Inflammation-Related Gene Signature for Predicting the Prognosis of Head and Neck Squamous Cell Carcinoma

Yilong Lu¹, Zengrong Jia²

¹School of Stomatology, Wenzhou Medical University, Wenzhou, Zhejiang, People’s Republic of China; ²Thyroid Surgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, People’s Republic of China

Correspondence: Zengrong Jia, Thyroid Surgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, 325015, People’s Republic of China, Tel +86 135 874 22709, Fax +86 577 55578033, Email immunity2@163.com

Purpose: The inflammatory response was associated with the prognosis of head and neck squamous cell carcinoma (HNSCC). This study aimed to perform a novel prognostic signature based on inflammation-related genes (IRGs) for a better understanding of the prognosis of HNSCC.

Patients and Methods: IRGs were obtained from The Cancer Genome Atlas (TCGA) database and the Gene Expression Omnibus (GEO) database. Functional enrichment analysis was performed to explore potential pathways. Univariate and multivariate Cox regression as well as the Least Absolute Shrinkage and Selection Operator (LASSO) were utilized to construct an IRGs-based prognostic model on TCGA database and the GEO database was utilized for outcome validation. The nomogram model was constructed based on independent prognostic factors after univariate and multivariate Cox regression. The immune cell infiltration level was analyzed via the Tumor Immune Estimation Resource (TIMER) database.

Results: In this study, we confirmed that 60% IRGs were abnormally expressed in HNSCC samples, and these were associated with important oncobiology. Then, a prognostic signature comprising 7 hub genes was generated based on TCGA database. The results were validated in 97 patients from GSE41613. A nomogram comprising risk score, age, M stage and N stage was generated to improve the accuracy of prognosis evaluation. The immune cell infiltration analysis suggested that 5 hub genes (ADGRE1, OLR1, TIMP1, GPR132 and CCR7) were negatively correlated with tumor purity and positively correlated with the infiltration of immune cells.

Conclusion: Our study established a novel signature consisting of 7 hub genes for the prognostic prediction in patients with HNSCC.

Keywords: TCGA, inflammation-related gene, prognosis, HNSCC

Introduction

Head and neck squamous cell carcinoma (HNSCC) represents a group of malignant tumors originating from oral cavity, as well as the nasopharynx, oropharynx, hypopharynx, and larynx. It was estimated that there would be nearly 932,000 new cases of HNSCC and almost 619,000 new deaths of HNSCC all over the world in 2020.¹ The main etiologies for HNSCC contain human papilloma virus infection, overindulgence of alcohol, tobacco and betel.²–⁵ HNSCC is a heterogeneous and complicated disease, and the reported 5-year rate is only 46% in the United States owing to the high probability of metastasis or recurrence or both.⁶,⁷ The stage, anatomical location and epidemiological factors rendered it a tough challenge to the prognostic evaluation.⁸ Hence, exploring a prognostic signature on the basis of novel and reliable biomarkers for HNSCC is needed imminently.

Known as “the seventh marker of cancer”,⁹ many studies have indicated that inflammation plays an important role in the occurrence and development of cancer. Since the relationship between inflammation and cancer was first pointed out in 1863, the function of inflammation in the occurrence and progression of cancer has been one of the focus points in cancer research.¹⁰–¹⁴ The relationship between cancer and inflammatory response could be reflected by the content of specific substances in blood or tumor specimens.¹⁵ Specific substances, including Interleukin (IL)-6, IL-8, growth-related...
oncogene-1, vascular endothelial growth factor, hepatocyte growth factor, and the increasing level of cytokine and growth factor, were related to cancer progress and recurrence.\textsuperscript{16–19} Further studies showed that polymorphism expression of inflammation-related genes (IRGs) was associated with the development of HNSCC, indicating that IRGs played an important role in HNSCC.\textsuperscript{20} Being an auxiliary role, tumor microenvironment was also related to tumor promotion and development, which might be associated with IRGs and could be reflected on immune cells to some extent.\textsuperscript{9} Recently, the rapid development of sequencing technology and statistics has allowed researchers to study the role of IRGs in the prognosis of specific cancer, such as hepatocellular carcinoma and colon cancer.\textsuperscript{21,22} However, there is still no previous study regarding IRGs as prognostic markers of HNSCC.

