Exploration of prognostic biomarkers in head and neck squamous cell carcinoma microenvironment from TCGA database

Ying Li#, Jianping Bi#, Guoliang Pi, Hanping He, Yanping Li, Guang Han

Department of Radiation Oncology, Hubei Cancer Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Contributions: (I) Conception and design: G Han, J Bi, Yanping Li; (II) Administrative support: None; (III) Provision of study materials or patients: Ying Li, G Pi, Yanping Li; (IV) Collection and assembly of data: J Bi, H He; (V) Data analysis and interpretation: G Han, J Bi, Ying Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#These authors contributed equally to this work.

Correspondence to: Guang Han. Department of Radiation Oncology, Hubei Cancer Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430079, China. Email: hg7913@hotmail.com.

Background: Immune checkpoint blockade (ICB) therapies have redefined human cancer treatment, including for head and neck squamous cell carcinoma (HNSCC). However, clinical responses to various immune checkpoint inhibitors are often accompanied by immune-related adverse events (irAEs). Therefore, it is crucial to obtain a comprehensive understanding of the association between different immune tumor microenvironments (TMEs) and the immunotherapeutic response.

Methods: The research data were obtained from The Cancer Genome Atlas (TCGA) database. We applied RNA-seq genomic data from tumor biopsies to assess the immune TME in HNSCC. As the TME is a heterogeneous system that is highly associated with HNSCC progression and clinical outcome, we relied on the Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) algorithm to calculate immune and stromal scores that were evaluated based on the immune or stromal components in the TME. Then, the Tumor Immune Dysfunction and Exclusion algorithm (TIDE) was used to predict the benefits of ICB to each patient. Finally, we identified specific prognostic tumor-infiltrating immune cells (TIICs) by quantifying the cellular composition of the immune response in HNSCC and its association to survival outcome, using the CIBERSORT algorithm.

Results: Utilizing the HNSCC cohort of the TCGA database and TIDE and ESTIMATE algorithm-derived immune scores, we obtained a list of microenvironment-associated lncRNAs that predicted different clinical outcomes in HNSCC patients. We validated these correlations in a different HNSCC cohort available from the TCGA database and provided insight into the prediction of response to ICB therapies in HNSCC.

Conclusions: This study confirmed that CD8+ T cells were significantly associated with better survival in HNSCC and verified that the top five significantly mutated genes (SMGs) in the TCGA HNSCC cohort were *TP53*, *TTN*, *FAT1*, *CDKN2A*, and *MUC16*. A high level of CD8+ T cells and high immune and stroma scores corresponded to a better survival probability in HNSCC.

Keywords: Head and neck squamous cell carcinoma (HNSCC); The Cancer Genome Atlas (TCGA); tumor microenvironment (TME); tumor-infiltrating immune cell (TIIC); immune checkpoint blockade (ICB)

Page 2 of 15

Li et al. Genes with prognostic value in HNSCC microenvironment

Introduction

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common type of cancer worldwide, accounting for approximately 4% of all cancers (1-3). Despite significant treatment progress in surgery, radiotherapy, and concomitant chemoradiotherapy, the mortality rate remains as high as 40–50% (4). Over the last decade, immune checkpoint inhibitors (ICIs) have emerged as a promising therapy for various cancers due to their prominent antitumor efficacy. Programmed death-1 (PD-1) inhibitors, such as pembrolizumab and nivolumab, programmed death-ligand 1 (PD-L1) inhibitors (durvalumab, avelumab, and atezolizumab) and Ipilimumab, a CTLA-4 inhibitor, soon followed, presenting promising results (5). However, clinical responses after anti-PD-1/PD-L1 treatment are heterogeneous; the majority of HNSCCs are resistant to immunotherapy ab initio, and serious immune-related adverse events (irAEs) are seen in patients experiencing ICI therapy (6,7). Malignant carcinomas, including HNSCC, are characterized not only by the intrinsic activities of cancer cells but also by the immune cells recruited to, and activated in, the tumor microenvironment (TME) (8). The sophisticated mechanistic basis remains largely unclear. According to the previous study, this may be due d to factors in the TME, such as a lack of proper rejection antigens, lack of immune surveillance, or the presence of immunosuppressive mediators (9). Therefore, a comprehensive investigation of immune profiling in the progression of HNSCC tumors may promote the development of new immunotherapeutic agents and novel treatment regimens. In this study, we mined The Cancer Genome Atlas (TCGA) database for genes with prognostic value and responses to various ICIs in the HNSCC microenvironment.

CD8[+] T cells have an affinity for major histocompatibility complex (MHC) class I molecules and are key anticancer immune cells (10,11). Additionally, CD8[+] T cells are the main effectors in PD-1 blockade-induced antitumor responses, and reinvigoration of exhausted CD8[+] T cells is one of the major determining factors of responsiveness to PD-1 blockade (12,13). For instance, in breast cancer, previous studies have shown a significantly increased number of CD8[+] T cells at tumor sites, which was negatively correlated with advanced tumor stages and positively correlated with clinical outcomes (14,15). CD8[+] T cells were also significantly associated with the clinical outcomes of HSCNN. In the study of Zhang *et al.*, they found that CD8[+] T cells exhaustion in the TME of HNSCC determines poor prognosis (16). Moreover, CD8[+] T cells differentiate to cytotoxic T cells, traffic into the TME, and exhibit cytotoxicity against tumor cells (17). Upon arrival in tumor beds, CD8[+] T cells inevitably face many obstacles, including the intrinsic checkpoint regulators PD1-PD-L1 and CD28-CTLA-4; an abnormal TME contain protumor inflammation; antigen loss and immune evasion of tumor targets; and tissue-specific alterations (18).

The TME is a complex microenvironment where the tumor cells are generated and located, consisting of immune cells, stromal cells, endothelial cells, surrounding blood vessels, inflammatory mediators and extracellular matrix (ECM) molecules (19). Cytotoxic T lymphocytes (CTL), natural killer (NK) cells, myeloid-derived suppressor cell (MDSC), regulatory T cell (Treg) and tumor-associated macrophage are immune cells which play vital roles in tumor biological process (20). Cancer-associated fibroblasts (CAFs) construct the main bulk of stromal cells in the TME to promote the growth of cancer cells (21,22). Also, the TME in HNSCC is comprised of heterogeneous non-malignant cells that integrated in a complex ECM (23). Increasing evidences indicated that tumors can be classified by the components of TME and stromal cell proportions or activations status (24,25). TME of different types of tumors have their own characteristics and also commonalities. However, the TME of HNSCC is characterized by some unique features compared to other cancer types (26).

Highlight box

Key findings
- A high level of CD8[+] T cells and high immune and stroma scores corresponded to a better survival probability in HNSCC.
- The top five significantly mutated genes in the TCGA HNSCC cohort were *TP53*, *TTN*, *FAT1*, *CDKN2A*, and *MUC16*.

What is known and what is new?
- Tumor-infiltrating immune cells were found to be significantly associated with prognosis and the identification of immunotherapy targets.
- This study is the first to utilize the HNSCC cohort of the TCGA database and TIDE and ESTIMATE algorithm-derived immune scores to obtain a list of TME-associated lncRNAs to predict clinical outcomes in HNSCC patients.

What is the implication, and what should change now?
- A comprehensive investigation of immune profiling in the progression of HNSCC tumors may promote the development of new immunotherapeutic agents and novel treatment regimens.

HNSCC patients have decreased absolute T cells count in tumor and circulation (27). Moreover, the interaction of various cellular components in the TME along with the production of cytokines and chemokines profoundly impact the function of T cells (26,28,29). Decreased number of NK cells is also common in HNSCC, leading to seriously immunodeficient tumors (30,31). *YTHDF1* expression was significantly associated with the TME in HNSCC, which was correlated with CD4+ T cells and CD8$^+$ T cells infiltration (32). In addition, HNSCC is characterized by desmoplastic stromal fibroblasts that promote tumor invasion and progression through autocrine and paracrine factors (33,34). In the TME, immune and stromal cells are the two main types of non-tumor components and have been suggested that they are valuable for the diagnostic and prognostic biomarkers of tumors. In recent year, algorithms have been developed to predict tumor purity using gene expression data based on the TCGA database. For example, Yoshihara *et al.* designed the Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) algorithm (35). Immune and stromal scores to predict the infiltration of non-tumor cells by analyzing the specific gene expression signatures of immune and stromal cells were calculated based on the ESTIMATE algorithm. Many studies have applied the ESTIMATE algorithm to prostate cancer (36,37), breast cancer (38,39), colorectal cancer (40-42) and hepatocellular carcinoma (43), which demonstrated that such big-data-based algorithms are effective, although its utility on immune and/or stromal scores in glioblastoma multiforme (GBM) has not been explored in detail.

