Computational and Structural Biotechnology Journal 20 (2022) 4786–4805

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Identification of new head and neck squamous cell carcinoma subtypes and development of a novel score system (PGSscore) based on variations in pathway activity between tumor and adjacent non-tumor samples

Yufan Zhang [a,1], Ying Liu [b,c,1], Junfei Huang [a,1], Zhiqi Hu [a,*], Yong Miao [a,*]

[a] Department of Plastic and Aesthetic Surgery, Nanfang Hospital of Southern Medical University, Guangzhou 510515, Guangdong Province, China
[b] Guangzhou Blood Centre, Guangzhou 510091, Guangdong, China
[c] Nanfang Hospital, Southern Medical University, Guangzhou 510515, Guangdong, China

ARTICLE INFO

Article history:
Received 29 March 2022
Received in revised form 20 August 2022
Accepted 27 August 2022
Available online 1 September 2022

Keywords:
Head and neck squamous cell carcinoma
Subtype
Gene set
PGSscore
Immunotherapy
Prognosis

ABSTRACT

The influence of adjacent non-tumor tissue characteristics on patient outcomes in head and neck squamous cell carcinoma (HNSC) remains unclear. In HNSC, subtype identification is generally tumor-based, and most prognosis-related studies focus on a single gene set. Here, we performed Gene Set Variation Analysis to comprehensively evaluate variations in diverse gene sets in tumor and non-tumor samples and converted a gene-centric matrix into a pathway-centric model. Three different prognostic subtypes correlated with previously identified subgroups, clinicopathologic features, risk factors, and tumor microenvironment (TME) were identified using the non-negative matrix factorization method. We also screened 2 and 11 gene sets from nontumor and tumor tissues, respectively based on representative gene sets. Interestingly, genes from nontumor gene sets were associated with serotonin secretion and P2Y receptors, while genes from tumor gene sets were associated with immunity and inflammation. The PGSscore was constructed to predict outcomes and immunotherapy responses. Low- and high-PGSscore groups showed significant differences in clinicopathological characteristics and TME. Our analysis indicates that the non-tumor tissue is indispensable for prognosis. PGSscore provides new avenue to evaluate overall survival and immunotherapy responses, and our method based on pathway-centric models can be extended to other diseases.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Head and neck squamous cell carcinoma (HNSC) is the sixth most common cancer worldwide, with 1.08 million new cases every year [1,2]. Surgery, radiation, chemotherapy, and immunotherapy are the key treatments for HNSC [2]; however, they are associated with disadvantages, such as disease recurrence, cervical node metastasis, chemotherapeutic drug resistance, and ineffective immunotherapy, all of which result in increased deaths from advanced HNSC. Therefore, it is important to identify novel HNSC subtypes and associated pathways and signatures to better predict patient overall survival and enhance outcomes. Such a strategy would help guide the treatment and improve immunotherapy responses.

Previous studies on tumor subtypes have generated some important results, and researchers have categorized HNSC into subtypes mainly based on gene expression profiling of the tumor tissue. In addition, there are several other criteria for subtype stratification, including copy number profiling and status of HPV16 DNA and RNA [3–7]. However, these studies did not take the adjacent non-tumor samples into account during the identification of molecular subtypes, although some previous research confirmed that the role of the non-tumor tissue is critical in tumor development. Increased epidermal cell proliferation in response to altered pathways in the parabasal layers adjacent to oral squamous cell carcinoma tissue may trigger early oncogenic events that favor invasion [8]. Further, the loss of E-cadherin in adjacent non-tumor epithelium is related to tumor invasiveness in oral squamous cell carcinoma [9]. Moreover, inflammatory signals and immune components in non-tumor samples may play an important role in facilitating tumor development [10]. Therefore, insights

* Corresponding authors.
E-mail addresses: huzhiqidr163@i.smu.edu.cn (Z. Hu), miaoyong123@i.smu.edu.cn (Y. Miao).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2022.08.057
2001-0370/© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

from adjacent non-tumor samples should be considered when determining the HNSC subtype.

Most previous studies have largely focused on one pathway, gene set, lncRNA set, or a single molecule that is considered to be associated with tumor development. For example, pyroptosis-related genes [11], pyroptosis-related lncRNA [12], ferroptosis-related genes [13], m6A-modified lncRNA [14], immune-related genes [15], aging-related genes [16], hypoxia-related genes [17], autophagy-related genes [18], inflammatory pathways [19], fatty acid metabolism pathway [20], Fascin Actin-bundling Protein 1 [21], and dual-specificity phosphatase 1 [22] have been used as a basis to predict HNSC outcomes. Hence, these publications tend to choose single emerging or popular gene sets to establish a molecular signature or model to predict outcomes, while ignoring other gene sets. Such studies may lead to superfluous information, as it would be hard to judge which gene/pathway/molecule gives the best results. Therefore, an approach of analyzing and integrating the data of abundant gene sets from tumor and adjacent non-tumor samples would provide a comprehensive profile of HNSC in an objective manner.

In this study, to comprehensively reveal subtle and pronounced variations in pathway activity in HNSCs, we transformed the gene expression matrix into a gene set enrichment score (ES) matrix, incorporating HNSC and adjacent non-tumor samples for quantitative analysis. We also determined the changes in pathway activity between tumor and adjacent non-tumor samples to identify novel HNSC subtypes and their prognostic correlations. We further explored the correlations of the subtypes with the levels of tumor microenvironment (TME) components, previously defined subgroups, clinicopathologic features, and risk factors. Based on these analyses, prognostic gene sets in HNSC and adjacent normal tissue were defined, enabling establishment of a new score (PGSscore) to predict overall survival and the immunotherapy response. Furthermore, we described the immune cell landscape, and the correlation between PGSscore and clinicopathological features, the TME, and other features. A nomogram was established to more precisely predict the overall survival rates of HNSC patients based on PGSscore. Our system can be used to detect changes in pathway activity in tumors and adjacent non-tumor tissues, offering a new method to identify subtypes and predict overall survival and immunotherapy responses in HNSC.

2. Materials and methods

2.1. Data collection

Public HNSC gene expression datasets with complete clinicopathological data for tumor and adjacent non-tumor tissues were downloaded from TCGA (https://portal.gdc.cancer.gov/) and the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/), i.e., TCGA head and neck squamous cell carcinoma, GSE41613, and GSE65858. For TCGA dataset, the fragments per kilobase per million (FPKM) values were transformed into transcripts per kilobase million (TPM). The GSE41613 and GSE65858 datasets were merged after background adjustment and quantitative normalization using the raw "cel" files of these datasets, and the "Affy" and "simpleaffy" R packages [23]). Subsequently, the SVA R package was used to eliminate batch effects in the merged datasets using the "ComBat" algorithm. Finally, 18,029 gene sets were downloaded from the Molecular Signatures Database v7.4 (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp. These datasets contained hallmark gene sets (50genesets), canonical pathways (CP; 2922 gene sets), gene ontology (GO; 10,185 gene sets), and the ImmuneSigDB subset of the C7 dataset [4]. Hallmark gene sets represent specific well-defined biological states or processes and display coherent expression [24]; CP represent biological processes, and were compiled by domain experts from BIOCARTA (292genesets), Kyoto Encyclopedia of Genes and Genomes (KEGG; 186 gene sets), PID (196genesets), REACTOME [1], and WikiPathways (615genesets). GO datasets, contained biological process [7], cellular component (996genesets), and molecular function [1] ontologies. The ImmuneSigDB subset of the C7 database represents chemical and genetic perturbations of the immune system and was generated using manual curation of published studies on human and mouse immunology [25]. The data of four molecular subtypes and six immune subtypes listed in the Pan-Cancer Atlas were downloaded from the Supplementary Materials of two previous studies [26–27].

2.2. Gene set variation analysis

The GSVA R package was used to condense the gene expression profiles into a gene set matrix with ESs using the GSVA algorithm in HNSC and adjacent non-tumor samples to investigate the variation in several gene sets or pathways [28]. Adjusted $p < 0.01$ was considered significant.

2.3. Cancer subtype analysis

An integrative R package called "CancerSubtypes," which includes methods such as non-negative matrix factorization (NMF) method, provides useful feature selection, subtype identification, and validation methods [29]. The NMF method was used to identify the optimal cluster number and to classify patients based on the calculated gene set matrix. The assignment accuracy and fitness were validated using the silhouette width index, ranging between −1 and 1. A silhouette width index closer to 1 indicates well-defined patterns or subtypes.

2.4. Differentially enriched gene sets associated with the three subtypes

The "limma" R package was used to obtain differentially enriched gene sets among the three subtypes and differentially expressed genes (DEGs) among the three PGS clusters. $P < 0.01$ was used to screen the differentially enriched gene sets and DEGs. Representative gene sets and overlapping genes were identified by selecting the intersection area of a Venn plot.

2.5. Analysis of tumor microenvironment

To identify the levels of tumor microenvironment, the CIBERSORT R package was used to evaluate the ImmuneScore, StromalScore, and ESTIMATE score for each patient.

2.6. Identification of prognostic gene sets

Twenty gene sets were identified using univariate Cox analysis ($p < 0.05$) from 48 representative gene sets. To avoid overfitting, 13 prognostic gene sets were identified using least absolute shrinkage and selection operator (LASSO) penalties based on the gene set ES matrix [30].

2.7. Unsupervised clustering of prognostic gene sets

Unsupervised clustering analysis was performed to identify patients based on 13 prognostic gene sets and their ESs using the "ConsensusClusterPlus" R package. The optimal number of clusters and their stability were also determined. 1000 repetitions were performed to confirm cluster stability [31].

2.8. Principal component analysis

The principal component analysis (PCA) R package was used to reduce the dimensionality of the gene set matrix and to identify the discriminative power of the PGS clusters.

2.9. Kaplan-Meier survival analysis

Kaplan-Meier survival analysis was used to evaluate overall survival for the three different subtypes, PGS clusters, low- and high-ES groups for each prognostic gene set, and low- and high-PGSscore. This analysis was performed using the "survminer" and "survival" R packages.

2.10. Functional enrichment analysis

The R package ClusterProfiler was used to conduct GO enrichment analysis and KEGG pathway analysis [32]. This analysis provided a functional interpretation of the overlapping DEGs in the PGS clusters, and the genes extracted from the gene sets in tumor (T gene sets) and gene sets in non-tumor (N gene sets).

2.11. Module and hub gene analysis

Protein-protein interaction (PPI) networks were analyzed and constructed using STRING (https://string-db.org/). Then, the PPI network was imported into Cytoscape version 3.7.2 and analyzed using the Molecular Complex Detection (MCODE) and CytoHubba plugins. MCODE identified the modules of the PPI network using the following parameters: degree cutoff, 2; node score cutoff, 0.2; K core, 2; maximum depth, 100. The hub genes of each module were obtained using the "degree" topological algorithm of CytoHubba.

2.12. Construction of the PGSscore

TCGA data set was randomized into a training set (50 % of the dataset) and a testing dataset (50 % of the dataset). Then, we used univariate Cox analysis to identify the prognosis-associated genes. The identified prognostic genes were screened using LASSO penalties. Finally, the correlation coefficients for 13 genes were obtained to calculate the PGSscore [PGSscore = coefficient (gene 1) × expression (gene 1) + coefficient (gene 2) × expression (gene 2) + coefficient (gene n) × expression (gene n)]. The median PGSscore was used as the cutoff value. The training set, testing set, GEO cohort, and the entire TCGA set were divided into high- and low-PGSscore groups based on the PGSscore cutoff. Then, the survminer R package was used to analyze overall survival. Receiver operating characteristic (ROC) curve analysis was performed using the "survivalROC" R package to estimate the predictive ability of the PGSscore.

2.13. Correlation between PGSscore and clinicopathologic parameters

Univariate Cox regression was used to identify prognostic genes. Multivariate Cox regression was performed to investigate whether the PGSscore was an independent prognostic factor in the training and testing sets. The "pheatmap" R package was used to visualize relative gene expression for the PGSscore in each patient and the relationship between the PGSscore and clinicopathologic characteristics. The "ggalluvial" R package was used to construct a ggalluvial plot to visualize the subtype, PGS cluster, PGSscore, and survival status in the entire TCGA set.

2.14. Estimation of the immune cell landscape

The "immunedeconv" R package, which contains the XCELL, TIMER, QUANTISEQ, MCPcounter, EPIC, CIBERSORT-ABS, and CIBERSORT R functions, was used to investigate the immune cell landscape among TCGA samples. Spearman analysis was used to evaluate the correlation between the PGSscore and immune cell profiles. The correlation coefficients are displayed using a lollipop diagram. Significance was set at $p < 0.05$.

2.15. Immunophenoscore and tumor immune dysfunction and exclusion scores

Individual immunophenoscore (IPS) data for HNSC patients were downloaded from the Cancer Immunome Atlas (https://tcia.at/home). The four immunogenicity-determining genes were effector cells, major histocompatibility complex molecules, immunosuppressive cells, and immunomodulators. The IPS for each individual was calculated using machine learning [33]; the IPS values ranged from 0 to 10 and were calculated according to z-scores representing gene expression. The IPS value was positively correlated with tumor immunogenicity [34]. IPS analysis was also performed to estimate the response to immune checkpoint inhibitors (ICIs) between the low- and high-PGSscore groups. Tumor immune dysfunction and exclusion (TIDE) data were downloaded from https://tide.dfci.harvard.edu, and were used to compare the ICI responses between low- and high-PGSscore groups [35].

2.16. Data collection and processing of immune-checkpoint blockade cohorts

Owing to the absence of data for an HNSC cohort treated with immune-checkpoint blockade, we referred to previous research methods [36]. Two cohorts with immune-checkpoint blockade were selected for further analysis, i.e., human renal cell carcinoma samples treated with anti-PD-1 (nivolumab) (cohort GSE67501 from GEO) [37] and advanced urothelial carcinoma treated with anti-PD-1 (atezolizumab) (cohort IMvigor210, https://research-pub.gene.com/IMvigor210CoreBiologies). For the GSE67501 cohort, the FPKM data were transformed into TPM values after standardization using the limma R package. The "DEseq2" R package was used to normalize the raw data of the IMvigor210 cohort and to convert the count values into TPM values.