In this study, we performed a systematic analysis of IRGs in HNSCC and identified HNSCC-associated IRGs. The potential function of these IRGs was explored to better understand the prognosis of HNSCC. Furthermore, clinical features and transcriptomic data were combined to further investigate the prognostic value of IRGs. Finally, we explored the relationship between IRGs and immune cells on the basis of the Tumor Immune Estimation Resource (TIMER) database.

\textbf{Materials and Methods}

\textbf{Data Collection}

Transcriptomic data and corresponding clinical information of 502 HNSCC patients were downloaded from The Cancer Genome Atlas (TCGA) data portal (\url{https://portal.gdc.cancer.gov/}), and a total of 97 samples from GSE41613 cohort were downloaded from the Gene Expression Omnibus (GEO) database (\url{https://www.ncbi.nlm.nih.gov/gds/}). A total of 200 IRGs were identified from the Molecular Signature database (Supplementary Table 1).\textsuperscript{21} A flowchart depicting the study is shown in Figure 1.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{flowchart.png}
\caption{The flowchart of our study.}
\end{figure}

\textbf{Abbreviations}: HNSCC, head and neck squamous cell carcinoma; DEGs, differentially expressed genes; DEIRGs, differentially expressed inflammation-related genes; LASSO, least absolute shrinkage and selection operator.
Identification of Differentially Expressed Inflammation-Related Genes (DEIRGs)

After the expression of 200 IRGs was obtained from TCGA cohort, “limma” package was performed to identify differentially expressed genes (DEGs) between HNSCC samples and adjacent normal samples. The criteria of DEGs were log fold change (logFC) ≥2 and false discovery rate (FDR) <0.01. Then, “ggplot2” and “pheatmap” packages were adopted to draw the volcano plot and heat map.

Functional Enrichment Analysis of DEIRGs

Although the relationship between DEIRGs and inflammation has been confirmed, other potential functions and closed pathways are unclear. Therefore, we performed Gene Ontology (GO) analysis to study the biological process, cellular component and molecular function of DEIRGs. In addition, Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was performed to identify the closed pathways of DEIRGs. The functional analyses, including GO and KEGG, were performed with “clusterProfiler” package.

Construction and Validation of an IRGs-Based Prognostic Signature for HNSCC

To evaluate the prognostic DEIRGs, univariate Cox regression analysis was conducted with a cutoff value of P < 0.05 in the TCGA cohort. Then, the Least Absolute Shrinkage and Selection Operator (LASSO) was used to minimize the probability of overfitting for the prognostic model.23,24 Finally, multivariate Cox regression was performed to select prognostic DEIRGs, and the selective DEIRGs were defined as hub genes and incorporated into the final IRGs-based prognostic model. The risk score formula of each patient was as follows: risk score = [Expression value of Gene1 × coefficient 1] + [Expression value of Gene2 × coefficient 2] +⋯+ [Expression value of Gene n × coefficient n].

In accordance with the median value of the risk score, patients were categorized into low- and high-risk groups. Kaplan–Meier (K-M) curve was used to analyze the survival difference between low- and high-risk groups. Time-dependent receiver operating characteristic (ROC) curve was generated to reflect the specificity and sensitivity of IRGs-based prognostic signature.

Establishment of a Clinical-IRGs Prognostic Nomogram for HNSCC

Clinicopathological parameters, such as age and AJCC stage, have been confirmed as robust prognostic predictors for cancer patients. Therefore, in order to confirm the independence of IRGs-based prognostic signature and further improve the accuracy of prognostic prediction, we incorporated the clinical data. The univariate and multivariate Cox analyses were orderly performed to study the independent overall survival (OS)-related variables. A nomogram was constructed based on the independent OS-related variables. The C-index was used to evaluate the discrimination of nomogram, and calibration curves were plotted to compare the actual and predictive survival for the evaluation of predictive performance.