Mounting research has revealed that long noncoding RNAs (lncRNAs) play a pivotal role in cancer progression (44-46). However, their potential involvement in HNSCC remains to be elucidated. In this paper, we investigated the CD8$^+$ T cell profiles in the TME and their clinical value in HNSCC patients using the TCGA RNA sequencing data. Signature lncRNAs combined with their coefficients in an elastic net model were used to calculate the risk score for every patient. The "ConsensusClusterPlus" R package clustered patients by the expression level of 22 selected lncRNAs. We further predicted the benefits of immune checkpoint blockade (ICB) to each patient using the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm (http://tide.dfci.harvard.edu/). Moreover, mutation patterns of significantly mutated genes (SMGs) in relation to HNSCC subtypes and immune score and stromal score were calculated to reveal the relationship between tumor immune status and prognosis. Finally, tumor-infiltrating immune cells (TIICs) were found to be significantly associated with prognosis and the identification of immunotherapy targets.

This study is the first to utilize the HNSCC cohort of the TCGA database and TIDE and ESTIMATE algorithm-derived immune scores to obtain a list of TME-associated lncRNAs to predict clinical outcomes in HNSCC patients. Notably, we validated these correlations in a different HNSCC cohort available from the TCGA database. And were able to provide insight into the prediction of response to ICB therapies in HNSCC. We present the following article in accordance with the REMARK reporting checklist (available at https://atm.amegroups.com/article/view/10.21037/atm-22-6481/rc).

Methods

Data profiling

The lncRNA expression microarray data and clinical characteristics, including medication history, histologic grade, pathological TNM stage, and survival information for 546 TCGA-HNSC samples from 528 TCGA-HNSC patients was retrieved from the TCGA database using the TCGA "biolinks" R package. Patients with complete survival information and genetic expression data at that time point were included in this study. The detailed clinicopathological information for the TCGA data sets are shown in Table S1. The lncRNA annotation information was downloaded from Ensemble (ftp://ftp.ensembl.org/pub/release-99/gtf/homo_sapiens/Homo_sapiens.GRCh38.99.gtf.gz). The lncRNA annotations were extracted from the Ensemble gtf annotation file based on genecode-defined lncRNA biotypes (https://www.gencodegenes.org/pages/biotypes.html). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

lncRNA differential expression analysis

The lncRNA expressions (fpkm values matrix) were extracted from the total RNA fpkm matrix by the Ensemble ID of the lncRNAs. Samples were classified into normal or tumor groups according to tissue definition. Differentially expressed lncRNAs between the normal and tumor samples were qualified using the "DESeq2" R package. lncRNAs with an adjusted P value <0.0001 were selected to plot the heatmap and volcano plot.

Page 4 of 15

Li et al. Genes with prognostic value in HNSCC microenvironment

Construct and validation of survival-related lncRNA signatures

To find lncRNAs that were most relevant to the prognosis of patients, we first selected the top 1,000 most significant genes using univariate Cox models (genes were ranked by the Wald test P value). We then selected the most related lncRNAs using an elastic net model in the "glmnet" R package with an alpha =0.1 and family = "Cox". These selected signature lncRNAs combined with their coefficients in the elastic net model were used to calculate the risk score for every patient. The risk score was defined as the sum of products from all signature lncRNAs and their coefficients. After calculating the risk score for every patient, these patients were then divided into two groups by the median value of all risk scores. The survival difference was calculated using the "survival" R package and plotted with the "survminer" R package.

$$\text{Risk score} = \sum_{i=1}^{n} \left(Coef_i * x_i \right) \qquad [1]$$

Consensus clustering

The "ConsensusClusterPlus" is an open resource, Bioconductor-compatible R software package for unsupervised class discovery (47). The algorithm starts by subsampling a proportion of items and a proportion of features from a data matrix. In this study, we performed consensus clustering using "ConsensusClusterPlus", where the lncRNAs for each HNSCC sample were divided into three subgroups and measured by the median absolute deviation (MAD). The most variable genes were used for subsequent clustering.

Differential expression level of 22 selected lncRNAs

After selecting the signature lncRNAs using the elastic net model, we clustered patients by the differential expression level of 22 selected lncRNAs using the "ConsensusClusterPlus" R package (50 iterations, 80% resampling rate, Pearson correlation, http://www.bioconductor.org/). The "PCA" R package (R v3.5.1) was used to investigate gene expression arrays in the HNSCC sample groups.

Estimation of tumor mutational burden (TMB)

TMB has been recognized as an emerging therapeutic measure of sensitivity to predict immunotherapy response in clinical oncology. The TMB score of each patient with HNSCC was calculated as previously described (48).

Immunotherapeutic response prediction

The programmed cell death 1 PD-1 and cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) pathways are associated with tumor immune evasion; therefore, ICIs targeting PD-1 and CTLA-4 enhance the antitumor immunity. To predict clinical responses to ICIs, we used the TIDE algorithm and subclass mapping as previously described (49,50).

Correlations between prognoses and stromal/immune scores

To investigate the relationship between tumor immune status and prognosis, we calculated the immune and stromal scores and identified the optimal score cutoff for grouping patients most significantly to the maximally selected rank statistics using the "maxstat" R package (51). Before calculating the scores, we first filtered genes that showed an fpkm value smaller than 5 in all samples. Patients were divided into 4 stages based on the stromal and immune scores estimated from each HNSCC sample and the prognoses for each group were examined. The log-rank tests were used to compare the survival outcomes.

Immune TME analysis

To explore the differences in immune cell subtypes for each HNSCC sample, the CIBERSORT package was used to value the fraction of 22 immune cell subtypes, based on an expression file as previously described (52). We set the number of permutations to 1,000. A total of 22 types of TIICs were quantified for each sample, together with CIBERSORT metrics, including CIBERSORT P value, Pearson's correlation coefficient and root mean square error. The "Genefilter" R package was used to screen each sample, and the threshold was set at a P value <0.01. The final CIBERSORT output was subsequently analyzed.

Statistical analyses

All statistical analyses were performed using R statistical software (https://www.r-project.org/, version 3.5.2) and GraphPad Prism (version 8.0, LaJolla, CA, USA).
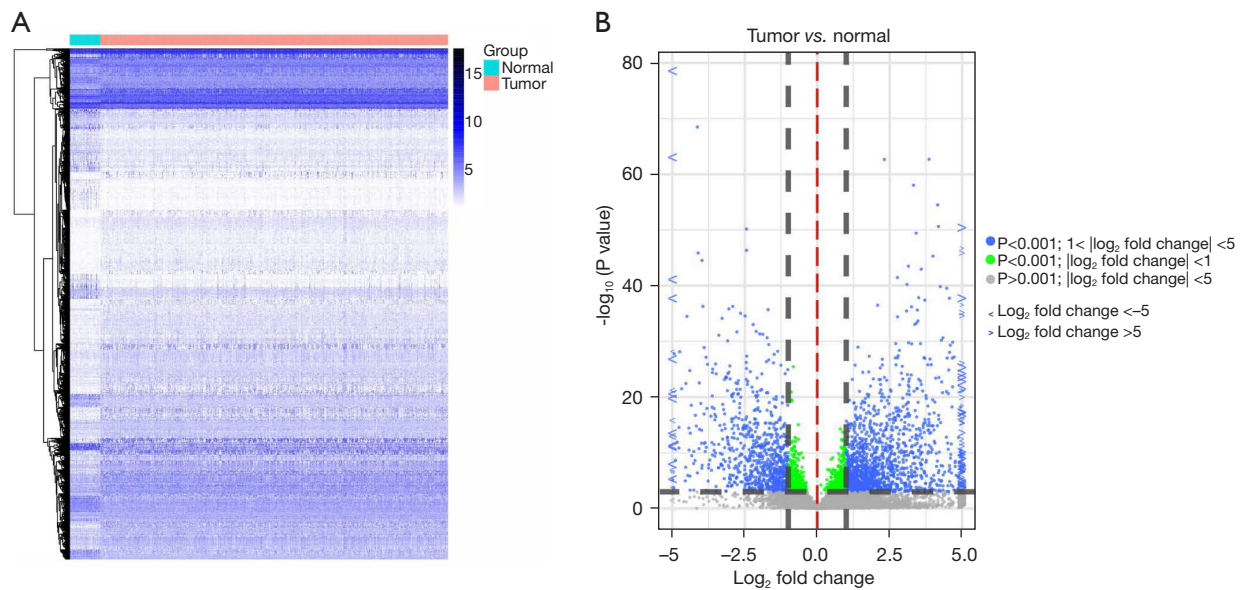
Figure 1 Comparison of lncRNA expression profiles between HNSCC tumor samples and normal samples. (A) Heatmap of differentially expressed lncRNA in normal and tumor tissue. Differentially expressed lncRNA are lncRNA show an adjusted P value ≤0.0001. Genes shown in red are upregulated and genes in blue are downregulated. (B) The significantly upregulated and downregulated lncRNA in HNSCC tumor samples and normal samples are shown in the volcano plot (log$_2$ fold change >1.0, P<0.01). HNSCC, head and neck squamous cell carcinoma; lncRNA, long noncoding RNA.