2.17. Nomogram establishment and interpretation

The "rms" R package was used to combine clinicopathologic characteristics and PGSscores to construct a nomogram that predicts the 1-, 3-, and 5-year overall survival. In the nomogram, each factor was matched with a specific score. The sum of the scores across all factors for each patient was the total score. Higher scores indicate greater risk [38]. The "ggDCA" R package was used to evaluate the clinical value of the nomogram by analyzing the net benefit at different threshold probabilities using the decision curve analysis (DCA) algorithm [39]. The concordance index (C-index) ranged from 0.5 to 1 and was used to assess the discriminative capacity. Larger C-index values suggest better prediction accuracy [38]. Meanwhile, ROC curve analysis was performed using the survivalROC R package to estimate the predictive ability of the nomogram. Finally, the consistency between the predicted and actual survival outcomes was explored using calibration analysis [40].

2.18. Statistical analysis

R version 4.1.0 (https://www.R-project.org) was used for statistical analysis. The differential analysis between two groups was calculated using Wilcoxon tests, while comparisons of three or more groups were analyzed using Kruskal-Wallis or one-way ANOVA. Differences in subtypes, clinicopathologic features, and risk factors were calculated using chi-square tests. Spearman's correlation was used to perform correlation tests. Survival analysis was performed using log-rank and Kaplan-Meier tests. Significance was set at $p < 0.05$.

3. Results

3.1. Subtype identification based on activity variations in HNSC and adjacent non-tumor samples in TCGA dataset

The detailed workflow for the identification process of gene set matrix, subtype, prognostic gene sets, PGS cluster and PGSscore is displayed in Fig. 1. First, the gene expression matrix was converted into a gene set matrix with ESs using the GSVA algorithm to comprehensively investigate variations in gene sets in HNSC and adjacent non-tumor samples, which are depicted as a heatmap, 13,479 of which were detected in TCGA dataset (Fig. 2A). We then used the CancerSubtypes R package, which integrates the current computational biology methods for cancer subtype identification and provides a standardized framework for cancer subtype analysis [29]. The results revealed three definite cancer subtypes (k = 3) based on the ESs of HNSC and adjacent nontumor tissues using the NMF method (Fig. 2B). The three subtypes were designated "subtypes 1–3" and were confirmed to be distinct (Fig. 2C). Survival analysis revealed substantial prognostic differences among the three subtypes ($p = 0.000128$). Subtype 3 was clearly associated with survival advantages, while subtype 1 had the poorest prognosis (Fig. 2D). The heatmap also showed the distribution of patients in the three subtypes (Fig. 2E). Finally, the average silhouette width (0.96) revealed an excellent match between an HNSC patient and its identified subtype. (Fig. 2F).

3.2. Correlation of our identified subtypes with clinicopathologic features, risk factors, and other subtypes in TCGA dataset

To identify correlations between the three subtypes, clinicopathologic characteristics, and risk factors, we performed chi-squared tests. Our results showed that disease stage, T stage, HPV status, and the frequency of alcohol consumption per week had significantly different distributions among the three subtypes (chi-squared test, $p < 0.05$) (Fig. 3A–D). More patients with stage I-II disease, T0-2 disease, positive HPV status, and lower alcohol consumption per week were grouped into subtype 3 than in the other subtypes, whereas patients with stage III-IV disease, T3-4 disease, negative HPV status, and high alcohol consumption per week were grouped into subtype 1, indicating that the subtypes can predict the clinicopathologic index and risk factors of HNSC in patients.

Next, we explored the correlation of our identified subtypes with HNSC molecular subgroups. HNSC molecular subgroups had significantly different distributions among the three subtypes (chi-squared test, $p = 0.001$). Specifically, 42 % of patients in subtype 3 were assigned to atypical subgroups, whereas only 13 % and 12 % patients in subtype 1 and 2, respectively, belonged to atypical subgroups (Fig. 3E). After integration with immune subtypes, approximately 98 % of patients belonged to the C1 (wound healing, 25 %) or *C*2 (IFN-c dominant, 73 %) immune subtypes, with only 2 % of patients associated with other immune subtypes. Furthermore, a previous study indicated that the *C*2 subtype was characterized by the highest lymphocyte and $CD8^+$ T cell infiltrates, along with a robust anti-tumor immune response [26]. As expected, we detected a higher proportion of patients in subtype 3 (78 %) belonging to the *C*2 immune subtype.

3.3. Differences between the three subtypes with respect to representative gene sets and TME components in TCGA dataset

To further explore the representative potential biological functions of the three HNSC subtypes based on gene set variations, we identified differentially enriched gene sets among the three subtypes ($p < 0.01$, $|\log_2 \text{foldchange}| > 1$). Forty-eight gene sets were acquired from the intersection of the differentially enriched gene sets from the three groups (Fig. 4A). Differences between the three subtypes in terms of the ESs of representative gene sets and various clinicopathological features are displayed in Fig. 4B. The 48 gene sets had the highest ESs in subtype 3 and the lowest ESs in subtype 1, suggesting that these gene sets could be related to disease prognosis. To investigate the enriched immune cells and stromal cells associated with TME, we conducted CIBERSORT and ESTIMATE analysis of the three different subtypes. As expected, there were considerable differences in the immune, stromal, and ESTIMATE scores among the three subtypes (Fig. 4C, D, and E). Subtype 3 had the highest scores, while subtype 1 had the lowest scores, which also explains why subtype 3 had the best survival advantages and subtype 1 had the poorest prognosis.

3.4. Thirteen prognostic gene sets identified from HNSC and non-tumor tissues in TCGA dataset

Next, we identified prognostic gene sets from the 48 representative gene sets in tumor and adjacent non-tumor tissue samples using univariate Cox (Fig. S1A) and LASSO methods and obtained 13 prognostic gene sets (Fig. S1B, C). The hazard ratio, confidence interval (CI), *p*-values, and related links are shown in Table 1. Among the 13 prognostic gene sets, 11 were in the tumor samples (T gene sets). Two gene sets were identified in adjacent non-tumor tissue samples (N gene sets). Among the three subtypes, the ESs of the 13 gene sets were statistically different ($p < 0.001$). Subtype 3 had the highest ESs and subtype 1 had the lowest ESs for all 13 gene sets (Fig. S1D). Based on the median ESs for each gene set, all the patients were classified into high- and low-ES groups. Meanwhile, prognostic analysis suggested that groups with high ESs have considerable survival advantages ($p < 0.05$) (Fig. 5A, B).

3.5. Comprehensive analysis of PGS clusters based on prognostic gene sets in TCGA dataset

Initially, a network of the 13 prognostic gene sets that comprehensively displayed the correlations and prognostic significance of each gene set obtained from the HNSC and non-tumor tissues was generated. This network revealed that all 13 gene sets were positively correlated with each other ($p < 0.0001$) (Fig. 6A). We used the ConsensusClusterPlus R package to further classify the patients into three distinct clusters (PGS clusters A-C; Fig. 6B and Fig. S2A–C). There were 210 patients in PGS cluster A, 141 in PGS cluster B, and 148 in PGS cluster C. Next, the classification of the PGS clusters was further verified using PCA (Fig. 6C). There were substantial differences in prognostic value among the three PGS clusters, and PGS cluster C displayed better overall survival than that by the other PGS clusters (log-rank test, $p = 0.008$) (Fig. 6D). The relationships between the three PGS clusters indicated the relative ESs of the 13 prognostic gene sets (Fig. 6E). To further investigate the potential biological factors influencing each PGS cluster, we identified 245 overlapping genes from the three DEG groups (Fig. S2D). To demonstrate how the 245 overlapping genes contribute to disease

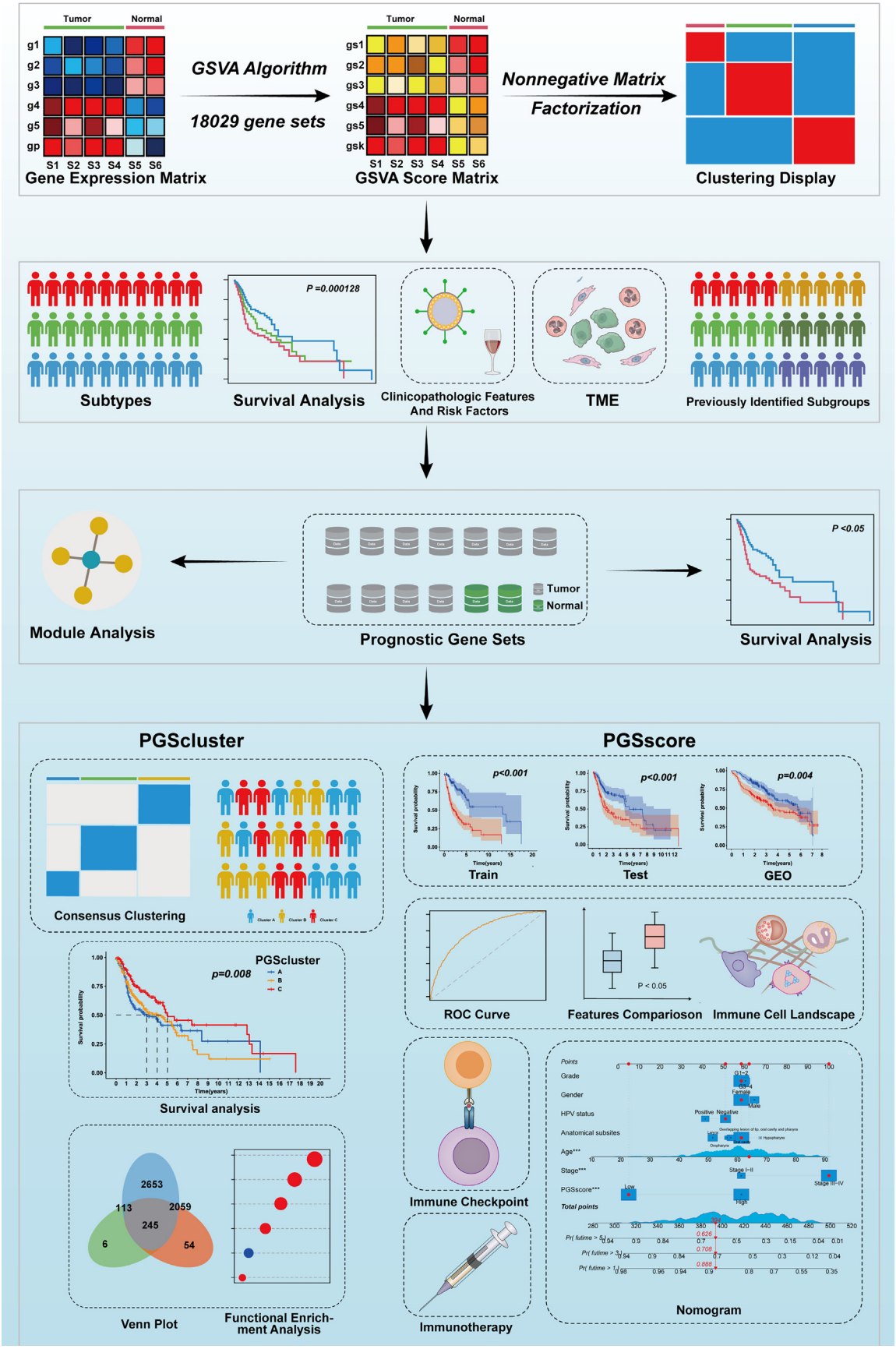


Fig. 1. Schematic diagram of the study. The identification process of the gene set matrix, subtypes, prognostic gene sets, PGS cluster, and PGSscore.

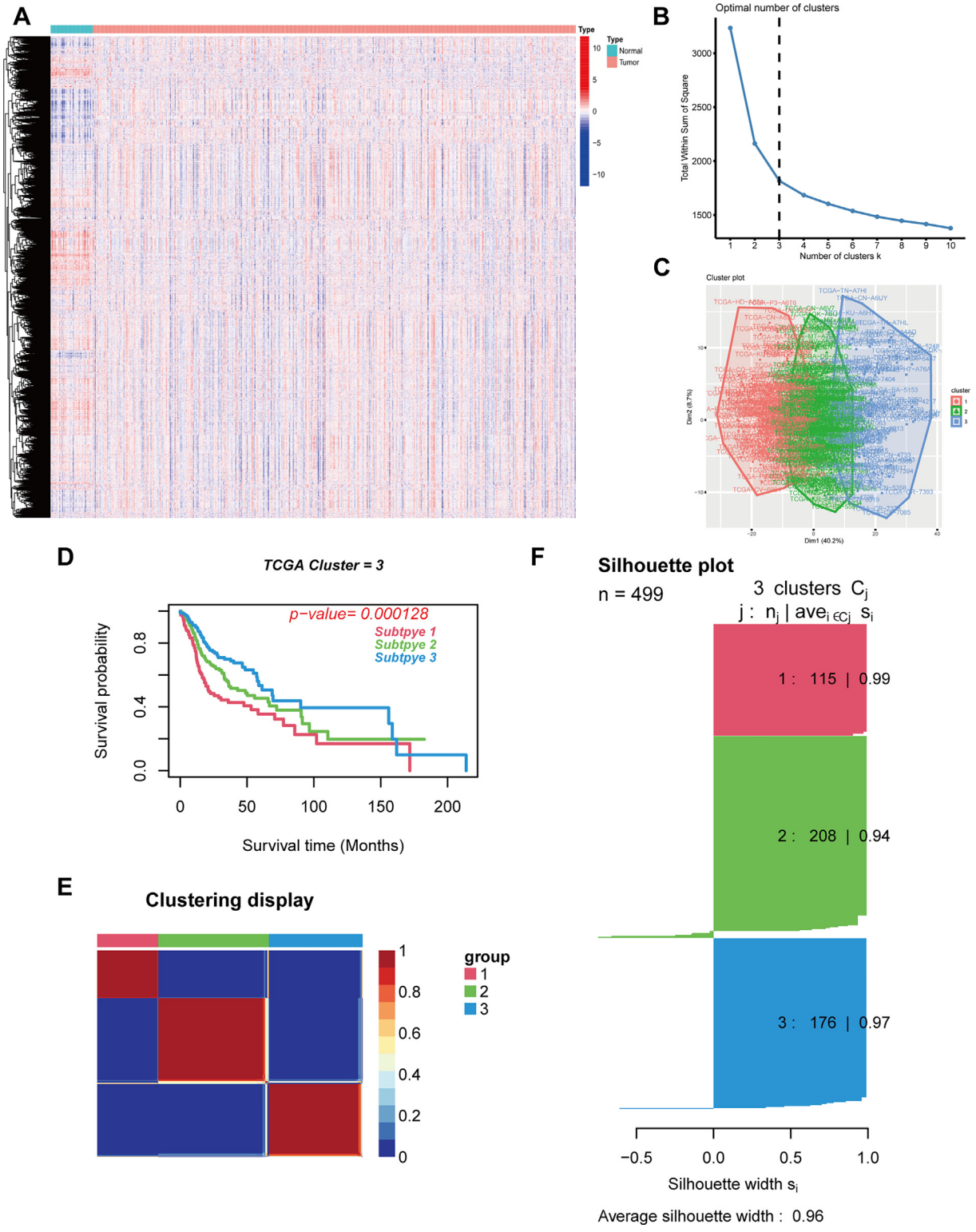


Fig. 2. Identification of subtypes based on variations in gene set activity in HNSC and adjacent non-tumor samples in TCGA dataset. (A) A heatmap showing the ESs of 18,029 in HNSC and adjacent non-tumor sample gene sets calculated using the GSVA method. (B) The optimal number of clusters (K) that identified using the NMF method. (C) Visualization of the three subtypes. The dots represent each patient. (D) Kaplan-Meier survival analysis revealed substantial differences among the three subtypes (log-rank test, p = 0.000128). (E) A heatmap defining the three subtypes (k = 3) and their correlation area in HNSC samples. (F) A silhouette width plot showing the subtype accuracy.

prognosis, we used clusterProfiler to perform GO enrichment and KEGG pathway analyses. The 245 overlapping genes were mostly correlated with the immune response-activating cell surface receptor signaling pathway and immune response-activating signal transduction (BP terms), T cell receptor complex (CC terms), and immune receptor activity (MF terms), which confirms that the distinction between the PGS clusters was based on immunity (Fig. S2E). KEGG analysis showed that cytokine-cytokine receptor interaction, chemokine signaling pathway, and cell adhesion molecules were mainly enriched in the gene sets (Fig. S2F).

