cancer and its surrounding normal tissues. Most patients in the TCGA database are white and Asian. Therefore, more public databases or experiments are needed to explore whether there is geographical significance in the expression levels of the six genes.

In addition to HNSCC, we found that the concept of selecting and validating ferroptosis-related genes or lncRNA signatures developed rapidly in various other cancers, such as hepatocellular carcinoma, bladder cancer, lung adenocarcinoma, and breast cancer during our paper submission. These papers used ferroptosis-related genes from FerrDb and developed prognosis prediction utilizing TCGA and GEO database-based univariate Cox regression analysis and LASSO regression analysis, as in our paper, further indicating the feasibility of the methodology and the importance of ferroptosis-
related genes in cancer development and prognosis. However, no common ferroptosis-related gene could predict prognosis in various cancers, indicating the cancer type-dependent role of ferroptosis. Thus, further investigation, such as increasing the number of samples, experimental validation, or mechanistic demonstration, would be helpful.

**Conclusion**

In conclusion, a new prognostic model comprising six ferroptosis-related genes that were independently related to OS in the derivation and validation cohorts was constructed, providing insights for the prediction of HNSCC patient prognosis. This model provides a suggestive reliable prognostic evaluation tool for clinicians and might need further deeper analyses on new datasets to identify the “core ferroptosis genes” strongly involved in prognosis and their different potential associations with different HNSCC subsites.

**Abbreviations**

HNSCC, Head and Neck Squamous Cell Carcinoma; TCGA, The Cancer Genome Atlas; LASSO, the Least Absolute Shrinkage and Selection Operator; GSEA, Gene Set Enrichment Analysis; GEO, Gene Expression Omnibus; KM, Kaplan–Meier; ROC, Receiver Operating Characteristic; HPV, Human Papilloma Virus.
Data Sharing Statement
The data are available from the corresponding author upon reasonable request.

Ethical Committee Approval and Informed Consent
The study was performed based on the TCGA and GEO public databases. Because no personal identifying information was used in the current study, it was granted an exemption from ethics approval from the Institutional Review Board of the Ningbo Medical Center Lihuili Hospital.

Acknowledgment
This work was supported by the Zhejiang Provincial Natural Science Foundation of China [LY19H160014; LQ21H130001], Ningbo Medical and Health Brand Discipline [No. PPXK2018-02], Medical and Health Research Project of Zhejiang Province [2019ZD018; 2021KY307], Ningbo Natural Science Foundation [20203N4239], and Ningbo “Technology Innovation 2025” Major Special Project [2020Z098; 2018B10015].

Disclosure
The authors declare no conflicts of interest.

References