Progression-free survival (PFS) and overall survival (OS) analyses were estimated by Kaplan-Meier curves, and the P value was determined by a log-rank test. Univariate and multivariate analyses of PFS and OS were performed with the "rms" package. P values were adjusted for multiple testing using the false discovery rate (FDR) approach by Benjamini and Hochberg. A P value <0.05 was considered statistically significant.

Results

lncRNA differential expression analysis

To investigate the differentially expressed lncRNAs in normal and tumor tissue samples, the heatmap of the lncRNA expression pattern was constructed by the "DESeq2" R package (*Figure 1A*). A total of 2,281 differentially expressed genes containing 1,466 upregulated genes and 815 downregulated genes were detected after analysis of the TCGA-HNSCC cohort (*Figure 1B*). The differently expressed lncRNA fold-change and corresponding adjusted P values are provided in https://cdn.amegroups.cn/static/public/atm-22-6481-1.xls. In addition, we downloaded the transcripts per million TPM values

of tumor and normal samples from the TCGA-HNSCC dataset and performed weighted gene co-expression network analysis (WGCNA) using the WGCNA R package. In the WGCNA analysis, a co-expression network was constructed according to the co-expression of all the genes in the samples, and 130 modules were identified, among which the MEroyalblue module was the most relevant to the normal samples, and the MElightyellow module was the most relevant to the tumor samples (Figure S1).

Prognostic signature construction with the LASSO Cox regression model using the training set

The LASSO Cox regression model was constructed using the "glmnet" package in R. Relying on a 10-fold cross-validation, we chose 0.645 as the minimum value criteria for λ (*Figure 2A*). We calculated the risk score of every patient for 1,000 selected signature lncRNAs combined with their coefficients in the elastic net model. Subsequently, we assigned patients into the low- and high-risk groups based on the median value of all risk scores. The survival curve verified that patients with a low-risk score had a preferable survival rate compared to those with high-risk scores (*Figure 2B*).
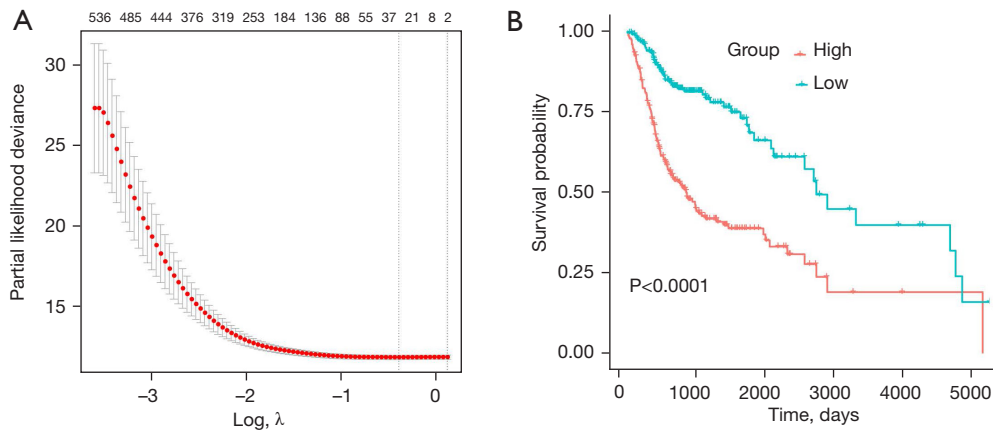
Page 6 of 15

Li et al. Genes with prognostic value in HNSCC microenvironment

Figure 2 LASSO Cox regression model construction and survival analysis. (A) λ selection by 10-fold cross-validation. The shrinkage curve of elastic net in selecting signature lncRNA. Adjusting parameter (lambda) screening in the LASSO regression model, lambda with lowest partial likelihood deviance were selected. Solid vertical lines represent partial likelihood deviance ± standard error. The dotted vertical lines are drawn correspond to the optimal values by minimum criteria and 1- standard error criteria. The value of 0.6457363 was chosen for λ by 10-fold cross-validation with the minimum criteria. (B) Survival curve of low- and high-risk groups stratified by their risk score. LASSO, Least Absolute Shrinkage and Selection Operator; lncRNA, long noncoding RNA.
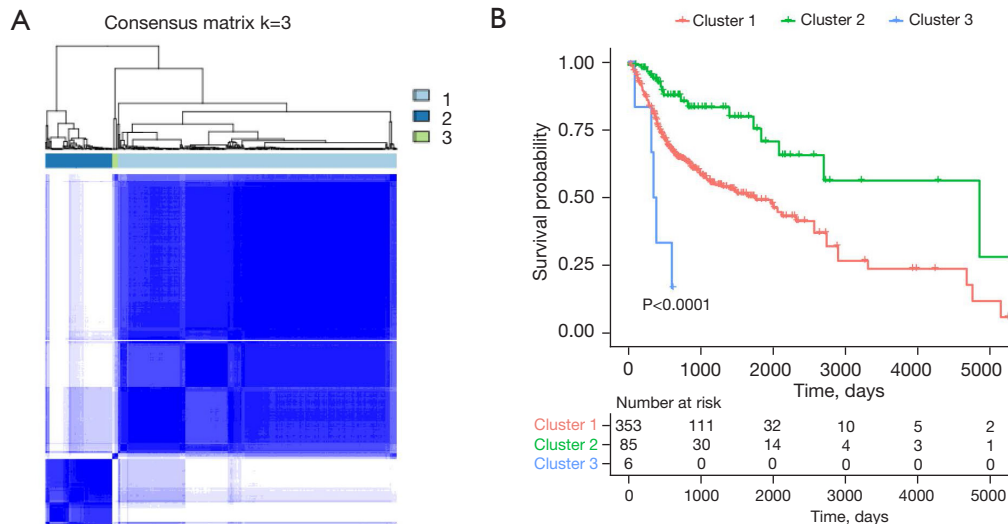

Figure 3 The selected 22 signature lncRNA and its association with predicted survival probability. (A) Consensus clustering matrix of 508 samples from TCGA dataset for k=3. (B) Survival difference among three groups of patients. Three groups of patients show significant survival difference. lncRNA, long noncoding RNA; TCGA, The Cancer Genome Atlas.