Fig. 3. Correlation of our identified subtypes with clinicopathologic features, risk factors and other subtypes in TCGA dataset. (A) Distribution of patients with stage I-II and stage III-IV tumors between our identified subtypes. (B) Distribution of patients with T0-2 and T3-4 tumors between our identified subtypes. (C) Distribution of patients with negative and positive HPV status between our identified subtypes. (D) Distribution of patients based on frequency of alcohol consumption per week between our identified subtypes. (E) Distribution of four previously identified HNSC molecular subtypes (basal, mesenchymal, atypical, classic) across our identified subtypes. (F) Distribution of six immune subtypes (C1: wound healing, C2: IFN-γ dominant, C3: inflammatory, C4: lymphocyte depleted, C5: immunologically quiet, and C6: TGF-β dominant) across our identified subtypes.

3.6. Functional enrichment analysis and module analysis of genes from the prognostic gene sets in TCGA dataset

We investigated the potential functions of the genes in the 13 prognostic gene sets obtained from HNSC and adjacent non-tumor tissue samples. The genes derived from the T gene sets (Table S1) were mostly associated with positive regulation of leukocyte activation, positive regulation of lymphocyte activation, and B cell activation (BP), immunoglobulin complex (CC), and immunoglobulin receptor binding (MF) (Fig. 7A). KEGG analysis showed enrichment in cytokine-cytokine receptor interactions and Th17 cell differentiation (Fig. 7B). These results indicate that the loss of immune cell activation and regulation may promote tumor development. Genes from the N gene sets (Table S2) were associated with monoamine transport and serotonin secretion (BP), integral component of presynaptic membrane (CC), and purinergic nucleotide receptor activity (MF) (Fig. 7C). KEGG analysis showed enrichment in serotonergic synapses (Fig. 7D). Next, we constructed a protein–protein interaction network (Figure S3), which was used in further module analysis. There were three clusters from the T gene sets (Fig. 7E, F, and G) and two clusters from the N gene sets (Fig. 7H and I). Meanwhile, the hub genes of each

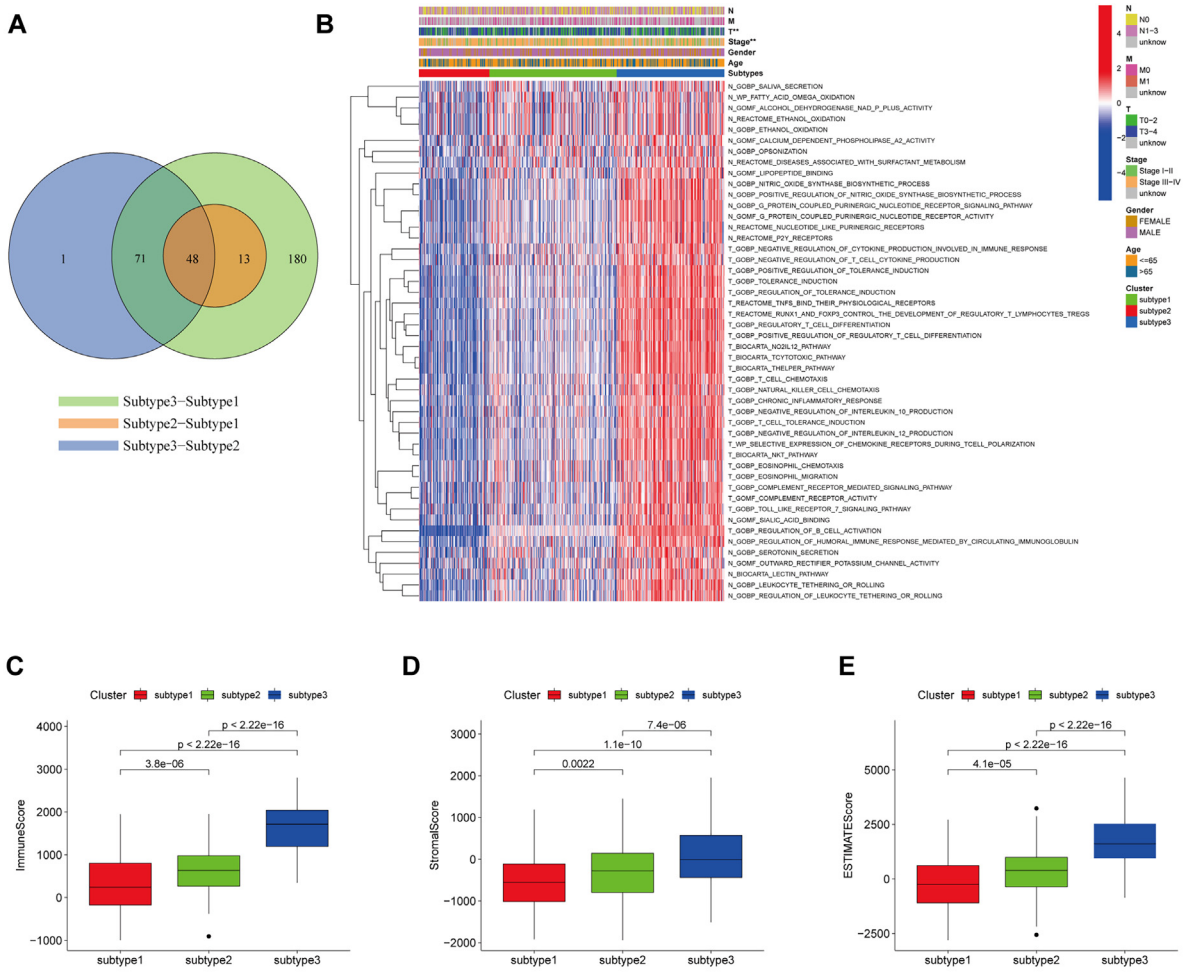


Fig. 4. Distinct properties of the three subtypes with respect to representative gene sets, TME, and immune cell infiltration in TCGA dataset. (A) Venn analysis identified overlapping representative gene sets from differentially enriched gene sets of subtypes 1 and 2, 2 and 3, and 1 and 3. (B) Heatmap showing the relative ESs of 48 representative gene sets of the three subtypes. Red represents high ESs and blue represents low ESs. Subtype, age, gender, stage, T, N, and M scores were used as sample annotations. (C–E) Correlation of the three subtypes with immune, ESTIMATE, and stromal scores. The line in the boxplot represents the median value. The statistical differences among the three subtypes were analyzed using one-way ANOVA. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cluster were obtained by selecting the highest degree value using CYTOHUBBA. In terms of the T gene sets, the hub genes from clusters 1, 2, and 3 were CD22, CD27, and CD247, respectively. With respect to the N gene sets, MAOB and P2RY6 were identified as hub genes for clusters 1 and 2, respectively.

3.7. Establishment and validation of the PGSscore for predicting patient outcomes

We next constructed a scoring system to predict the outcome for each HNSC patient. Initially, a total of 381 genes were extracted from the 13 prognostic gene sets. Univariate Cox regression analysis was performed to determine the prognostic genes from the 318 genes of TCGA dataset; finally, we obtained 72 prognostic genes that were significantly associated with overall survival (Fig. 8A). The heatmap in Fig. 8B shows the expression of the 72 genes in the HNSC and normal samples. Next, the entire TCGA set was randomly divided into a training set and a testing set. LASSO Cox regression analysis was performed to identify the optimal genes among the 72 and their coefficients in the training set (Fig. S4A and B). These results revealed 13 genes that were appropriate for constructing a prognostic model, which we termed the PGSscore. The correlation coefficients are listed in Table S3. Based on the median PGSscore, the training set, testing set, entire TCGA dataset, and GEO cohort (the integration of GSE41613 and GSE65858) were classified into high- and low-PGSscore groups. Fig. 8C and Fig. S4C–E show the distribution of patients in the training and testing sets according to their PGSscore, survival status, and survival time. In the training set, 56 % of patients died in the high-PGSscore groups and 27 % in the low-PGSscore group (Fig. 8D). We confirmed that the PGSscores of dead patients was significantly higher than that of living patients in the training set ($p < 0.001$; Fig. 8E). Survival analysis illustrated that the overall survival of patients with a high PGSscore was lower than that of patients with a low PGSscore ($p < 0.01$) in the training set (Fig. 8F), testing set (Fig. 8G), and GEO cohort (Fig. 8H). ROC curve analysis indicated that the PGSscore predicts the 1-, 3-, and 5-year survival rates in patients in the training set (Fig. 8I) and testing set (Fig. 8J), and the 1-year prognosis in the GEO cohort (Fig. 8K).

3.8. Identification of PGSscore as an independent prognostic factor correlated with clinicopathologic characteristics, TME, subtype, and PGS cluster in TCGA dataset

To determine whether the PGSscore could independently predict patient outcome, univariate and multivariate Cox regression analyses were conducted in a training set and a testing set. The hazard ratios

Table 1
Thirteen prognostic gene sets.

Prognostic gene sets	Source	HR	HR.95L	HR.95H	*p*-Value	links
GOBP_REGULATION_OF_B_CELL_ACTIVATION	Tumor	0.012118828	0.001318015	0.111429691	0.000718752	https://amigo.geneontology.org/amigo/term/GO:0050864
REACTOME_TNFS_BIND_THEIR_PHYSIOLOGICAL_RECEPTORS	Tumor	0.014707259	0.000499211	0.433290304	0.04534151	https://www.reactome.org/content/detail/R-HSA-5669034
GOBP_TOLERANCE_INDUCTION	Tumor	0.006054859	0.000171635	0.21360029	0.007916449	https://amigo.geneontology.org/amigo/term/GO:0002507
BIOCARTA_THELPER_PATHWAY	Tumor	0.032767433	0.004171448	0.257393737	0.014544625	https://data.broadinstitute.org/gsea-msigdb/msigdb/biocarta/human/h_thelperPathway.gif
REACTOME_RUNX1_AND_FOXP3_CONTROL_THE_DEVELOPMENT_OF_REGULATORY_T_LYMPHOCYTES_TREGS	Tumor	0.023384956	0.002234648	0.244716923	0.013938655	https://reactome.org/PathwayBrowser/#/R-HSA-8877330
BIOCARTA_NO2IL12_PATHWAY	Tumor	0.034205375	0.004525225	0.258552379	0.000119703	https://data.broadinstitute.org/gsea-msigdb/msigdb/biocarta/human/h_no2il12Pathway.gif
WP_SELECTIVE_EXPRESSION_OF_CHEMOKINE_RECEPTORS_DURING_TCELL_POLARIZATION	Tumor	0.04201833	0.003277525	0.538680953	0.020300962	https://www.wikipathways.org/instance/WP4494_r103559
GOBP_REGULATORY_T_CELL_DIFFERENTIATION	Tumor	0.006537493	0.000154828	0.276039811	0.025234999	https://amigo.geneontology.org/amigo/term/GO:0045066
GOBP_REGULATION_OF_TOLERANCE_INDUCTION	Tumor	0.019627105	0.000544318	0.707717068	0.042516185	https://amigo.geneontology.org/amigo/term/GO:0002643
GOBP_CHRONIC_INFLAMMATORY_RESPONSE	Tumor	0.002432552	4.32E-05	0.136832463	0.015498973	https://amigo.geneontology.org/amigo/term/GO:0002643
GOBP_POSITIVE_REGULATION_OF_TOLERANCE_INDUCTION	Tumor	0.018650268	0.001141307	0.304766735	0.012122161	https://amigo.geneontology.org/amigo/term/GO:0002645
GOBP_SEROTONIN_SECRETION	Non-tumor	0.020151167	0.000893779	0.454328891	0.048379596	https://amigo.geneontology.org/amigo/term/GO:0001820
REACTOME_P2Y_RECEPTORS	Non-tumor	0.033425315	0.002045168	0.546288412	0.017618262	https://reactome.org/PathwayBrowser/#/R-HSA-417957

of the PGSscore and its 95 % CI were 8.010 and 2.898–22.145 (p < 0.001; univariate Cox regression) (Fig. 9A) and 8.401 and 3.081–22.909 (p < 0.001; multivariate Cox regression), respectively, in the training set (Fig. 9B). These values were 7.851 and 1.759–35.039 (p = 0.007; univariate Cox regression) (Fig. S4F), and 6.879 and 1.494–31.677 (p = 0.013; multivariate Cox regression), respectively, in the testing set (Fig. S4G). To further explore the correlation between the PGSscore, clinicopathologic characteristics, and TME in the entire TCGA cohort, we analyzed the data using Wilcoxon rank sum tests. Our results indicate that more patients in the high-PGSscore group had advanced T (p < 0.001) (Fig. 9C) and N (p < 0.05) stage disease (Fig. 9D). Patients with stage III – IV disease also showed higher PGSscores than did those with stage I – II disease (p < 0.001) (Fig. 9E). Additionally, there was a negative correlation between the PGSscore and the immune score (p < 0.001) (Fig. 9F) and ESTIMATE score (p < 0.01) (Fig. 9J), whereas no correlation was noted between the PGSscore and the stromal score (p > 0.05) (Fig. 9H). Fig. S4H displays the correlation between the PGSscore, various clinicopathologic features, and gene expression in the entire TCGA set. To further validate the predictive ability of the PGSscore in groups divided by the clinicopathologic index, we performed a stratification analysis, which revealed that the prognosis of patients in the low- and high-PGSscore groups differed significantly according to stratification in the entire TCGA cohort (Fig. S5A-S5K). Subsequently, an alluvial diagram was constructed to better visualize the relationship among the subtypes, PGS cluster, PGSscore, and patient outcomes (Fig. 9I). We then examined the relationship between the PGSscore, subtypes 1–3, and PGS cluster A-C in more detail. As expected, subtype 3 had a low PGSscore and better prognosis, while subtype 1 had a high PGSscore (Fig. 9J). PGS cluster C showed a low PGSscore, whereas cluster B had a high PGSscore (Fig. 9K).