Analysis of Gene Expression and Tumor-Infiltrating Immune Cells

To analyze the relationship between the expression of hub genes and tumor infiltrating immune cells, we selected TIMER, which allows researchers to systematically analyze immune infiltrations in multiple malignancies. Six immune cells, including B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages and dendritic cells were analyzed. Pearson’s correlation coefficient and the estimated P value were adopted to analyze the correlation between gene expression and infiltration level of immune cells.

External Validation of IRGs-Based Prognostic Signature in an Independent Cohort

In this study, GSE41613, which is a publicly available cohort from GEO database, including 97 HNSCC patients, was selected to validate our IRGs-based signature. The validation procedure included two steps. Firstly, according to the IRGs-based prognostic signature established in the TCGA cohort, the risk score of 97 patients in the GSE41613 was calculated. Then, the K-M survival curve and ROC curve of GSE41613 were generated to validate the performance of IRGs-based prognostic signature.
Results
Identification of DEIRGs in HNSCC Patients and Enrichment Analyses

The expression of 200 IRGs was obtained in the TCGA-HNSCC cohort. Through a differential analysis between 502 HNSCC samples and 44 adjacent normal samples, a total of 120 DEIRGs were determined, and most of DEIRGs were upregulated in tumor samples (Figure 2A, B and Supplementary Table 2). KEGG pathway analysis demonstrated that 120 DEIRGs were significantly correlated with the cytokine-, cancer- and immune microenvironment-related pathways (Figure 2C). In addition, the GO analysis also confirmed that these DEIRGs were significantly associated with inflammatory and immune functions, including regulation of leukocyte activation, secretory granule membrane and cytokine receptor binding were the most significant terms in the facet of biological process, cellular component and molecular function, respectively (Figure 2D). Generally, our study confirmed 60% IRGs were abnormally expressed in HNSCC samples, and these were associated with important oncobiology.

Figure 2 Identification of DEIRGs in HNSCC patients and function enrichment analysis. Volcano plot and heat map were drawn to show the differentially expressed genes in HNSCC versus normal samples (A and B); Functional enrichment analysis uncovered the potential function and closed pathway (C and D).
Establishment of an Inflammation-Related Gene Signature (IRGS)

After excluding three patients without follow-up data, a total of 499 HNSCC patients were included in the survival analysis, and 217 patients died during the follow-up period. The median follow-up time is 1.75 years, and the median OS of the present cohort is 4.57 years. In the univariate Cox analysis, 32 DEIRGs were significantly related with OS for HNSCC patients (Supplementary Table 3). Then, a total of 15 DEIRGs were selected according to the LASSO analysis (Figure 3A and B). Finally, by performing the multivariate Cox regression, seven DEIRGs were identified as hub genes to establish IRGS (Supplementary Table 4). The calculation of risk score was shown as follows: Risk score = [Expression level of ADGRE1 × 0.283652] + [Expression level of OLR1 × 0.146693] + [Expression level of GPR132 × (−0.292147)]  + [Expression level of TIMP1 × 0.156694] + [Expression level of CCR7 × (−0.327125)] + [Expression level of PSEN1 × 0.401136] + [Expression level of NAMPT × 0.127155].

Patients were divided into low- and high-risk groups based on the median risk score (Figure 4A). The K-M survival curve indicated that patients in the high-risk group had significantly worse OS than patients in the low-risk group (Figure 4B). The scatter chart demonstrated that patients in high-risk group were more likely and early to die than ones in low-risk group (Figure 4C). Time-dependent ROC curves showed that the values of area under curve (AUC) of IRGS were 0.685, 0.704 and 0.650 in 1-year, 3-year and 5-year, respectively. These results indicated the good discrimination of IRGS (Figure 4D). The heat map showed that ADGRE1, OLR1, TIMP1, PSEN1 and NAMPT were highly expressed in high-risk group, while GPR132 and CCR7 were high expression in low-risk group (Figure 4E).