Consensus clustering

The elbow method and gap statistic were used to group the HNSCC samples into three clusters. With K=3, the elbow curve smoothly decreased (*Figure 3A*). Consensus clustering of related gene expressions grouped the HNSCC sample into three clusters with different pathological features and clinical outcomes. From the clustering results, we also observed a significant survival difference among the different groups of patients. Specifically, patients of cluster 3 exhibited the best survival probability, followed by those of cluster 1, and patients of cluster 2 showed the worst survival probability (*Figure 3B*).
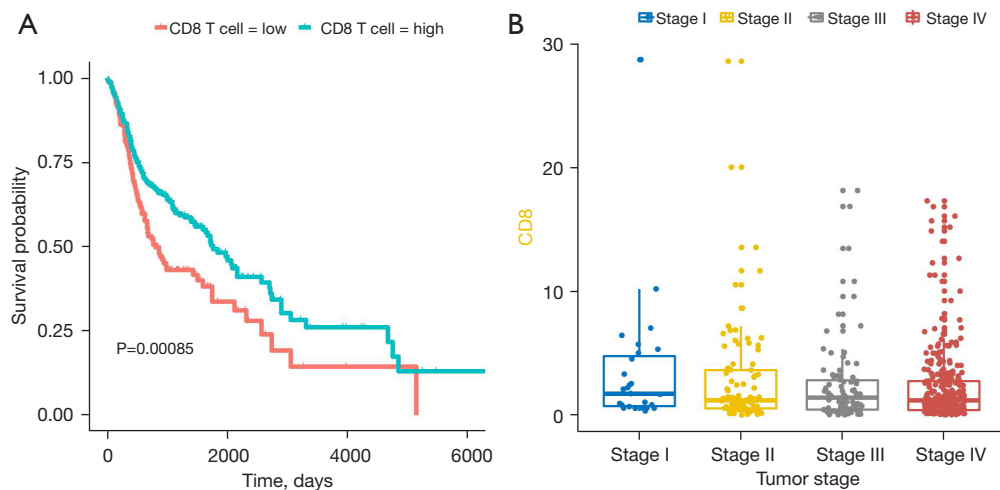
Figure 4 Immune therapy analysis. (A) The expression level of CD8 is significantly related the survival of patients. (B) Expression level of CD8 in different HNSCC stages. HNSCC, head and neck squamous cell carcinoma.

The TIDE signature predicted ICB response

$CD8^+$ T cells are considered the major anti-cancer effector cells as they can give rise to CTLs that kill cancer cells presenting a specific pMHC complex. To investigate the response of these patients to immune therapy, the TIDE (http://tide.dfci.harvard.edu/) algorithm was used to predict the ICB of each patient. A high expression of $CD8^+$ T cells was significantly related to a better prognosis in HNSCC (*Figure 4A*). When investigating the expression level of $CD8^+$ T cells in patients at different stages, the expression level of $CD8^+$ T cells showed a decrease from stage I to stage IV, which implied that the proportion of $CD8^+$ T cells experienced a decline in line with HNSCC progression (*Figure 4B*).

Mutation patterns of SMGs in relation to HNSCC subtypes

The TMB is an independent indicator of a better response to ICB treatment (53). The average TMB was 3.48 mutants per Mb and the detailed TMB information can be found in *Figure 5A* and https://cdn.amegroups.cn/static/public/atm-22-6481-2.xls. Gene expression has emerged as one of the most robust and reliable approaches to differentiate human lymphomas. Using manual selection, we divided 508 tumor tissues into the training and validation cohorts. We identified 20 SMGs in the TCGA HNSCC cohort. An oncoplot of the top 50 mutated genes is shown in *Figure 5B*. *TP53*, *TTN*, *FAT1*, *CDKN2A*, *and MUC16* were much more frequently mutated in the high-risk

subtype (Fisher exact test, OR >1, P<0.05; *Figure 5B*). The associations between these SMGs and the high-risk subtype were still significant after including the factors of age, gender, stage, and prognostic status (OR >1, P<0.01).

Immune and stromal scores were significantly associated with HNSCC subtypes

To investigate the relationship between tumor immune status and prognosis, we calculated the immune and stromal scores using the "estimate" R package After excluding patients without complete clinical information and the normal control group, 508 patients with HNSCC were enrolled in this research. Relying on the pathological diagnosis given in the TCGA database, HNSCC was divided into four stages. The stromal scores among the different HNSCC stages decreased from stage I to stage III, and increased from stage III to stage IV (*Figure 6A*). However, the distribution of the immune scores among the different HNSCC stages was quite stable from stage I to stage IV (*Figure 6B*).

Association of stromal and immune scores with HNSCC patient tumor stages and prognosis

After excluding patients without complete clinical information and the normal control group, 508 HNSCC patients were enrolled in this research. Relying on the pathological diagnosis given in the TCGA database, HNSCC was classified into four stages. The association
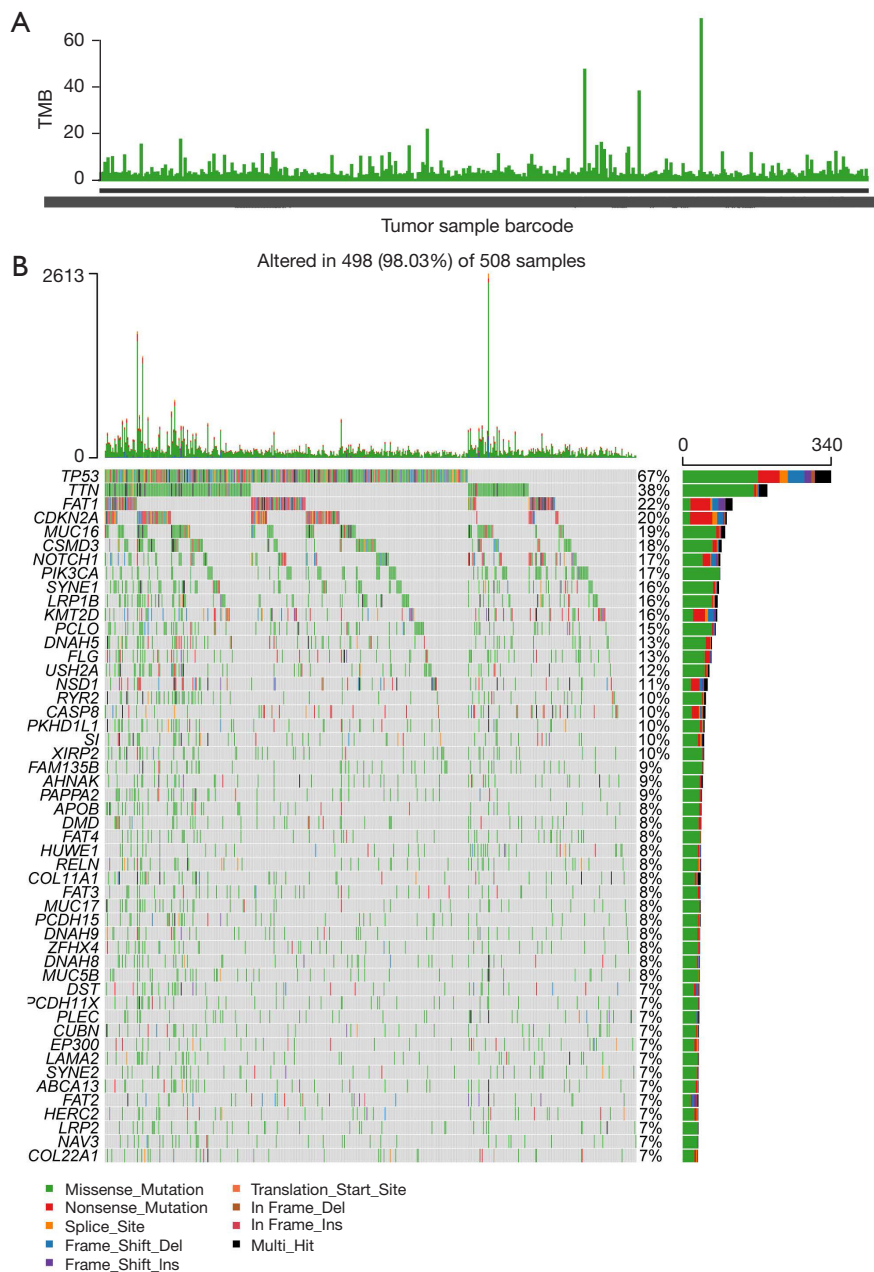
Figure 5 TMB score and mutation status of HNSCC patients. (A) TMB for every patient. (B) Oncoplot for the top 50 mostly mutated genes. TMB, tumor mutational burden; HNSCC, head and neck squamous cell carcinoma.

between the stromal and immune scores with HNSCC patient pathologic characteristics was qualified by comparing the score distributions in different tumor stages (*Figure 6A,6B*). A roughly increased stromal and immune scores with the advancing of tumor stage could be seen. There was a significant association between stromal (P=0.015) and immune scores (P=0.00088) and HNSCC

patients survival probability (*Figure 6C,6D*).