3.9. The immune cell landscape and the ability of the PGSscore to predict immunotherapy response

To characterize the relationship between the immune cell landscape and the PGSscore, we calculated the Spearman correlation between the PGSscore and tumor-infiltrating immune cells in TCGA dataset using XCELL, TIMER, QUANTISEQ, MCPcounter, EPIC, CIBERSORT-ABS, and CIBERSORT. Immune infiltrating cells such as B cells, $CD8^+$ T cells, Tregs, and activated NK cells were negatively correlated with the PGSscore, while common lymphoid progenitor M2 macrophages and resting NK cells were positively correlated with the PGSscore. In addition, the correlation coefficient of most immune infiltrating cell types was<0, suggesting that the counts of most types of immune cells were increased with the reduction of PGSscore (Fig. 10A and Table S4). The relative proportions of tumor-infiltrating immune cells between the low- and high-PGSscore groups are displayed in Fig. S6A and B. To estimate the capacity of the PGSscore to predict responses to ICI therapy in TCGA samples, we determined the relationship between the low- and high-PGSscore groups and immune checkpoint-related genes. Almost all ICI-related genes had significantly higher expression in the low-PGSscore group (Fig. 10B). The predictive ability of the PGSscore was also assessed by IPS analysis. IPS-PD1-CTLA4 scores and IPS-PD1 scores were significantly higher in the low-PGSscore group in TCGA dataset (Fig. 10C). Fig. 10D shows that the TIDE score in the low-PGSscore group was significantly lower than that in the other groups, suggesting that the PGSscore is related to the outcome of ICI therapy, and the low-PGSscore group had better responses to ICIs in TCGA dataset. To examine the utility of the PGSscore in predicting immunotherapy outcomes, the PGSscore for individuals in the GSE67501 and IMvigor210 cohorts were calculated based on the candidate genes. The correlation coefficients were calculated using TCGA dataset. The GSE67501 and IMvigor210 cohorts were divided into low- and high-PGSscore groups. Our results illustrate that the non-responsive group had a higher PGSscore (Wilcoxon test, p = 0.033) (Fig. 10E) and that a larger proportion of high-PGSscore patients belonged to the non-response group (Fig. 10F) in the GSE67501 cohort. Notably, the low-PGSscore group in the IMvigor210 cohort had a better prognosis than the high-PGSscore group (log rank test, p = 0.033) (Fig. 10G). The complete response/partial response group had a lower PGSscore (Wilcoxon test, p = 0.027) (Fig. 11H) and a larger

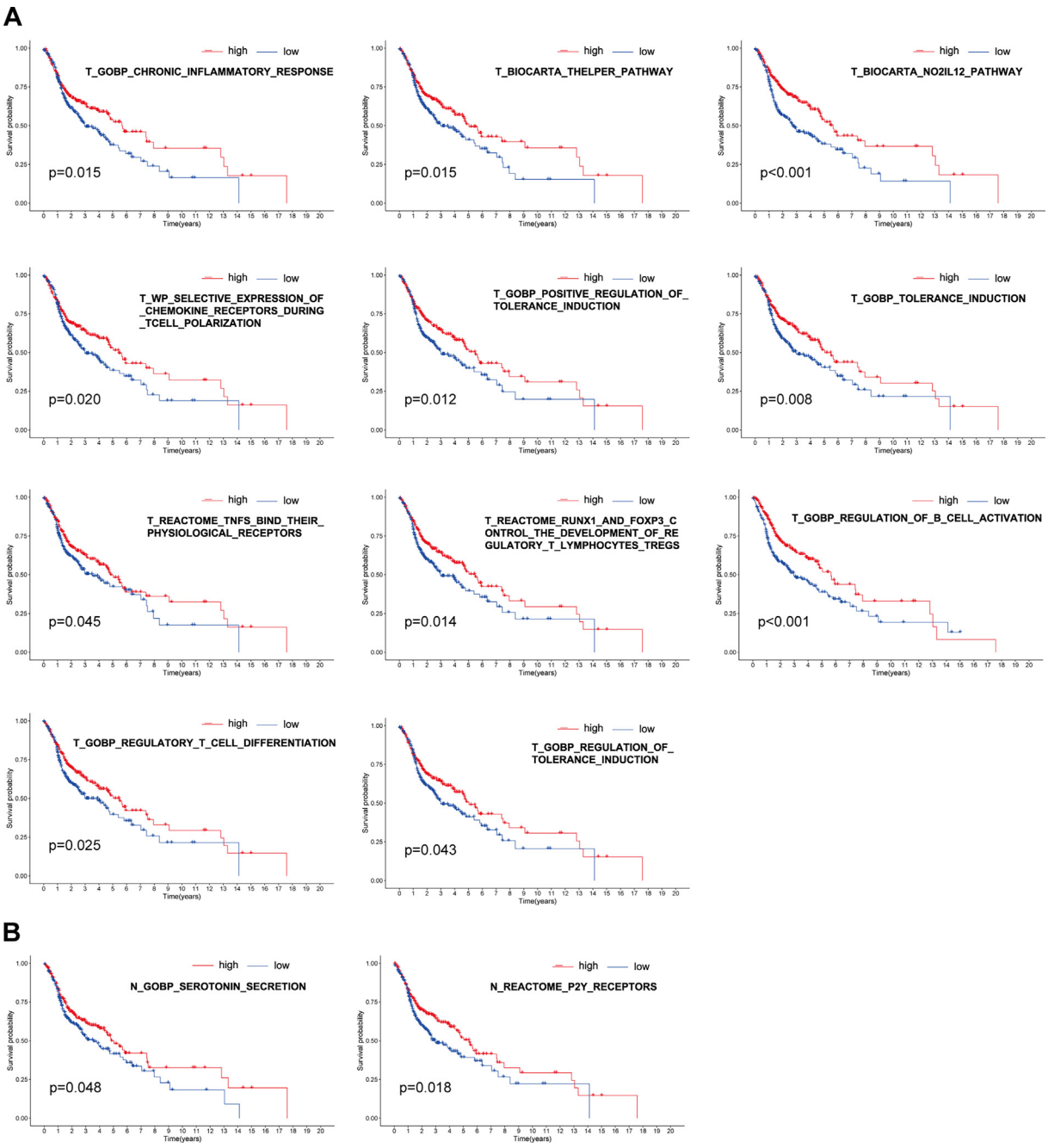


Fig. 5. Survival analysis between high- and low-ES groups of HNSC patients in TCGA dataset. Patients were divided into high- and low-ES groups according to the median ES of each prognostic gene set obtained from tumor sample data (A) and adjacent non-tumor tissue sample data (B). Kaplan-Meier survival analysis revealed significant differences between two gene sets (log-rank test, $p < 0.05$).

percentage of low-PGSscore patients (Fig. 10I) compared with the stable disease/progressive disease group in the IMvigor210 cohort.

3.10. Establishment and evaluation of the HNSC nomogram

We combined various predictive factors, including clinicopathologic characteristics and the PGSscore based on TCGA dataset, to increase utility and achieve higher prognostic accuracy of a nomogram that predicts the 1-, 3-, and 5-year overall survival. The predicting factors included grade, sex, age, HPV status, tumor anatomical subsites, stage, and PGSscore (Fig. 11A). To consider the clinicopathologic utility of the model, we conducted DCA to integrate patient preferences or decision points into the analysis. When the threshold probability was above 0.2, the nomogram was superior at predicting prognosis than individual factors (Fig. 11B). In addition, both the C-index and ROC curves indicated that the nomogram has high predictive accuracy. The C-index of the nomogram fluctuated by 0.80 from 1 to 5 years (Fig. 11C), while the area under the ROC curve (AUC) was 0.816, 0.803, and 0.771 for 1-, 3-, and 5-year survival, respectively (Fig. 11D–F). The calibration plots revealed that the results predicted by the nomogram were very close to the actual results for 1, 3, and 5 years (Fig. 11G–I). Finally,

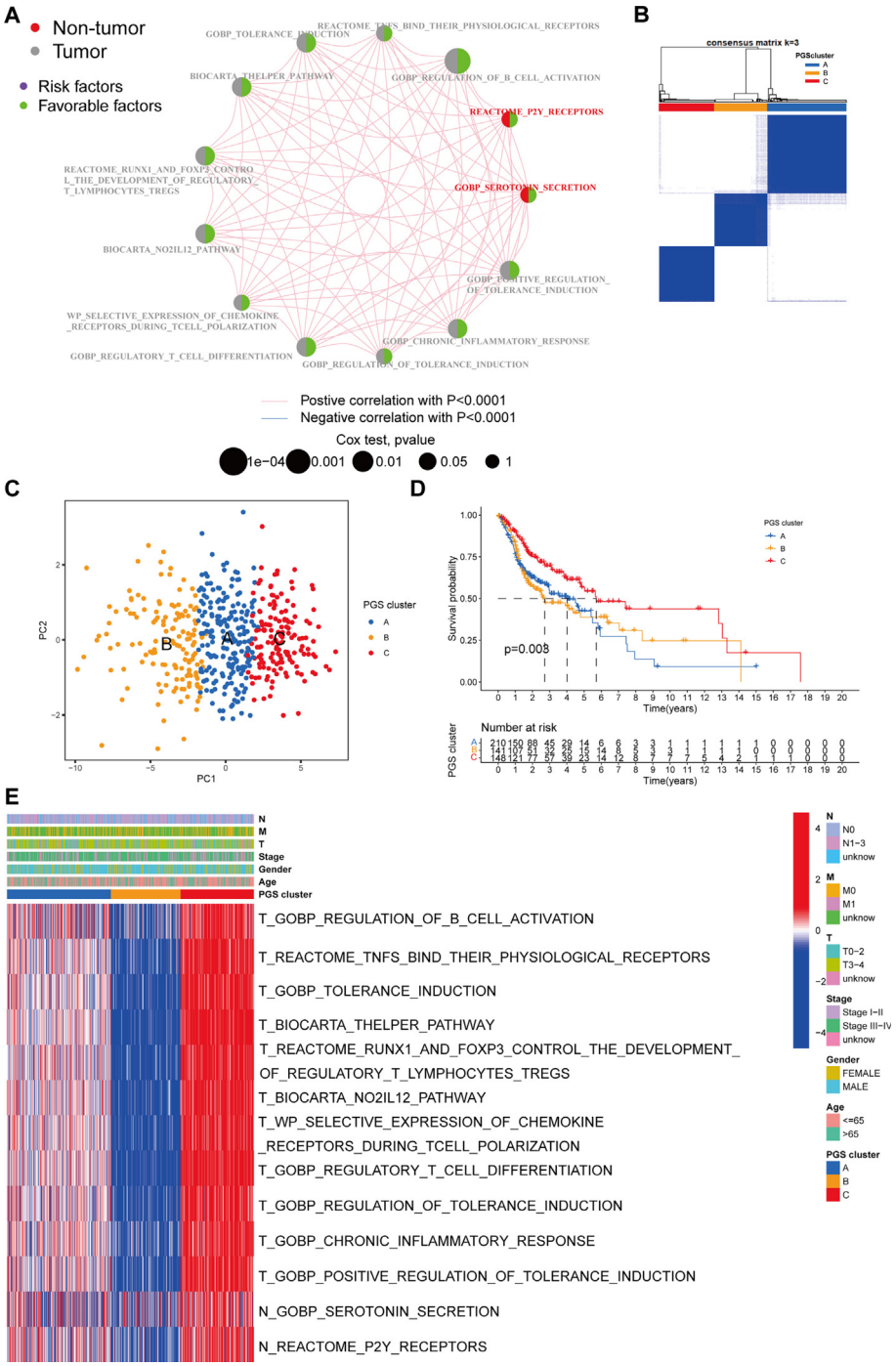


Fig. 6. Identification of PGS clusters and functional analysis using HNSC samples in TCGA dataset. (A) Correlations of 13 prognostic gene sets in HNSC. Red and grey semi-circles represent gene sets derived from adjacent non-tumor and tumor tissue samples, respectively. The size of the circle represents the prognostic value of each gene set. The *p*-value was calculated by log-rank test. Green semi-circles represent favorable factors, while purple semi-circles represent risk factors. Links between genes indicate their correlations. Red lines represent positive correlations and blue lines represent negative correlations. (B) Consensus clustering matrix for k = 3 was calculated using the ConsensusClusterPlus R package. (C) Three significantly distinct clusters named PGS clusters A–C were identified using PCA based on the ESs of the prognostic gene sets in HNSC samples. (D) Kaplan-Meier survival analysis of the three PGS clusters revealed statistical differences between the three clusters (log-rank test, $p = 0.008$). (E) A heatmap showing the ESs of the 13 prognostic gene sets in the three PGS clusters. Red represents high ESs and blue represents low ESs. PGS cluster, age, gender, stage, T, N, and M were used as sample annotations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

we compared the 1-year AUC value of the nomogram developed based on the PGSscore with that of other existing signatures [14,41–46], indicating that our nomogram had higher predictive accuracy than individual signatures (Fig. 11J).