Validation of IRGS in the GSE41613 Cohort

To test the reliability of IRGS in an independent cohort, patients in the GSE41613 cohort were categorized into low- and high-risk groups according to the cut-off value (Figure 5A). Similarly, high-risk patients showed worse OS than patients in the low-risk group, and were more likely to die earlier than low-risk patients (Figure 5B and C). The ROC curves of GSE41613 cohort also confirmed favorable discrimination of IRGS, and the AUC values were 0.686, 0.615 and 0.620 in 1-year, 3-year and 5-year, respectively (Figure 5D). Consistent with previous results in TCGA cohort, ADGRE1, OLR1, TIMP1, PSEN1 and NAMPT were highly expressed in high-risk group, whereas GPR132 and CCR7 got high expression in low-risk group (Figure 5E).

Figure 3 LASSO regression. LASSO regression was performed to select 15 DEIRGs (A and B).
Construction and Evaluation of a Nomogram Model

In the univariate Cox analysis, age, sex, M stage, N stage and risk score were selected as prognostic factors (Supplementary Table 5). Then, a multivariate Cox analysis was performed, and risk score, age, M stage and N stage were identified as independent prognostic factors (Supplementary Table 6). Combined with independent prognostic factors, a prognostic nomogram was constructed to provide survival analysis for individual HNSCC patients (Figure 6A). The C-index of the nomogram was 0.670. The calibration curves of this nomogram indicate a good agreement between predictive and actual 1-, 3-, 5-year survival analysis (Figure 6B–D).

Analysis of Gene Expression and Tumor-Infiltrating Immune Cells

In order to further explore the relationship between hub genes and tumor immune microenvironment, we analyzed the correlation of 7 hub genes and the infiltration level of immune cells (B cells, CD4+ T cells, CD8+ T cells, macrophages, neutrophils and dendritic cells) based on TIMER database. We discovered that most of the hub genes were significantly related to the infiltration level of immune cells, especially ADGRE1 (EMR1), OLR1, TIMP1, GPR132 and CCR7, which was negatively correlated with tumor purity and positively correlated with the infiltration of immune cells (Figure 7). This result suggested that ADGRE1, OLR1, TIMP1, GPR132 and CCR7 played specific roles in the immune infiltration of HNSCC.

Figure 4 Construction and Validation of IRGS for HNSCC. Patients were divided into low- and high-risk groups based on the median risk score (A); K-M survival curve indicated that patients in the high-risk group had significantly worse OS than patients in the low-risk group (B); Scatter chart showed that patients in high-risk group were more likely and earlier to die than ones in low-risk group (C); The 1-year, 3-year and 5-year AUC of ROC curve was 0.685, 0.704 and 0.650, indicating good discrimination (D); Different expression levels of 7 hub genes between two risk groups (E).

Abbreviations: TP, true positive; FP, false positive; AUC, area under the ROC curve; ROC, receiver operating characteristic.
Discussion

HNSCC is a complex group of malignant tumors and brings severe challenges to doctors in the therapy and prognosis evaluation. Being reputed as the seventh hallmark, inflammation plays a significant role in the development process of cancer and the essence of cancer development is genetic changes. Nevertheless, there is still no study regarding IRGs as prognostic markers of HNSCC. In this study, we systematically analyzed the expression of 200 IRGs in HNSCC tissues and their relationship with OS. 120 DEIRGs were screened out from the TCGA cohort and 32 DEIRGs were significantly associated with OS. Then, a prognostic model incorporating 7 hub genes was constructed and validated. Patients were classified into high- and low-risk groups. High-risk group was significantly associated with worse OS. Risk score was an independent prognostic factor of OS. Combined with risk score and some clinical features, a nomogram was established to predict the prognosis of HNSCC patients. We also identified a significant relationship between 7 hub genes and immune cells based on TIMER database.

Our IRGs was based on hub genes including ADGRE1, OLR1, GPR132, TIMP1, CCR7, PSEN1 and NAMPT. ADGRE1, named EMR1 in the past, was reported to be associated with the alpha fetoprotein level in hepatocellular carcinoma and plays a crucial role in carcinoma. It was reported that afucosylated chimeric antibodies directed against ADGRE1 offered a promising target for immune cell ablation, which might affect the normal immune system. OLR1

Figure 5 Validation of IRGS in the GSE41613 cohort. Patients were categorized into low- and high-risk groups according to the cut-off value (A); High-risk patients showed worse OS than patients in the low-risk group, and were likely to die earlier than low-risk patients (B and C). The 1-year, 3-year and 5-year AUC of ROC curve was 0.686, 0.615 and 0.620, indicating good discrimination (D); Different expression levels of 7 hub genes between two risk groups (E).