Composition difference and prognosis value of TIICs in HNSCC samples

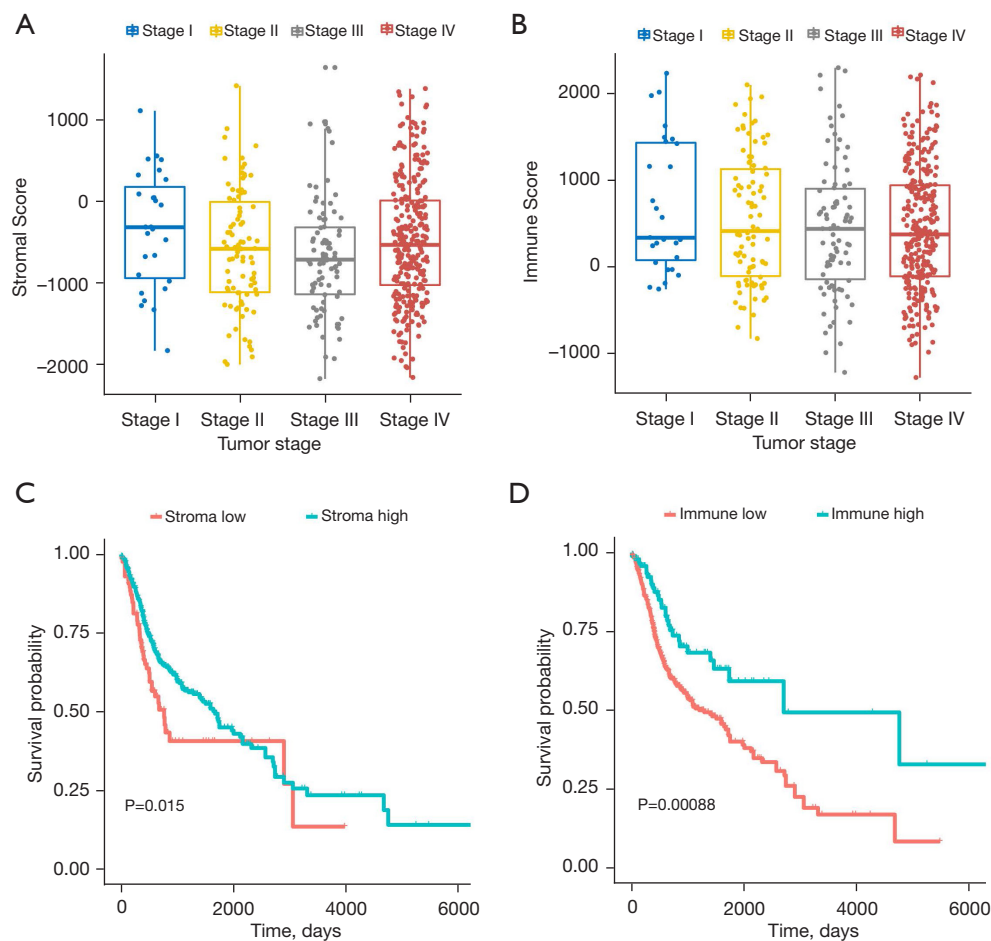The CIBERSORT analytical tool allowed us to specifically analyze the intrinsic fractions of 22 subpopulation TIICs in

Figure 6 The relationship between HNSCC tumor immune status and prognosis. (A) Distribution of stromal score among different stage of tumors. (B) Distribution of immune score among different stage of patients. (C) The effects of stromal score on the survival of patients. (D) Prediction of HNSCC patients' prognosis corresponding with different immune score. HNSCC, head and neck squamous cell carcinoma.

the bulk tumor samples. Insight into TIICs may be helpful in explaining the initiation and development of HNSCC. The total value of the 22 subset immune cells in each sample as one, and the fraction of all 22 subpopulations of immune cells in each sample is depicted in *Figure 7A*. Calculating the average fraction of each subset of TIICs among the HNSCC samples, the top five average fractions were M0 macrophages, CD8 T cells, M2 macrophages, CD4 resting memory T cells, and M1 macrophages. Detailed information is shown in https://cdn.amegroups.cn/static/public/atm-22-6481-3.xls. Moreover, as presented in *Figure 7B*, we used unsupervised hierarchical clustering based on the above identified cell subsets to divide the HNSCC samples into two discrete groups. It was clear that the fractions of immune cells in the different HNSCC samples were significantly varied. Hence, we inferred

that variation in TIIC fractions might be an essential characteristic of HNSCC. We further identified the prognostic subsets of TIICs in HNSCC. The Kaplan-Meier plots and log-rank tests for the above-identified TIIC subpopulations showed that high expression levels of immune cells, including CD8 T cells (P=0.0014), follicular helper T cells (P=0.0024), and regulatory T cells (P=0.00034) were associated with a better OS; whereas high expression levels of CD4[+] memory T cells (P=0.00035) corresponded to a poor OS (*Figure 7C-7F*).

Discussion

As increasing application of immunotherapy in cancer, the information of individual's immune status could help identify those who will resist to ICB therapies. TME has
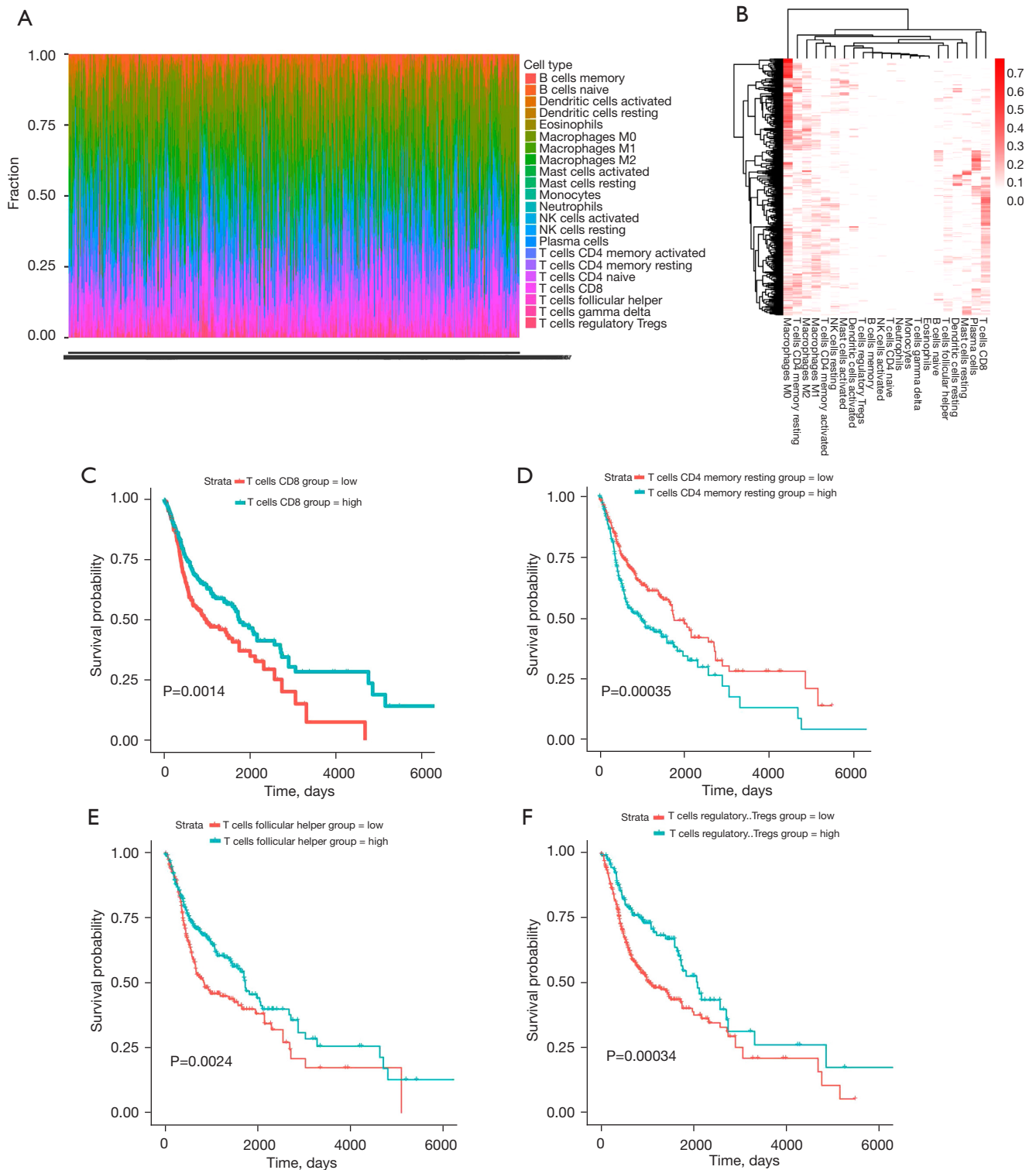
Figure 7 The landscape of immune infiltration in HNSCC. (A) The difference of immune infiltration in HNSCC samples. (B) Heat map of the 22 immune cell fractions in TCGA-HNSCC cohort. The vertical axis depicts the clustering data of samples which were divided into two discrete groups. (C) Survival plot of median of CD8 T cells. (D) Survival plot of median of CD4+ memory T cells. (E) Survival plot of median of follicular helper T cells. (F) Survival plot of median of regulatory T cells. The P values are from log-rank tests. HNSCC, head and neck squamous cell carcinoma; TCGA, The Cancer Genome Atlas.