4. Discussion

In this study, we used GSVA to establish a pathway-centric model of HNSC and obtain novel insights into variations in path-

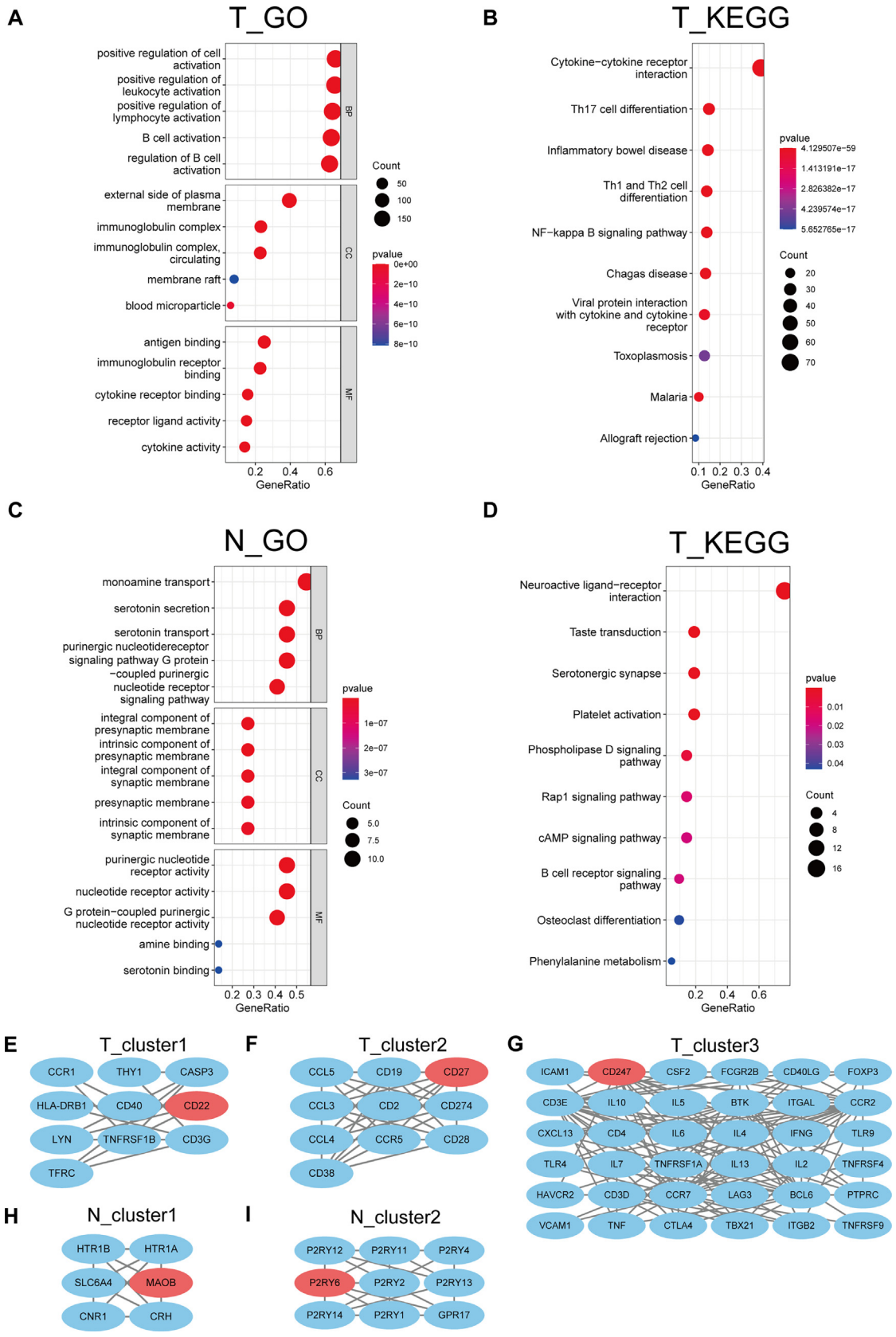
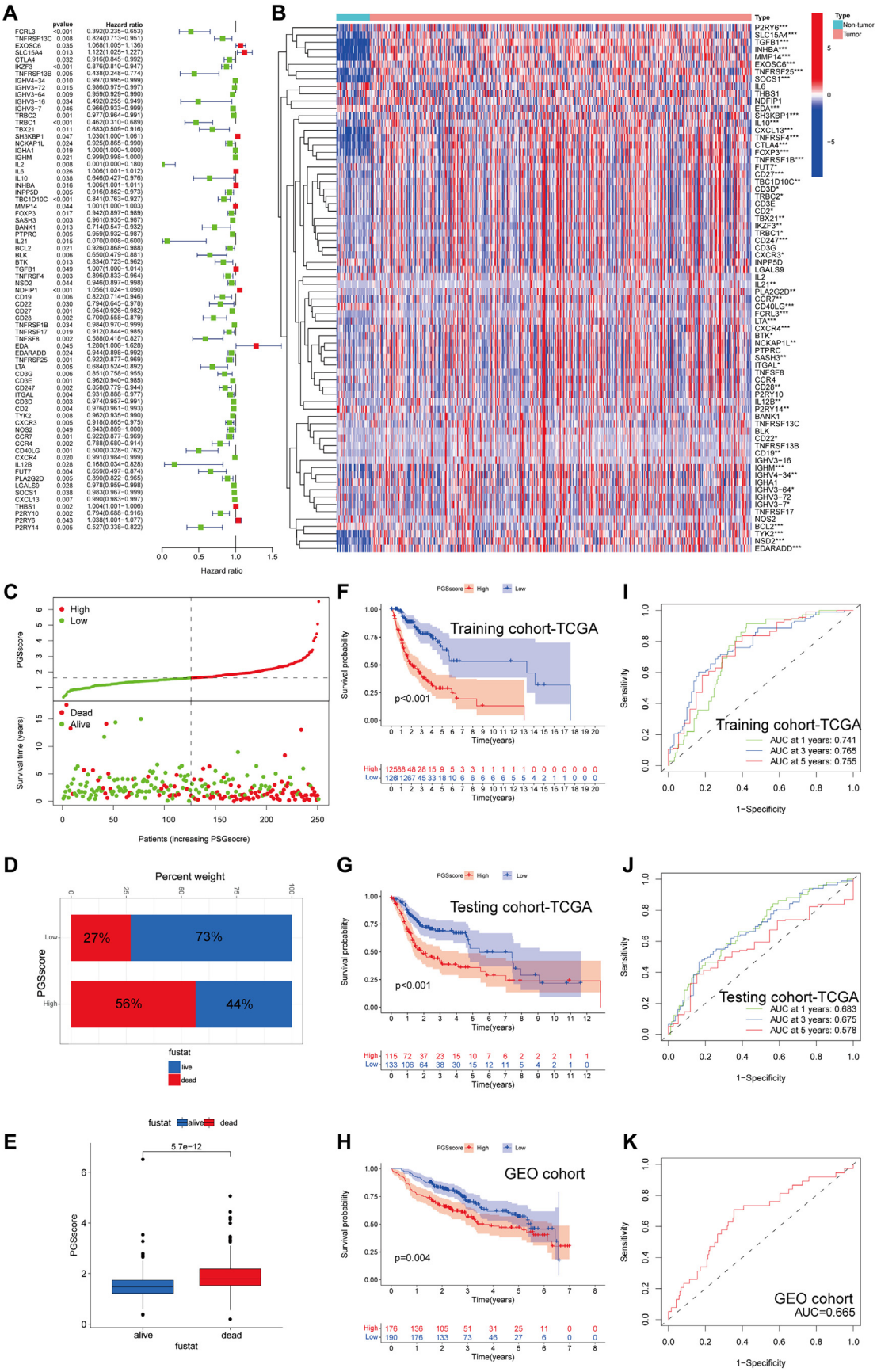


Fig. 7. Functional enrichment analysis and module analysis for the genes in the prognostic gene sets in TCGA dataset. The GO enrichment and KEGG pathway analyses of genes in the T gene sets (A and B) and N gene sets (C and D). The size of the bubbles represents the number of enriched gene sets, and the color represents the *p*-value. BP, biological process; CC, cellular component; MF, molecular function. Networks based on the genes in the T gene sets (E-G) and N gene sets (H and I) were calculated using the MCODE application. Genes in red were identified by hub genes using CYTOHUBBA. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)


A
pvalue
Hazard ratio
FCRL3 <0.001 0.392(0.235−0.653)
TNFRSF13C 0.008 0.824(0.713−0.951)
EXOSC6 0.035 1.068(1.005−1.136)
SLC15A4 0.013 1.122(1.025−1.227)
CTLA4 0.032 0.916(0.845−0.992)
IKZF3 <0.001 0.876(0.810−0.947)
TNFRSF13B 0.005 0.438(0.248−0.774)
IGHV4−34 0.010 0.997(0.995−0.999)
IGHV3−72 0.015 0.986(0.975−0.997)
IGHV3−64 0.005 0.959(0.929−0.990)
IGHV3−16 0.034 0.492(0.255−0.949)
IGHV3−7 0.046 0.966(0.933−0.999)
TRBC2 0.001 0.977(0.964−0.991)
TRBC1 <0.001 0.462(0.310−0.689)
TBX21 0.011 0.683(0.509−0.916)
SH3KBP1 0.047 1.030(1.000−1.061)
NCKAP1L 0.024 0.925(0.865−0.990)
IGHA1 0.019 1.000(1.000−1.000)
IGHM 0.021 0.999(0.998−1.000)
IL2 0.008 0.001(0.000−0.180)
IL6 0.026 1.006(1.001−1.012)
IL10 0.038 0.646(0.427−0.976)
INHBA 0.016 1.006(1.001−1.011)
INPP5D 0.005 0.916(0.862−0.973)
TBC1D10C <0.001 0.841(0.763−0.927)
MMP14 0.044 1.001(1.000−1.003)
FOXP3 0.017 0.942(0.897−0.989)
SASH3 0.003 0.961(0.935−0.987)
BANK1 0.013 0.714(0.547−0.932)
PTPRC 0.005 0.959(0.932−0.987)
IL21 0.015 0.070(0.008−0.600)
BCL2 0.021 0.926(0.868−0.988)
BLK 0.006 0.650(0.479−0.881)
BTK 0.013 0.834(0.723−0.962)
TGFB1 0.049 1.007(1.000−1.014)
TNFRSF4 0.003 0.896(0.833−0.964)
NSD2 0.044 0.946(0.897−0.998)
NDFIP1 <0.001 1.056(1.024−1.090)
CD19 0.006 0.822(0.714−0.946)
CD22 0.030 0.794(0.645−0.978)
CD27 0.001 0.954(0.926−0.982)
CD28 0.002 0.700(0.558−0.879)
TNFRSF1B 0.034 0.984(0.970−0.999)
TNFRSF17 0.019 0.912(0.844−0.985)
TNFSF8 0.002 0.588(0.418−0.827)
EDA 0.045 1.280(1.006−1.628)
EDARADD 0.024 0.944(0.898−0.992)
TNFRSF25 0.001 0.922(0.877−0.969)
LTA 0.005 0.684(0.524−0.892)
CD3G 0.006 0.851(0.758−0.955)
CD3E 0.001 0.962(0.940−0.985)
CD247 0.002 0.858(0.779−0.944)
ITGAL 0.004 0.93(0.888−0.977)
CD3D 0.003 0.974(0.957−0.991)
CD2 0.004 0.976(0.961−0.993)
TYK2 0.008 0.962(0.935−0.990)
CXCR3 0.005 0.918(0.865−0.975)
NOS2 0.049 0.943(0.889−1.000)
CCR7 0.001 0.922(0.877−0.969)
CCR4 0.002 0.788(0.680−0.914)
CD40LG 0.001 0.500(0.328−0.762)
CXCR4 0.020 0.991(0.984−0.999)
IL12B 0.028 0.168(0.034−0.828)
FUT7 0.004 0.659(0.497−0.874)
PLA2G2D 0.005 0.890(0.822−0.965)
LGALS9 0.028 0.978(0.959−0.998)
SOCS1 0.038 0.983(0.967−0.999)
CXCL13 0.007 0.990(0.983−0.997)
THBS1 0.002 1.004(1.001−1.006)
P2RY10 0.002 0.794(0.688−0.916)
P2RY6 0.043 1.038(1.001−1.077)
P2RY14 0.005 0.527(0.338−0.822)
0.0 0.5 1.0 1.5
Hazard ratio
B
Type
Non-tumor
Tumor
5
0
−5
P2RY6*** SLC15A4*** TGFB1*** INHBA*** MMP14*** EXOSC6*** TNFRSF25*** SOCS1*** IL6 THBS1 NDFIP1 EDA*** SH3KBP1*** IL10*** CXCL13*** TNFRSF4*** CTLA4*** FOXP3*** TNFRSF1B*** FUT7* CD27*** TBC1D10C** CD3D* TRBC2* CD3E CD2* TBX21** IKZF3** TRBC1* CD247*** CD3G CXCR3* INPP5D LGALS9 IL2 IL21** PLA2G2D** CCR7** CD40LG*** FCRL3*** LTA*** CXCR4*** BTK* NCKAP1L** PTPRC SASH3** ITGAL* TNFSF8 CCR4 CD28** P2RY10 IL12B** P2RY14** BANK1 TNFRSF13C BLK CD22* TNFRSF13B CD19** IGHV3−16 IGHM*** IGHV4−34** IGHA1 IGHV3−64* IGHV3−72 IGHV3−7* TNFRSF17 NOS2 BCL2*** TYK2*** NSD2*** EDARADD***
C
High
Low
PGSscore
Dead
Alive
Survival time (years)
0 50 100 150 200 250
Patients (increasing PSGsocre)
D
Percent weight
0 25 50 75 100
Low
High
PGSscore
27%
73%
56%
44%
fustat
live
dead
E
fustat alive dead
5.7e−12
PGSscore
alive
dead
fustat
F
PGSscore High Low
Training cohort-TCGA
p<0.001
Survival probability
Time(years)
High 125 88 48 28 15 9 5 3 3 1 1 1 1 1 0 0 0 0 0 0 0
Low 125 126 77 45 33 18 10 6 6 6 6 6 5 5 4 2 1 1 0 0 0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
G
PGSscore High Low
Testing cohort-TCGA
p<0.001
Survival probability
Time(years)
High 115 72 37 23 15 10 7 6 2 2 2 1 1
Low 133 106 64 38 30 15 12 11 5 4 2 1 0
0 1 2 3 4 5 6 7 8 9 10 11 12
H
PGSscore High Low
GEO cohort
p=0.004
Survival probability
Time(years)
High 176 136 105 51 31 25 11 0 0
Low 190 176 133 73 46 27 6 0 0
0 1 2 3 4 5 6 7 8
I
Sensitivity
Training cohort-TCGA
AUC at 1 years: 0.741
AUC at 3 years: 0.765
AUC at 5 years: 0.755
1−Specificity
J
Sensitivity
Testing cohort-TCGA
AUC at 1 years: 0.683
AUC at 3 years: 0.675
AUC at 5 years: 0.578
1−Specificity
K
Sensitivity
GEO cohort
AUC=0.665
1−Specificity

way activity between tumor and adjacent non-tumor samples to better identify HNSC subtypes, detect critical gene sets or pathways, and establish a score system to predict overall survival and immunotherapy responses.