Abbreviations: TP, true positive; FP, false positive; AUC, area under the ROC curve; ROC, receiver operating characteristic.
was regarded to have several pro-oncogenic functions including inhibition of apoptosis and stimulation of proliferation by activating the NF-κB signaling pathway, activation of de novo lipogenesis and promoting the efficiency of adhesion and transendothelial migration. OLR1 encodes LOX-1 receptor protein, and LOX-1 plays some role in VEGF induction and HIF-1α Activation and MMP-9/MMP-2 expression, which promotes neoangiogenic and the epithelial–mesenchymal transition process of several types of cancer. Meanwhile, the knockdown of OLR1 may reduce the expression of SULT2B1 by downregulating c-MYC, therefore inhibiting glycolytic metabolism to restrain the proliferation and chemoresistance of colon cancer cells. TIMP1 could inhibit matrix metalloproteinase to restrict tumor progression, while the poor prognosis of almost all known cancer types was closely related to the high expression of TIMP1 with the utility of independent prognostic factor. TIMP1 was highly expressed in colon cancer samples, which led to tumor proliferation, metastasis and anti-apoptosis through FAK-PI3K/AKT and MAPK pathway. Animal experiments also revealed that the down-regulation of TIMP1 effectively weakened the tumorigenicity and metastasis in vivo.

PSEN1, one of the known causative genes for Alzheimer’s disease, was notably overexpressed in HNSCC tissues and cells, and was associated with poor prognosis and radioresistance of HNSCC. BLACA T1 was reported to mediate the radiosensitivity of HNSCC by regulating PSEN1. PSEN1 protein plays an indispensable role in Notch signaling. The activation of Notch signal is very important for the differentiation and maintenance of tumor stem cell-like cells in a...
variety of cancer types, including HNSCC. PSEN1 level can be reduced by miR-654-3p targeting CREB1 to inhibit sinonasal squamous cell carcinoma cells proliferation, invasion and migration. NAMPT was reported to be highly expressed in many sorts of cancers and associated with chemoresistance. NAMPT gene encodes nicotinamide phosphoribosyltransferase, which acts as a rate-limiting enzyme in the biosynthesis of nicotinamide adenine dinucleotide

Figure 7 Analysis of gene expression and tumor-infiltrating immune cells in HNSCC.
(NAD). NAD can regulate many cellular processes, such as DNA repair and gene expression to maintain tumor growth and survival and energy requirements. Combined with SIRT1, NAMPT activates OSKM factor and stem cell signaling pathway to promote the dedifferentiation and stemness from cancer cells to cancer stem-like cells. A subgroup of cancer stem-like cells is important for the occurrence, development, drug resistance and metastasis of tumors.

GPR132 and CCR7 were the two hub genes with negative coefficients in the risk score formula and comparatively highly expressed in the low-risk group. G protein-coupled receptor, encoded by GPR132, is a key macrophage sensor of the rising lactate in the acidic tumor microenvironment. It is reported that the expression of GPR132 is related to breast cancer and skin tumors. However, the exact function of GPR 132 in HNSCC is still unclear, and more exploration should be performed in the future. The CCR7 chemokine axis played two opposite roles in cancer. On the one hand, the axis helped the transportation of numerous effector cells involved in immune responses to combat the spread of cancer. On the other hand, it played a significant part in the lymph node metastasis and distant metastasis. In HNSCC, the high expression of CCR7 in low-risk group might be attributed to the predominance of the former role.

In univariate and multivariate Cox regression analyses, we identified that risk score was an independent factor of HNSCC. Then, some clinical characteristics were selected. A prognostic nomogram model was constructed to calculate individual risk score and predict OS, which would be conducive to prognosis evaluation and decision-making. In other words, the nomogram would contribute to the diagnosis and treatment of HNSCC patients.