been reported to seriously affect gene expression of tumor cells, thus the clinical outcomes. The TME is a complex microenvironment where the tumor cells are generated and located. It is comprised of mesenchymal cells, stromal cells, immune cells, endothelial cells, surrounding blood vessels, as well as inflammatory moderators and ECM molecules (24,54). In the TME, immune and stromal cells are two notable kinds of non-tumor components, which have magnificent value for tumors diagnosis and prognosis assessment. HNSCC is one of the most common malignant tumors in the world, with high metastasis rate, low survival rate and poor prognosis (55). Importantly, most HNSCCs have significant immune infiltration, and the close interaction between tumor cells and their microenvironment significantly impairs the development and progression of malignant tumors, therefore influences the prognosis of patients (56).

In this study, we identified TME related genes that contribute to HNSCC OS and conducted prediction of immunotherapeutic response using data from TCGA database. Specifically, by comparing lncRNA expression level of 546 TCGA-HNSC samples from 528 TCGA-HNSC patients, we then extracted 258 genes involved in prognoses and stromal/immune Scores and immunotherapeutic response prediction. We used the ESTIMATE algorithm to calculate immune and stromal scores in the TME. Therefore, we demonstrated that patients with low stroma risk score in line with a relatively higher survival probability, and low immune risk score corresponds a relatively higher survival probability.

In our study, the average TMB was 3.48 mutants per Mb from 508 HNSCC patients, which was similar to background TMB value in HNSCC (57). We verified top five mutation SMGs in TCGA HNSCC cohort were *TP53*, *TTN*, *FAT1*, *CDKN2A* and *MUC16*. Early studies verified that *TP53* is a ubiquitous tumor suppressor which is critically mutated in different malignant tumors, with about two-thirds of HNSCC accompanying mutations in exons (58,59), and these alterations were intimately associated with resistance to radiation and cisplatin-based chemotherapeutics (60,61). Many studies have revealed that *TP53* is one of the most frequently mutated genes in HNSCC (62-64). The *TTN* was included in old age specific clusters and pathway enrichment in HNSCC SMGs (65). Based on previous studies, *FAT1* plays different roles in different tissues or cancer types and its expression level is downregulated in HNSCC and other cancers, which has been considered as a tumor suppressor (66,67). The highest

mutation rate of *FAT1* was approximately 23% in HNSCC, ranking as the second most mutated gene after *TP53* (68,69). Several studies have shown that suppression of *CDKN2A* expression by methylation is involved in the development of HNSCC, which could be a potential diagnosis and prognosis biomarker for HNSCC (70-72). Although not much is known about the specific effects of *MUC16* in HNSCC, many studies have been verified that deregulated expression of *MUC16* was associated with several cancers, such as breast cancer (73), ovarian cancer (74), non-small-cell lung cancer (75) and pancreatic cancer (76).

Furthermore, both CIBERSORT and TIDE algorithm outcome agree that a high level of CD8[+] T cells was significantly corresponding to a preferable survival in HNSCC, which was in a line with that high CD8[+] intratumoural counts exhibited a remarkable association with relapse free survival in breast cancer and Triple-negative breast cancer (14,15). And this tendency is consistent with previous study that increased numbers of intraepithelial CD8[+] TIL was associated with favorable outcome in HNSCC (77). Besides, the CD8[+] T cells would experience a declination from stage I to stage IV. Though there are contradictory study point out, instead of CD8[+] T cell, CD4[+] T cells in cancer stroma are closely related to a favorable prognosis in human non-small cell lung cancers (78), We speculate that exogenous reactivation of CD8[+] T cells might be theoretically feasible to alleviate HNSCC patients suffering using rational immunotherapy strategies (17,79).

Over the recent years, the TME was characterized as a pivotal role in determining tumor progression and treatment outcomes. As an indispensable part of the TME, the tumor stroma greatly affects tumorigenesis, cancer progression, metastasis, and therapy resistance in various cancers (80). In our study, even though the survival difference between high immune score group and low immune score group was quite slight, the high immune score corresponding to a high survival probability (P=0.015). Moreover, there are a huge survival gap between high stroma score group and low stroma score group, high stroma score group has obvious better survival than low stroma score group as depicted in Kaplan-Meier OS analysis (P=0.00088).

Conclusions

In conclusion, we performed a comprehensive analysis of the TME in HNSCC using RNA-seq genomic data from TCGA database. Immune and stromal scores were calculated

Page 12 of 15

Li et al. Genes with prognostic value in HNSCC microenvironment

by ESTIMATE algorithm. The responses of ICB to each HNSCC patient were predicted by TIDE algorithm. Finally, we used CIBERSORT algorithm identified specific prognostic TIICs by quantifying the cellular composition of the immune response in HNSCC and its association to survival outcome, using the CIBERSORT algorithm. The results showed that high level of CD8+ T cells was significantly corresponding to a preferable survival and high level of CD4+ T cells was significantly associated with poor survival in HNSCC. Furthermore, high immune score and stroma score corresponded to a better survival probability in HNSCC. In addition, we verified that top five SMGs in TCGA HNSCC cohort were *TP53*, *TTN*, *FAT1*, *CDKN2A* and *MUC16*. Therefore, the expression level of CD8+ T cells and CD4+ T cells could be used as prognostic biomarkers for HSCNN patients in clinical applications. The expression levels of characteristic genes that used by CIBERSORT and ESTIMATE algorithms, or the expression levels of characteristic lncRNAs of LASSO Cox regression model in this study can be detected through high-throughput sequencing methods, and then the relative T cells proportion, immune score or stromal score can be calculated based on the algorithm models. Hence, the prognosis of patients can be stratified using the comprehensive immune approaches at the transcriptome level. Our study results consolidate the role TME in the progression of HNSCC and keep investigating the mechanisms of TME that preciously mediate tumorigenesis will augment the new targets for cancer therapy.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the REMARK reporting checklist. Available at https://atm.amegroups.com/article/view/10.21037/atm-22-6481/rc

Peer Review File: Available at https://atm.amegroups.com/article/view/10.21037/atm-22-6481/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-22-6481/coif). The authors report technical support from the Shanghai Tongshu Biotechnology Co., Ltd. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Mody MD, Rocco JW, Yom SS, et al. Head and neck cancer. Lancet 2021;398:2289-99.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019;69:7-34.
3. Bilal M, Raza SEA, Azam A, et al. Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. Lancet Digit Health 2021;3:e763-72.
4. Haddad RI, Shin DM. Recent advances in head and neck cancer. N Engl J Med 2008;359:1143-54.
5. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science 2018;359:1350-5.
6. Darvin P, Toor SM, Sasidharan Nair V, et al. Immune checkpoint inhibitors: recent progress and potential biomarkers. Exp Mol Med 2018;50:1-11.
7. Long B, Brém E, Koyfman A. Oncologic Emergencies: Immune-Based Cancer Therapies and Complications. West J Emerg Med 2020;21:566-80.
8. Xiong Y, Wang K, Zhou H, et al. Profiles of immune infiltration in colorectal cancer and their clinical significant: A gene expression-based study. Cancer Med

2018;7:4496-508.