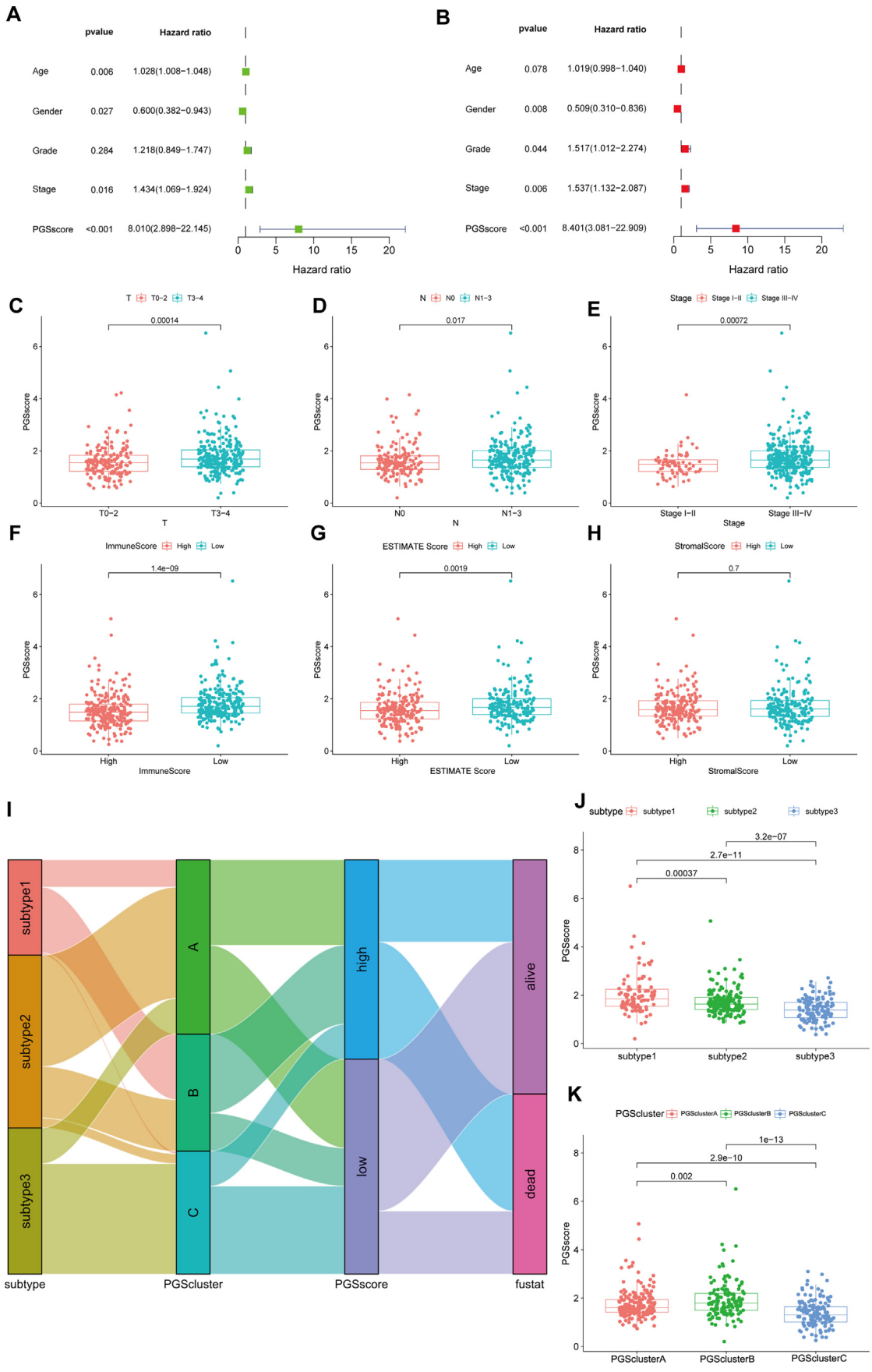
Using the NMF method, we classified HNSC patients into three different subtypes according to the ESs of the gene sets between tumor and adjacent non-tumor samples. In comparison with previously reported HNSC molecular subtypes [5], our subtype 3 was more common in the less aggressive, atypical subtype, which was consistent with the favorable overall survival of subtype 3. This relationship may partially reveal the distinct characteristics between molecular subtypes. In 2018, Thorsson et al. [26] reported six immune subtypes for a pan-cancer cohort. After integration of these immune subtypes with our data, we determined that the majority of patients with HNSC had wound healing (25 %) or IFN-c dominant (73 %) immune subtypes, with only 2 % of patients associated with the other four immune subtypes. Thus, the pan-cancer immune subtypes may not be completely applicable to HNSC. This new subtype stratification offers added value, as it is more comprehensive given that we did not only concentrate on the tumor itself but also considered pathway activity in non-tumor tissues. In contrast, Walter et al. [5] determined four molecular classes using genomic analysis in tumor samples only, while ignoring non-tumor tissues. Therefore, our identified subtypes add a functional perspective to adjacent non-tumor samples. Furthermore, the hierarchical clustering analyses reported by Walter et al. [5] and Chung et al. [3] were based on gene expression profiles. However, gene function is exerted collectively and continuously altered in response to different environmental conditions, disease states, and genetic modifications [28]. Moreover, well-annotated gene sets representing an extensive collection of biological processes are critical for insightful interpretation of large-scale genomic data [47]. Therefore, gene expression (molecular) profiles could have lower interpretability compared to well-annotated pathway-centered profiles. Our method represents an improvement over previous approaches as the genes were condensed into gene sets before clustering analysis; thus, the gene sets profile provides a more intuitive and stable context for evaluating biological activity [48]. In addition, we included a wider, more diverse range of gene set types rather than a single popular and emerging gene set [11]. Finally, we observed that clinicopathologic characteristics and risk factors were significantly and distinctly distributed among the three subtypes. Patients with factors beneficial to survival, such as a lower disease stage, lower T stage, positive HPV status [49], and lower alcohol consumption [50], tended to have subtype 3 HNSC, whereas patients with factors detrimental to survival tended to have subtype 1 HNSC. By contrast, such clinical associations were not significant in a previous study [5].

From the representative gene sets determined from the intersection of differentially enriched genes among the three subtypes, we identified 11 prognostic gene sets in HNSC and 2 in non-tumor samples. The low ESs of these gene sets were highly correlated with negative outcomes. However, we emphasize that low ESs do not indicate downregulated pathways. PGS clusters of patients differing in overall survival based on the relative ESs of the 13 gene sets were similar to the clustering of subtypes established based on the relative ESs of all gene sets, with three groups identified in both cases. In addition, most of the patients in PGS cluster C, with the best survival outcomes, were classified in subtype 3, whereas the greatest proportion of patients in PGS cluster B, with the lowest ESs and worst outcomes, were classified in subtype 1. This suggests that the 13 gene sets we obtained are representative of all gene sets and are closely associated with prognosis.

Our results indicate that serotonin secretion and P2Y receptor activation in the non-tumor tissue may significantly affect tumor progression. Serotonin stimulates cancer cell growth [51], differentiation [52], migration and metastasis [53], and angiogenesis [54]. Woods et al. [55] showed that purinergic receptor P2Y2 (P2Y2R) positively regulates the responses of HNSC cells to extracellular nucleotides, thus promoting cancer cell proliferation, and further summarized how different P2Y receptors influence tumorigenesis [56]. However, no studies have investigated the role of serotonin and P2Y receptors in non-tumor tissues adjacent to HNSC tissues. We hypothesize that changes in serotonin secretion and P2Y receptor activity could be regarded as a prognostic biomarker for HNSC.

Among the 11 gene sets in tumor tissues, we hypothesized that those from the T gene sets are correlated with immune-related pathway activity changes. Regulation of B cell activation was identified as a determinant of HNSC prognosis. Previous studies have also indicated that increased B cell counts are associated with good prognosis in various solid tumors [57–59]). B cell-mediated secretion of inflammatory cytokines such as interferon-γ has been suggested as a potential mechanism underlying the inhibition of tumor growth and progression [60]. Interestingly, specific features of B cell genes are detected in HPV^+, but not HPV^- tumors, which is consistent with the fact that patients with HPV^+ HNSC have better outcomes than those with HPV^- HNSC [61]. Further, tolerance induction seems to be an important determinant of HNSC prognosis. HNSC cells may use different strategies to escape immune surveillance, such as downregulation of histocompatibility molecules and production of cytokines and chemotactic factors, potentially contributing to the high reoccurrence rate of HNSC [62]. The T helper pathway, development of Tregs (regulated by *FOXP3* and *RUNX1*, T cell differentiation, and selective expression of chemokine receptors during T cell polarization were also included in tumor tissue gene sets; these may play an important role in the anti-tumor process [62–67]. The NO_2-dependent IL-12 pathway was also associated with these gene sets, which is a functional pathway in NK cells that exhibit potent cytotoxicity and play a key role in anti-tumor immunity. These results indicate that the critical immune characteristics of HNSC tissues could be accurately determined using our method. In addition to immunity, our results suggest that changes in the expression of these gene sets are associated with inflammation. The remaining two gene sets included TNF superfamily members and the chronic inflammation response. TNF-α induces apoptosis, regulates immune responses and the functioning of tumor tissue vascular systems, and induces necrosis [68]. Moreover, various cytokines produced during the inflamma-

Fig. 8. PGSscore based on candidate genes from prognostic gene sets predict patient outcomes. (A) Forest plot showing 72 genes correlated with clinical prognosis were selected using univariate Cox regression analysis in TCGA dataset. (B) Heatmap showing differences in the expression of 72 genes between HNSC and non-tumor samples in TCGA dataset. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. (C) Distribution of the patient survival status ordered by PGSscores in the training set. (D) Overall survival in the low- and high-PGSscore groups in the training set. (E) Correlation between the PGSscore and survival status in the training set. The lines represent the median values. Statistically significant differences among the three subtypes were analyzed using Wilcoxon tests. Survival analysis of low- and high-PGSscore groups in the training set (F), testing set (G), and GEO cohort (H). Kaplan-Meier survival analysis revealed substantial differences among the two groups (log-rank test, $p < 0.01$). The ROC curve of the PGSscores predicted the 1-, 3-, and 5-year outcomes in the training set (I) and testing set (J), and the predicted 1-year outcomes in the GEO cohort (K).


A
pvalue
Hazard ratio
Age 0.006 1.028(1.008−1.048)
Gender 0.027 0.600(0.382−0.943)
Grade 0.284 1.218(0.849−1.747)
Stage 0.016 1.434(1.069−1.924)
PGSscore <0.001 8.010(2.898−22.145)
0 5 10 15 20
Hazard ratio
B
pvalue
Hazard ratio
Age 0.078 1.019(0.998−1.040)
Gender 0.008 0.509(0.310−0.836)
Grade 0.044 1.517(1.012−2.274)
Stage 0.006 1.537(1.132−2.087)
PGSscore <0.001 8.401(3.081−22.909)
0 5 10 15 20
Hazard ratio
C
T T0−2 T3−4
0.00014
PGSscore
T0−2
T3−4
T
D
N N0 N1−3
0.017
PGSscore
N0
N1−3
N
E
Stage Stage I−II Stage III−IV
0.00072
PGSscore
Stage I−II
Stage III−IV
Stage
F
ImmuneScore High Low
1.4e−09
PGSscore
High
Low
ImmuneScore
G
ESTIMATE Score High Low
0.0019
PGSscore
High
Low
ESTIMATE Score
H
StromalScore High Low
0.7
PGSscore
High
Low
StromalScore
I
subtype1
subtype2
subtype3
A
B
C
high
low
alive
dead
subtype
PGScluster
PGSscore
fustat
J
subtype subtype1 subtype2 subtype3
3.2e−07
2.7e−11
0.00037
PGSscore
subtype1
subtype2
subtype3
K
PGScluster PGSclusterA PGSclusterB PGSclusterC
1e−13
2.9e−10
0.002
PGSscore
PGSclusterA
PGSclusterB
PGSclusterC

tory response play important roles in promoting HNSC cell proliferation, invasion, and metastasis [10]. Additionally, we identified three hub genes from the T gene sets and two hub genes from the N gene sets. Among them, *CD22*, *CD27*, *CD247*, and *MAOB* have been reported to be correlated with HNSC tumorigenesis and prognosis [69–72].