Previous studies indicated that the antagonism between inflammation and immunity affected the development of cancer and treatment outcomes. Inflammation in the tumor microenvironment can promote tumor process through regulating antitumor immunity and subverting adaptive immune responses. The immune infiltration analysis indicated that IRGs including ADGRE1, OLR1, GPR132, TIMP1 and CCR7 were negatively correlated with tumor purity and positively associated with the infiltration level of immune cells. It demonstrated that those genes were mainly expressed in immune cells rather than tumor cells and played specific roles in the regulation of tumor microenvironment. What is more, tumor microenvironment, a necessary part to promote the neoplastic process, was largely regulated by inflammatory cells. In short, these hub genes were not only associated with inflammatory responses but also immune cells, which might be potential therapeutic targets for HNSCC patients.

There are numerous similar studies based on biological markers of HNSCC. However, there are two main advantages in our study. Firstly, a multi-gene signature was chosen for our study. MAGE-A11 was reported to be a potentially valuable prognostic biomarker in HNSCC. With the combination of two independent cohorts, PIK3CA was identified as a potential prognostic marker for recurrence in HNSCC. However, the expression of a single gene may vary among the population due to the heterogeneity of the samples, while a multi-gene signature is available to compensate for individual differences and reduces possible errors. Therefore, compared with markers based on single-gene expression, a multi-gene signature has higher predictability and better accuracy. Hence, our research constructed a multi-gene signature for better predictive performance and accuracy. Second, both immune infiltration analysis and nomogram establishment were utilized while maintaining similar AUC in our study. Previous studies indicated that ferroptosis-related gene signature, autophagy-related gene signature, glycolysis-related gene signature and immune checkpoint-related gene signature predicted 5-year OS for HNSCC with AUC at 0.645, 0.622, 0.675 and 0.625, respectively, which were similar to our study. What is more, we also obtained 499 samples in an attempt to reduce potential errors and promote reliability, which reduces the deviation of small-scale samples to some extent. Compared with the previous studies, both immune infiltration analysis and nomogram establishment were performed in our research to explore the relationship between hub genes and immune cells and provide a tool for individual prognosis evaluation.

A novel signature consisting of 7 hub genes was constructed for the prognostic prediction of HNSCC patients. However, there are still some limitations. Firstly, the samples in our study were retrospective; in this case, prospective samples are needed to verify. To compensate this point to some extent, we endeavored to study HNSCC in as many samples as possible. Secondly, the biomarkers acquired from the study should be combined with clinical studies to refine their functions and lay a good foundation for better therapy. We are looking forward to the cooperation with research centers and hospitals for further exploration.
Conclusion
In conclusion, our study established a novel signature consisting of seven hub genes, which has good AUC values in both training set and validation set. The results suggested that high-risk score was associated with worse prognosis in HNSCC patients. Risk score, age, M stage and N stage were selected to establish a nomogram for individual prognosis evaluation. Hub genes including ADGRE1, OLR1, TIMP1, GPR132 and CCR7 contributed to the immune process, which is consistent with the results of functional enrichment analyses.

Abbreviations
HNSCC, head and neck squamous cell carcinoma; IRGs, inflammation-related genes; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; LASSO, Least Absolute Shrinkage and Selection Operator; TIMER, Tumor Immune Estimation Resource; IRGS, inflammation-related gene signature; IL, Interleukin; DEIRGs, differentially expressed inflammation-related genes; DEGs, differentially expressed genes; logFC, log fold change; FDR, false discovery rate; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes, K-M, Kaplan–Meier; ROC, receiver operating characteristic; OS, overall survival; AUC, area under curve; NAD, nicotinamide adenine dinucleotide.

Ethics Approval
The present study has been submitted to our institutional review board for review, and the ethics approval was not required due to the data of present study were obtained from public databases. The need for review was waived by the Ethics Committee of The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang Province, China.

Author Contributions
All authors contributed to study design, acquisition, and analysis of data, and writing and revising manuscripts and reached a consensus on the journal to which the article was submitted, approved the final version to be published and agreed to be responsible for all content of the article.

Disclosure
The authors report no conflicts of interest in this work.

References


40. Navas LE, Carnero A. NAD metabolism, stemness, the immune response, and cancer. Signal Transduct Target Ther. 2021;6(1). doi:10.1038/s41392-020-00354-w