9. Dos Santos LV, Abrahão CM, William WN Jr. Overcoming Resistance to Immune Checkpoint Inhibitors in Head and Neck Squamous Cell Carcinomas. Front Oncol 2021;11:596290.

10. Vacca P, Munari E, Tumino N, et al. Human natural killer cells and other innate lymphoid cells in cancer: Friends or foes? Immunol Lett 2018;201:14-9.

11. Raskov H, Orhan A, Christensen JP, et al. Cytotoxic CD8(+) T cells in cancer and cancer immunotherapy. Br J Cancer 2021;124:359-67.

12. Arasanz H, Gato-Cañas M, Zuazo M, et al. PD1 signal transduction pathways in T cells. Oncotarget 2017;8:51936-45.

13. Patsoukis N, Bardhan K, Chatterjee P, et al. PD-1 alters T-cell metabolic reprogramming by inhibiting glycolysis and promoting lipolysis and fatty acid oxidation. Nat Commun 2015;6:6692.

14. Matsumoto H, Thike AA, Li H, et al. Increased CD4 and CD8-positive T cell infiltrate signifies good prognosis in a subset of triple-negative breast cancer. Breast Cancer Res Treat 2016;156:237-47.

15. Rathore AS, Kumar S, Konwar R, et al. CD3+, CD4+ & CD8+ tumour infiltrating lymphocytes (TILs) are predictors of favourable survival outcome in infiltrating ductal carcinoma of breast. Indian J Med Res 2014;140:361-9.

16. Zhang Y, Li L, Zheng W, et al. CD8(+) T-cell exhaustion in the tumor microenvironment of head and neck squamous cell carcinoma determines poor prognosis. Ann Transl Med 2022;10:273.

17. He QF, Xu Y, Li J, et al. CD8+ T-cell exhaustion in cancer: mechanisms and new area for cancer immunotherapy. Brief Funct Genomics 2019;18:99-106.

18. Palucka AK, Coussens LM. The Basis of Oncoimmunology. Cell 2016;164:1233-47.

19. Xiao Y, Yu D. Tumor microenvironment as a therapeutic target in cancer. Pharmacol Ther 2021;221:107753.

20. Hinshaw DC, Shevde LA. The Tumor Microenvironment Innately Modulates Cancer Progression. Cancer Res 2019;79:4557-66.

21. Öhlund D, Elyada E, Tuveson D. Fibroblast heterogeneity in the cancer wound. J Exp Med 2014;211:1503-23.

22. Kalluri R. The biology and function of fibroblasts in cancer. Nat Rev Cancer 2016;16:582-98.

23. Elmusrati A, Wang J, Wang CY. Tumor microenvironment and immune evasion in head and neck squamous cell carcinoma. Int J Oral Sci 2021;13:24.

24. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. Cancer Cell 2012;21:309-22.

25. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646-74.

26. Wang G, Zhang M, Cheng M, et al. Tumor microenvironment in head and neck squamous cell carcinoma: Functions and regulatory mechanisms. Cancer Lett 2021;507:55-69.

27. Duray A, Demoulin S, Hubert P, et al. Immune suppression in head and neck cancers: a review. Clin Dev Immunol 2010;2010:701657.

28. Echarti A, Hecht M, Büttner-Herold M, et al. CD8+ and Regulatory T cells Differentiate Tumor Immune Phenotypes and Predict Survival in Locally Advanced Head and Neck Cancer. Cancers (Basel) 2019;11:1398.

29. Espinosa-Cotton M, Rodman Iii SN, Ross KA, et al. Interleukin-1 alpha increases anti-tumor efficacy of cetuximab in head and neck squamous cell carcinoma. J Immunother Cancer 2019;7:79.

30. Pitt JM, Vétizou M, Daillère R, et al. Resistance Mechanisms to Immune-Checkpoint Blockade in Cancer: Tumor-Intrinsic and -Extrinsic Factors. Immunity 2016;44:1255-69.

31. Echarri MJ, Lopez-Martin A, Hitt R. Targeted Therapy in Locally Advanced and Recurrent/Metastatic Head and Neck Squamous Cell Carcinoma (LA-R/M HNSCC). Cancers (Basel) 2016;8:27.

32. Huang Y, Liao J, Wu S, et al. Upregulated YTHDF1 associates with tumor immune microenvironment in head and neck squamous cell carcinomas. Transl Cancer Res 2022;11:3986-99.

33. Kunz-Schughart LA, Knuechel R. Tumor-associated fibroblasts (part I): Active stromal participants in tumor development and progression? Histol Histopathol 2002;17:599-621.

34. Wheeler SE, Shi H, Lin F, et al. Enhancement of head and neck squamous cell carcinoma proliferation, invasion, and metastasis by tumor-associated fibroblasts in preclinical models. Head Neck 2014;36:385-92.

35. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4:2612.

36. Gu CY, Dai B, Zhu Y, et al. The novel transcriptomic signature of angiogenesis predicts clinical outcome, tumor microenvironment and treatment response for prostate adenocarcinoma. Mol Med 2022;28:78.

Page 14 of 15

Li et al. Genes with prognostic value in HNSCC microenvironment

37. Zhang X, Tian C, Cheng J, et al. LTBP2 inhibits prostate cancer progression and metastasis via the PI3K/AKT signaling pathway. Exp Ther Med 2022;24:563.

38. Zheng S, Zou Y, Tang Y, et al. Landscape of cancer-associated fibroblasts identifies the secreted biglycan as a protumor and immunosuppressive factor in triple-negative breast cancer. Oncoimmunology 2022;11:2020984.

39. Zhang D, Lu W, Zhuo Z, et al. Construction of a breast cancer prognosis model based on alternative splicing and immune infiltration. Discov Oncol 2022;13:78.

40. Yang J, Chen H, Wang Y, et al. Development and validation of a robust necroptosis related classifier for colon adenocarcinoma. Front Genet 2022;13:965799.

41. Wang J, Akter R, Shahriar MF, et al. Cancer-Associated Stromal Fibroblast-Derived Transcriptomes Predict Poor Clinical Outcomes and Immunosuppression in Colon Cancer. Pathol Oncol Res 2022;28:1610350.

42. Xu M, Chang J, Wang W, et al. Classification of colon adenocarcinoma based on immunological characterizations: Implications for prognosis and immunotherapy. Front Immunol 2022;13:934083.

43. Ning YM, Lin K, Liu XP, et al. NAPSB as a predictive marker for prognosis and therapy associated with an immuno-hot tumor microenvironment in hepatocellular carcinoma. BMC Gastroenterol 2022;22:392.

44. Tee AE, Liu B, Song R, et al. The long noncoding RNA MALAT1 promotes tumor-driven angiogenesis by up-regulating pro-angiogenic gene expression. Oncotarget 2016;7:8663-75.

45. Fu WM, Lu YF, Hu BG, et al. Long noncoding RNA Hotair mediated angiogenesis in nasopharyngeal carcinoma by direct and indirect signaling pathways. Oncotarget 2016;7:4712-23.

46. Zhou C, Ye L, Jiang C, et al. Long noncoding RNA HOTAIR, a hypoxia-inducible factor-1 activated driver of malignancy, enhances hypoxic cancer cell proliferation, migration, and invasion in non-small cell lung cancer. Tumour Biol 2015;36:9179-88.

47. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 2010;26:1572-3.

48. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med 2017;9:34.

49. Zhou R, Zhang J, Zeng D, et al. Immune cell infiltration as a biomarker for the diagnosis and prognosis of stage I-III colon cancer. Cancer Immunol Immunother 2019;68:433-42.

50. Xu F, Zhang H, Chen J, et al. Immune signature of T follicular helper cells predicts clinical prognostic and therapeutic impact in lung squamous cell carcinoma. Int Immunopharmacol 2020;81:105932.

51. Hothorn T, Zeileis A. Generalized maximally selected statistics. Biometrics 2008;64:1263-9.

52. Chen F, Yang Y, Zhao Y, et al. Immune Infiltration Profiling in Nonsmall Cell Lung Cancer and Their Clinical Significance: Study Based on Gene Expression Measurements. DNA Cell Biol 2019;38:1387-401.

53. Rizvi H, Sanchez-Vega F, La K, et al. Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. J Clin Oncol 2018;36:633-41.

54. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100:57-70.

55. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.

56. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. Nat Rev Cancer 2009;9:239-52.

57. Xing L, Guo M, Zhang X, et al. A transcriptional metabolic gene-set based prognostic signature is associated with clinical and mutational features in head and neck squamous cell carcinoma. J Cancer Res Clin Oncol 2020;146:621-30.

58. Gasco M, Crook T. The p53 network in head and neck cancer. Oral Oncol 2003;39:222-31.

59. Morris LGT, Chandramohan R, West L, et al. The Molecular Landscape of Recurrent and Metastatic Head and Neck Cancers: Insights From a Precision Oncology Sequencing Platform. JAMA Oncol 2017;3:244-55.

60. Ohnishi K, Ota I, Takahashi A, et al. Transfection of mutant p53 gene depresses X-ray- or CDDP-induced apoptosis in a human squamous cell carcinoma of the head and neck. Apoptosis 2002;7:367-72.

61. Sharma K, Vu TT, Cook W, et al. p53-independent Noxa induction by cisplatin is regulated by ATF3/ATF4 in head and neck squamous cell carcinoma cells. Mol Oncol 2018;12:788-98.

62. Boyle JO, Hakim J, Koch W, et al. The incidence of p53 mutations increases with progression of head and neck cancer. Cancer Res 1993;53:4477-80.

63. Nishio M, Koshikawa T, Kuroishi T, et al. Prognostic significance of abnormal p53 accumulation in primary,

resected non-small-cell lung cancers. J Clin Oncol 1996;14:497-502.

64. Alsner J, Sørensen SB, Overgaard J. TP53 mutation is related to poor prognosis after radiotherapy, but not surgery, in squamous cell carcinoma of the head and neck. Radiother Oncol 2001;59:179-85.

65. Meucci S, Keilholz U, Tinhofer I, et al. Mutational load and mutational patterns in relation to age in head and neck cancer. Oncotarget 2016;7:69188-69199.

66. Peng Z, Gong Y, Liang X. Role of FAT1 in health and disease. Oncol Lett 2021;21:398.

67. Chen ZG, Saba NF, Teng Y. The diverse functions of FAT1 in cancer progression: good, bad, or ugly? J Exp Clin Cancer Res 2022;41:248.

68. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 2015;517:576-82.

69. Campbell JD, Yau C, Bowlby R, et al. Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas. Cell Rep 2018;23:194-212.e6.

70. Herman JG, Merlo A, Mao L, et al. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. Cancer Res 1995;55:4525-30.

71. El-Naggar AK, Lai S, Clayman G, et al. Methylation, a major mechanism of p16/CDKN2 gene inactivation in head and neck squamous carcinoma. Am J Pathol 1997;151:1767-74.

72. Zhou C, Shen Z, Ye D, et al. The Association and Clinical Significance of CDKN2A Promoter Methylation in Head and Neck Squamous Cell Carcinoma: a Meta-Analysis. Cell Physiol Biochem 2018;50:868-82.

73. Lakshmanan I, Ponnusamy MP, Das S, et al. MUC16 induced rapid G2/M transition via interactions with JAK2 for increased proliferation and anti-apoptosis in breast cancer cells. Oncogene 2012;31:805-17.

74. Yin BW, Lloyd KO. Molecular cloning of the CA125 ovarian cancer antigen: identification as a new mucin, MUC16. J Biol Chem 2001;276:27371-5.

75. Cedrés S, Nuñez I, Longo M, et al. Serum tumor markers CEA, CYFRA21-1, and CA-125 are associated with worse prognosis in advanced non-small-cell lung cancer (NSCLC). Clin Lung Cancer 2011;12:172-9.

76. Haridas D, Chakraborty S, Ponnusamy MP, et al. Pathobiological implications of MUC16 expression in pancreatic cancer. PLoS One 2011;6:e26839.

77. Pretscher D, Distel LV, Grabenbauer GG, et al. Distribution of immune cells in head and neck cancer: CD8+ T-cells and CD20+ B-cells in metastatic lymph nodes are associated with favourable outcome in patients with oro- and hypopharyngeal carcinoma. BMC Cancer 2009;9:292.

78. Wakabayashi O, Yamazaki K, Oizumi S, et al. CD4+ T cells in cancer stroma, not CD8+ T cells in cancer cell nests, are associated with favorable prognosis in human non-small cell lung cancers. Cancer Sci 2003;94:1003-9.

79. Farhood B, Najafi M, Mortezaee K. CD8(+) cytotoxic T lymphocytes in cancer immunotherapy: A review. J Cell Physiol 2019;234:8509-21.

80. Valkenburg KC, de Groot AE, Pienta KJ. Targeting the tumour stroma to improve cancer therapy. Nat Rev Clin Oncol 2018;15:366-81.

Table S1 Clinicopathological information for the TCGA data sets

| Characteristic	N
Median age in years [range]	61 [19–90]
Male/Female	386/142
Median survival (months) [range]	24 [1–213]
Race	
White	452
Black	48
Asian	11
Never-drinker/Ever drinker	176/352
Tumor stage	
Stage I–II	102
Stage III	82
Stage IVA–IVB	269
Tumor site	
Oral cavity	Tongue (n=156) + Gum (n=11) + Mouth (n=79) + Mandible (n=1) + Cheek Mucosa (n=19) =266
Oropharynx	Larynx (n=116) + Supraglottis (n=1) 117
Tonsil	49
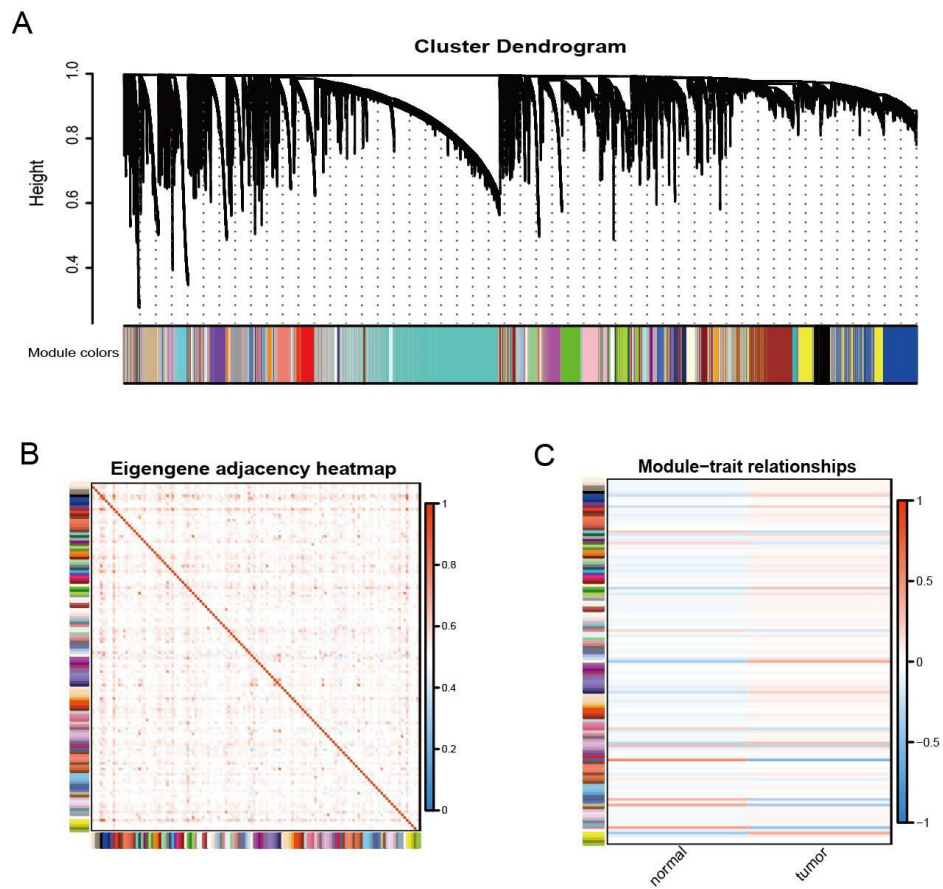
Figure S1 Identification of modules associated with HNSCC samples. (A) The cluster dendrogram and module colors of HNSCC samples. (B) Eigengene adjacency heatmap of HNSCC samples. (C) Heatmap plot of genes differentially expressed in tumor and normal tissues.