More, we used the extracted genes to develop a scoring system, i.e., the PGSscore. Apart from PGSscores' robust prognostic predictive ability, it also could assist to assess immune status. By integrating the results of these analyses, we found that the PGSscore was negatively associated with immune cell infiltration and expression of immune checkpoint genes. Other methods, such as IPS [33] and TIDE analysis [35], can also predict the responses of patients to immunotherapy. These results indirectly indicate that patients in the low-PGSscore group may have an immunogenic tumor microenvironment. We further confirmed the predictive value of the PGSscore using data from the IMvigor210 and GSE67501 cohorts (subjected to anti-PD-1 immunotherapy). A significant difference in the PGSscore was observed between non-responders and responders. This provides direct evidence of the fact that the low-PGSscore group had a higher distribution of better-response patients. Previous studies also identified potential immune classes of HNSC. Chen et al. [73] identified an immune class divided into active and exhausted immune subgroups. The active immune subgroup may be similar to our low-PGSscore group, as they are both characterized by B cell-related immune signatures and proinflammatory M1 macrophages. This revealed that the humoral immune response may play a crucial role in intratumoral immune response activation and exhaustion in HNSC. Song et al. [65] demonstrated that patients with HNSC can be classified into immunity-H, immunity-M, or immunity-L groups. Immunity-H patients had the highest levels of immune infiltration and PD-L1 expression, and had good prognoses. Therefore, the low-PGSscore group may be highly relevant to the Immunity-H group, since the PD-L1 expression and overall survival in the low-PGSscore group were significantly higher than that of the high-PGSscore group. Compared with previous immunophenotype- and immunotherapy-related studies, the advantage of our study is that the adjacent non-tumor tissues were also considered in the development of PGSscore. These tissues may play an important role in immunotherapy, and are worth further exploration in future research.

In terms of clinical application, HNSC can be divided into subtypes 1, 2, and 3, according to pathway activity changes between tumor and adjacent non-tumor tissues, and HNSC patients can also be classified into low- and high-PGSscore groups. These two stratifications are interrelated, but may be distinct in clinical practice. They are similar in that they can both predict the clinical outcomes of patients. Notably, nomograms based on PGSscore not only improved the predictive performance but also outperformed individual factors in terms of a clinical net benefit [39]. Since adjacent non-tumor tissues are directly involved in the analysis, the subtype stratification may be helpful to find effective anti-tumor targets in non-tumor tissues and implement precise treatment, either by activating or inhibiting the corresponding signaling pathway according to the patient's profile. In addition, applying these subtypes can offer a simple method for clinicians to generate preliminary conjectures about the clinicopathological characteristics and living habits of patients before specialized clinical examinations. Moreover, stratification by low- and high-PGSscore groups may provide precise and tailored immunotherapeutic strategies, as it can offer an improved method to evaluate a patient's response to immunotherapy. These findings and prospects warrant further exploration in larger HNSC cohorts receiving immune checkpoint therapies.

There are some limitations to this study. First, we did not include other datasets with adjacent non-tumor HNSC samples, except for TCGA dataset. A dataset with both tumor and non-tumor tissue samples would better verify the optimal number of subtype clusters. Second, the clinicopathologic information from the two GEO datasets is inconsistent with that from TCGA cohort, so it cannot be used to verify the nomogram. Finally, since no immunotherapy-treated HNSC cohort could be used for confirmation, we made use of existing immunotherapy cohorts of patients with other tumor types.

5. Conclusion

We used GSVA to establish a pathway-centric matrix and comprehensively indicated subtle and pronounced pathway activity variations that can be used to develop a more intuitive and stable context for evaluating biological activity between HNSC and non-tumor samples. Based on the obtained tumor and non-tumor data, we identified three HNSC subtypes that exhibited distinct clinicopathologic features and were associated with risk factors and previously established subgroups. We also screened 2 and 11 gene sets from non-tumor and tumor samples, respectively, based on the representative gene sets. Genes extracted from non-tumor gene sets were associated with serotonin secretion and P2Y receptors, while genes from tumor gene sets were related to immunity and inflammation. Our results emphasize the indispensable role of non-tumor samples in influencing prognosis, and they indicate that the current strategy used for selecting the appropriate HNSC therapy should also consider pathway activity variation in non-tumor tissues. Therapies targeting serotonin and P2Y receptors in non-tumor tissues could serve as a novel treatment for HNSC in the future. We also illustrate that the PGSscore is an independent prognostic biomarker that can evaluate the clinicopathological characteristics and TME of each patient. In addition, we found that the PGSscore can be used to predict patient response to immunotherapy and it was used to establish a quantitative nomogram. This study provides novel insights for cancer treatment and suggests that targeting changes in pathway activity in non-tumor and tumor tissues may be used to develop immunotherapeutic agents. Meanwhile, our method to estimate pathway-centric models should be extended to other tumors and diseases.

Funding

This study received no funding from any governmental or private institutions.

Fig. 9. Identification of the PGSscore as an independent prognostic factor correlated with clinicopathologic characteristics, TME, subtypes, and PGS clusters in TCGA dataset. Forest plots of prognostic predictors in the training group identified via univariate Cox regression analysis (A) and multivariate Cox regression analysis (B). The correlation between the PGSscore and the T (C), N (D), and pathologic stages (E), immune score (F), ESTIMATE score (G); stromal score (H) in the entire TCGA set. The line in each box represents the median value. Statistical differences were analyzed by Wilcoxon tests. (I) Alluvial diagram displaying the relationship or transformation of the subtype, PGS cluster, PGSscore, and survival state in the entire TCGA set. (J) Difference of PGSscores among the 3 subtypes in the entire TCGA set ($p < 0.001$, Kruskal-Wallis test). (K) Differences in PGSscores among the 3 PGS clusters in the entire TCGA set ($p < 0.01$, Kruskal-Wallis test).

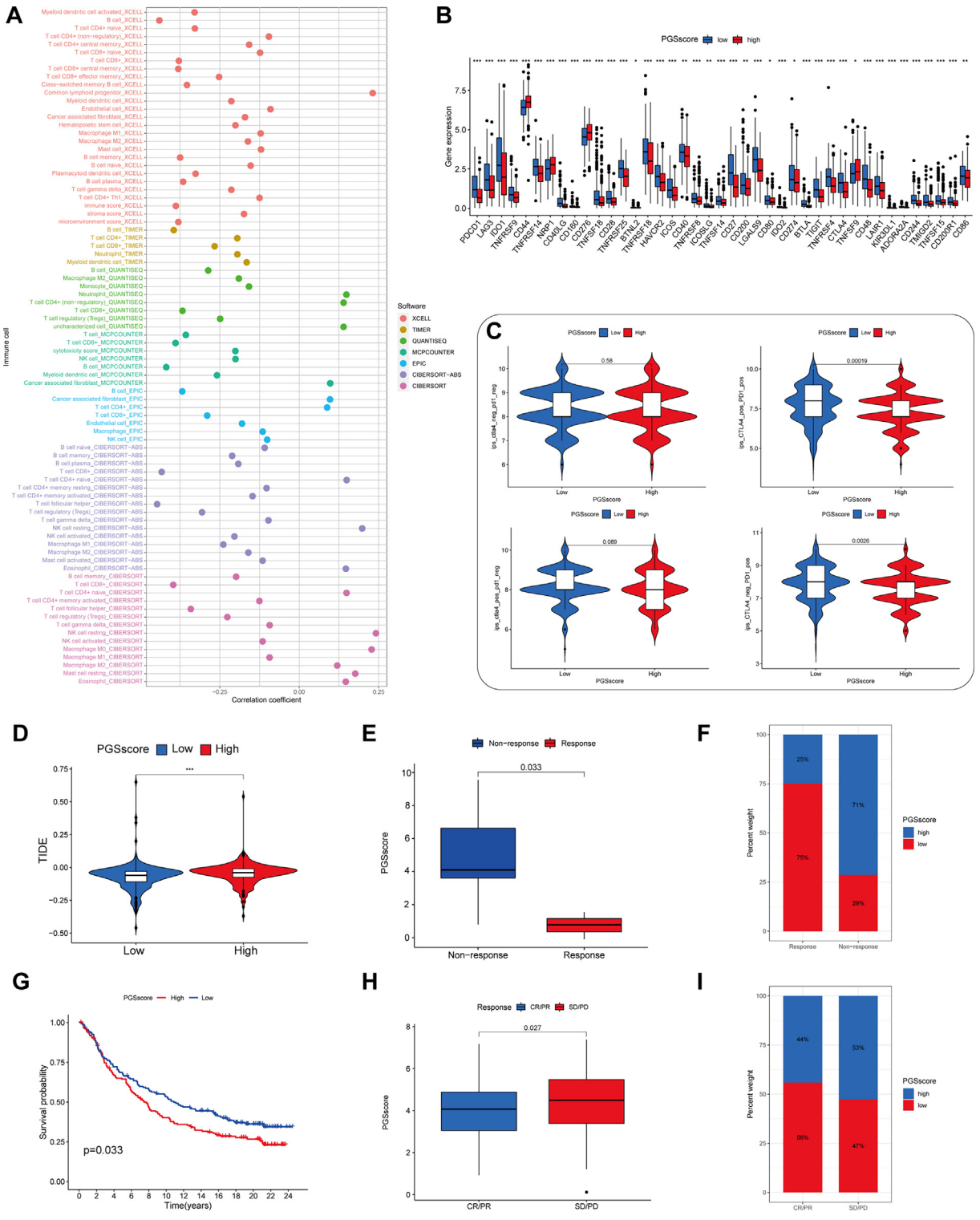


Fig. 10. Immune cell landscape and the ability of the PGSscore to predict the response to immunotherapy. (A) Lollipop plot showing the Spearman correlation between PGSscore and tumor-infiltrating immune cells in TCGA cohort. (B) The expression of immune checkpoint genes between the low- and high-PGSscore groups in TCGA cohort. (C) The relationship between IPS and PGSscore in TCGA cohort. The line in each box represents the median value (Wilcoxon test). (D) TIDE scores between low- and high-PGSscore groups in TCGA cohort. (E) PGSscores between non-responsive and responsive immunotherapy groups in the GSE67501 dataset. (F) The distribution of low- and high-PGSscore groups in the non-responsive and responsive groups in the GSE67501 dataset. (G) Survival analysis of low- and high-PGSscore groups in the IMvigor210 cohort. Kaplan-Meier survival curve revealed significant differences between the groups (log-rank test, p = 0.033). (H) Differences in PGSscores of the complete response/partial response group and stable disease/progressive disease group in the IMvigor210 cohort. (I) The proportion of low- and high-PGSscore groups in the response/partial response and stable disease/progressive disease groups in the IMvigor210 cohort. The line in each box represents the median value. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; Wilcoxon test.

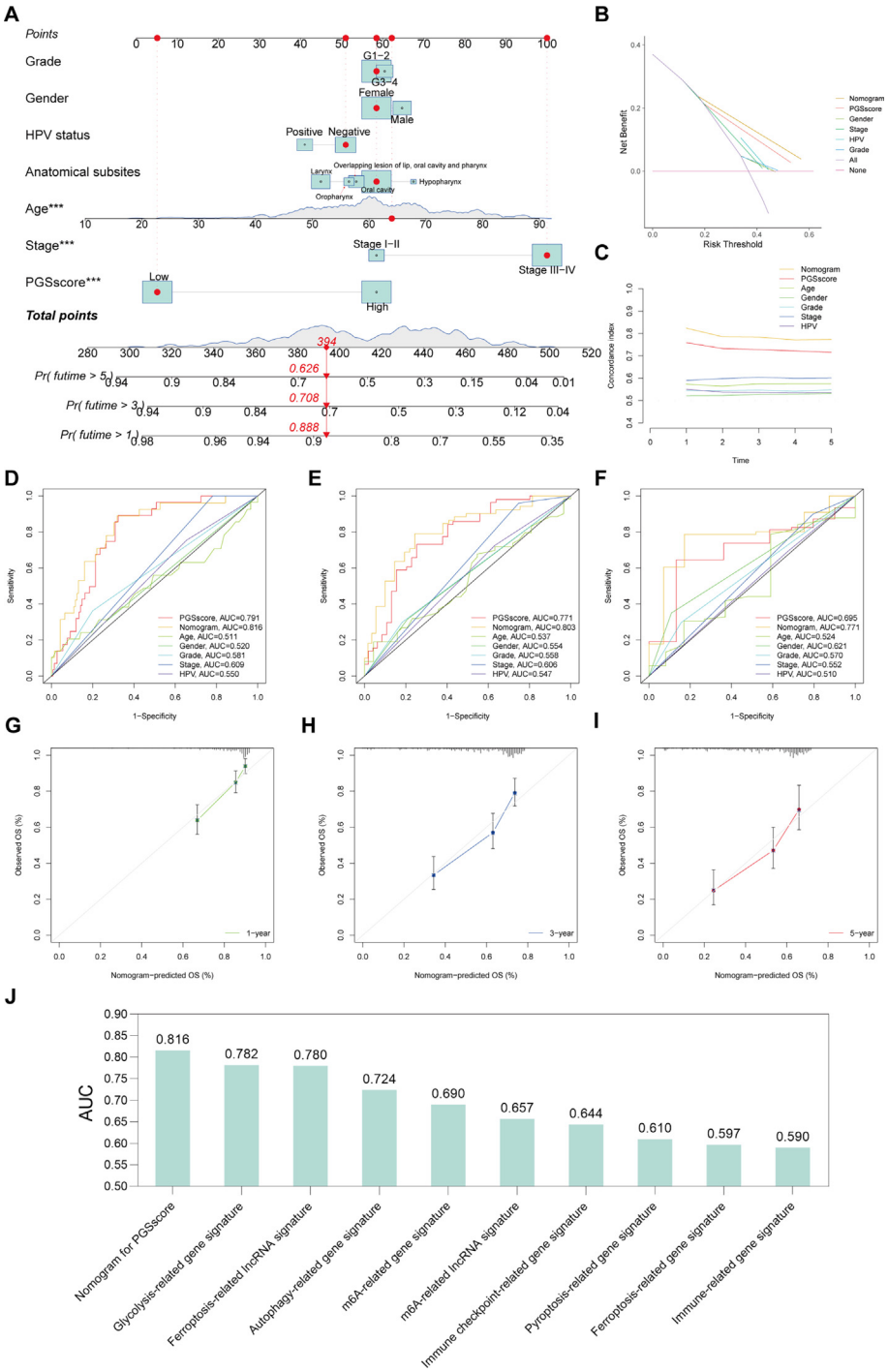


Fig. 11. Establishment and evaluation of a nomogram from the entire TCGA set. (A) A nomogram was established to predict overall survival after 1, 3, and 5 years. (B) Decision curve analysis of the nomogram compared with PGSscore and clinicopathologic features after 1 year. (C) C-index of the nomogram, PGSscore, and other clinicopathologic features after 1–5 years. (D-F) The ROC curve of the nomogram, PGSscore, and other clinicopathologic features predict overall survival after 1, 3, and 5 years. (G-I) Calibration plots of the nomogram predicting the 1-, 3-, 5-year overall survival compared with actual outcomes. (J) The 1-year AUC value of the nomogram based on the PGSscore (compared with existing signatures).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We would like to thank Editage (www.editage.cn) for English language editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.08.057.

References

[1] Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: globocan sources and methods. Int J Cancer 2019;144(8):1941–53. https://doi.org/10.1002/ijc.31937.

[2] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68(6):394–424. https://doi.org/10.3322/caac.21492.

[3] Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell 2004;5(5):489–500. https://doi.org/10.1016/s1535-6108(04)00112-6.

[4] De Cecco L, Nicolau M, Giannoccaro M, Daidone MG, Bossi P, Locati L, et al. Head and neck cancer subtypes with biological and clinical relevance: meta-analysis of gene-expression data. Oncotarget 2015;6(11):9627–42. https://doi.org/10.18632/oncotarget.3301.

[5] Walter V, Yin X, Wilkerson MD, Cabanski CR, Zhao N, Du Y, et al. Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes. PLoS ONE 2013;8(2):e56823.

[6] Wichmann G, Rosolowski M, Krohn K, Kreuz M, Boehm A, Reiche A, et al. The role of hpv rna transcription, immune response-related gene expression and disruptive Tp53 mutations in diagnostic and prognostic profiling of head and neck cancer. Int J Cancer 2015;137(12):2846–57. https://doi.org/10.1002/ijc.29649.

[7] Keck MK, Zuo Z, Khattri A, Stricker TP, Brown CD, Imanguli M, et al. Integrative analysis of head and neck cancer identifies two biologically distinct Hpv and three non-Hpv subtypes. Clin Cancer Res 2015;21(4):870–81. https://doi.org/10.1158/1078-0432.Ccr-14-2481.

[8] González-Moles MA, Bravo M, Ruiz-Avila I, Acebal F, Gil-Montoya JA, Brener S, et al. Ki-67 expression in non-tumour epithelium adjacent to oral cancer as risk marker for multiple oral tumours. Oral Dis 2010;16(1):68–75. https://doi.org/10.1111/j.1601-0825.2009.01611.x.

[9] González-Moles MA, Bravo M, Ruiz-Avila I, Gil-Montoya JA, Acebal F, Esteban F. E-cadherin in non-tumor epithelium adjacent to oral cancer as risk marker for the development of multiple tumors. Br J Oral Maxillofac Surg 2013;51(2):157–63. https://doi.org/10.1016/j.bjoms.2012.05.001.

[10] Bonomi M, Patsias A, Posner M, Sikora A. The role of inflammation in head and neck cancer. Adv Exp Med Biol 2014;816:107–27. https://doi.org/10.1007/978-3-0348-0837-8_5.

[11] Shen Y, Li X, Wang D, Zhang L, Li X, Xia T, et al. Novel prognostic model established for patients with head and neck squamous cell carcinoma based on pyroptosis-related genes. Transl Oncol 2021;14(12):. https://doi.org/10.1016/j.tranon.2021.101233101233.

[12] Zhu W, Ye Z, Chen L, Liang H, Cai Q. A pyroptosis-related lncrna signature predicts prognosis and immune microenvironment in head and neck squamous cell carcinoma. Int Immunopharmacol 2021;101(Pt B):. https://doi.org/10.1016/j.intimp.2021.108268108268.

[13] He D, Liao S, Xiao L, Cai L, You M, He L, et al. Prognostic value of a ferroptosis-related gene signature in patients with head and neck squamous cell carcinoma. Front Cell Dev Biol 2021;9:. https://doi.org/10.3389/fcell.2021.739011739011.

[14] Feng ZY, Gao HY, Feng TD. Immune infiltrates of M(6)a Rna methylation-related lncrnas and identification of Pd-L1 in patients with primary head and neck squamous cell carcinoma. Front Cell Develop Biol 2021;9:. https://doi.org/10.3389/fcell.2021.672248672248.

[15] Prat A, Navarro A, Paré L, Reguart N, Galván P, Pascual T, et al. Immune-related gene expression profiling after Pd-1 blockade in non-small cell lung carcinoma, head and neck squamous cell carcinoma, and melanoma. Cancer Res 2017;77(13):3540–50. https://doi.org/10.1158/0008-5472.Can-16-3556.

[16] Yang J, Jiang Q, Liu L, Peng H, Wang Y, Li S, et al. Identification of prognostic aging-related genes associated with immunosuppression and inflammation in head and neck squamous cell carcinoma. Aging 2020;12(24):25778–804. https://doi.org/10.18632/aging.104199.

[17] Jonathan RA, Wijffels KI, Peeters W, de Wilde PC, Marres HA, Merkx MA, et al. The prognostic value of endogenous hypoxia-related markers for head and neck squamous cell carcinomas treated with arcon. Radiother Oncol 2006;79(3):288–97. https://doi.org/10.1016/j.radonc.2006.04.008.

[18] Jin Y, Qin X. Development of a prognostic signature based on autophagy-related genes for head and neck squamous cell carcinoma. Arch Med Res 2020;51(8):860–7. https://doi.org/10.1016/j.arcmed.2020.09.009.

[19] Li Y, Zhu L, Yao H, Zhang Y, Kong X, Chen L, et al. Association of inflammation-related gene polymorphisms with susceptibility and radiotherapy sensitivity in head and neck squamous cell carcinoma patients in northeast china. Front Oncol 2021;11:. https://doi.org/10.3389/fonc.2021.651632651632.

[20] Su YW, Wu PS, Lin SH, Huang WY, Kuo YS, Lin HP. Prognostic value of the overexpression of fatty acid metabolism-related enzymes in squamous cell carcinoma of the head and neck. Int J Mol Sci 2020;21(18). https://doi.org/10.3390/ijms21186851.

[21] Wang L, Jia Y, Jiang Z, Gao W, Wang B. Fscn1 is upregulated by snai2 and promotes epithelial to mesenchymal transition in head and neck squamous cell carcinoma. Cell Biol Int 2017;41(8):833–41. https://doi.org/10.1002/cbin.10786.

[22] Zhang X, Hyer JM, Yu H, D'Silva NJ, Kirkwood KL. Dusp1 phosphatase regulates the proinflammatory milieu in head and neck squamous cell carcinoma. Cancer Res 2014;74(24):7191–7. https://doi.org/10.1158/0008-5472.Can-14-1379.

[23] Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy-analysis of affymetrix genechip data at the probe level. Bioinformatics 2004;20(3):307–15. https://doi.org/10.1093/bioinformatics/btg405.

[24] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (Msigdb) hallmark gene set collection. Cell systems 2015;1(6):417–25. https://doi.org/10.1016/j.cels.2015.12.004.

[25] Godec J, Tan Y, Liberzon A, Tamayo P, Bhattacharya S, Butte AJ, et al. Compendium of immune signatures identifies conserved and species-specific biology in response to inflammation. Immunity 2016;44(1):194–206. https://doi.org/10.1016/j.immuni.2015.12.006.

[26] Network T. The Immune Landscape of Cancer. (2018).

[27] Comprehensive Genomic Characterization of Head and Neck Squamous Cell Carcinomas. *Nature* (2015) 517(7536):576-82. doi: 10.1038/nature14129.

[28] Hänzelmann S, Castelo R, Guinney J. Gsva: gene set variation analysis for microarray and rna-seq data. BMC Bioinform 2013;14:7. https://doi.org/10.1186/1471-2105-14-7.

[29] Xu T, Le TD, Liu L, Su N, Wang R, Sun B, et al. Cancersubtypes: an r/bioconductor package for molecular cancer subtype identification, validation and visualization. Bioinformatics 2017;33(19):3131–3. https://doi.org/10.1093/bioinformatics/btx378.

[30] Gao J, Kwan PW, Shi D. Sparse kernel learning with lasso and bayesian inference algorithm. Neural Netw 2010;23(2):257–64. https://doi.org/10.1016/j.neunet.2009.07.001.

[31] Wilkerson MD, Hayes DN. Consensusclusterplus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 2010;26(12):1572–3. https://doi.org/10.1093/bioinformatics/btq170.

[32] Yu G, Wang LG, Han Y, He QY. Clusterprofiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16(5):284–7. https://doi.org/10.1089/omi.2011.0118.

[33] Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep 2017;18(1):248–62. https://doi.org/10.1016/j.celrep.2016.12.019.

[34] Hajiran A, Chakiryan N, Aydin AM, Zemp L, Nguyen J, Laborde JM, et al. Reconnaissance of Tumor immune microenvironment spatial heterogeneity in metastatic renal cell carcinoma and correlation with immunotherapy response. Clin Exp Immunol 2021;204(1):96–106. https://doi.org/10.1111/cei.13567.

[35] Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med 2018;24(10):1550–8. https://doi.org/10.1038/s41591-018-0136-1.

[36] Zeng D, Li M, Zhou R, Zhang J, Sun H, Shi M, et al. Tumor microenvironment characterization in gastric cancer identifies prognostic and immunotherapeutically relevant gene signatures. Cancer Immunol Res 2019;7(5):737–50. https://doi.org/10.1158/2326-6066.Cir-18-0436.

[37] Ascierto ML, McMiller TL, Berger AE, Danilova L, Anders RA, Netto GJ, et al. The intratumoral balance between metabolic and immunologic gene expression is associated with anti-pd-1 response in patients with renal cell carcinoma. Cancer Immunol Res 2016;4(9):726–33. https://doi.org/10.1158/2326-6066.Cir-16-0072.

[38] Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 2008;26(8):1364–70. https://doi.org/10.1200/jco.2007.12.9791.

[39] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26(6):565–74. https://doi.org/10.1177/0272989x06295361.

[40] Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. JAMA 2017;318(14):1377–84. https://doi.org/10.1001/jama.2017.12126.

[41] Fang J, Yang Z, Xie J, Li Z, Hu C, Yang M, et al. Identification and validation of autophagy-related prognostic signature for head and neck squamous cell carcinoma. Transl Oncol 2021;14(7):. https://doi.org/10.1016/j.tranon.2021.101094101094.

[42] Tang Y, Li C, Zhang YJ, Wu ZH. Ferroptosis-related long non-coding rna signature predicts the prognosis of head and neck squamous cell carcinoma. Int J Biol Sci 2021;17(3):702–11. https://doi.org/10.7150/ijbs.55552.

[43] Zhu W, Zhang J, Wang M, Zhai R, Xu Y, Wang J, et al. Development of a prognostic pyroptosis-related gene signature for head and neck squamous cell carcinoma patient. Cancer Cell Int 2022;22(1):62. https://doi.org/10.1186/s12935-022-02476-3.

[44] He F, Chen Z, Deng W, Zhan T, Huang X, Zheng Y, et al. Development and validation of a novel ferroptosis-related gene signature for predicting prognosis and immune microenvironment in head and neck squamous cell carcinoma. Int Immunopharmacol 2021;98:. https://doi.org/10.1016/j.intimp.2021.107789107789.

[45] Yang Z, Ming X, Huang S, Yang M, Zhou X, Fang J. Comprehensive analysis of M (6)a regulators characterized by the immune cell infiltration in head and neck squamous cell carcinoma to aid immunotherapy and chemotherapy. Front Oncol 2021;11:. https://doi.org/10.3389/fonc.2021.764798764798.
[46] Zhang D, Zheng Y, Yang S, Li Y, Wang M, Yao J, et al. Identification of a novel glycolysis-related gene signature for predicting breast cancer survival. Front Oncol 2020;10:. https://doi.org/10.3389/fonc.2020.596087596087.
[47] Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular Signatures Database (Msigdb) 30 2011;27:1739–40.
[48] Hung JH, Yang TH, Hu Z, Weng Z, DeLisi C. Gene set enrichment analysis: performance evaluation and usage guidelines. Brief Bioinform 2012;13 (3):281–91. https://doi.org/10.1093/bib/bbr049.
[49] Cillo AR, Kürten CHL, Tabib T, Qi Z, Onkar S, Wang T, et al. Immune landscape of viral- and carcinogen-driven head and neck cancer. Immunity 2020;52 (1):183-99.e9. https://doi.org/10.1016/j.immuni.2019.11.014.
[50] Leemans CR, Snijders PJF, Brakenhoff RH. The molecular landscape of head and neck cancer. Nat Rev Cancer 2018;18(5):269–82. https://doi.org/10.1038/nrc.2018.11.
[51] Sarrouilhe D, Clarhaut J, Defamie N, Mesnil M. Serotonin and cancer: what is the link? Curr Mol Med 2015;15(1):62–77. https://doi.org/10.2174/1566524015666150114113411.
[52] Dizeyi N, Hedlund P, Bjartell A, Tinzl M, Austild-Taskén K, Abrahamsson PA. Serotonin activates map kinase and Pi3k/Akt signaling pathways in prostate cancer cell lines. Urol Oncol 2011;29(4):436–45. https://doi.org/10.1016/j.urolonc.2009.09.013.
[53] Merzak A, Koochekpour S, Fillion MP, Fillion G, Pilkington GJ. Expression of serotonin receptors in human fetal astrocytes and glioma cell lines: a possible role in glioma cell proliferation and migration. Brain Res Mol Brain Res 1996;41(1–2):1–7. https://doi.org/10.1016/0169-328x(96)00058-7.
[54] Zamani A, Qu Z. Serotonin activates angiogenic phosphorylation signaling in human endothelial cells. FEBS Lett 2012;586(16):2360–5. https://doi.org/10.1016/j.febslet.2012.05.047.
[55] Woods LT, Jasmer KJ, Muñoz Forti K, Shanbhag VC, Camden JM, Erb L, et al. P2y (2) receptors mediate nucleotide-induced egfr phosphorylation and stimulate proliferation and tumorigenesis of head and neck squamous cell carcinoma cell lines. Oral Oncol 2020;109:. https://doi.org/10.1016/j.oraloncology.2020.104808104808.
[56] Woods LT, Forti KM, Shanbhag VC, Camden JM, Weisman GA. P2y receptors for extracellular nucleotides: contributions to cancer progression and therapeutic implications. Biochem Pharmacol 2021;187:. https://doi.org/10.1016/j.bcp.2021.114406114406.
[57] Zhang Y, Gallastegui N, Rosenblatt JD. Regulatory B cells in anti-tumor immunity. Int Immunol 2015;27(10):521–30. https://doi.org/10.1093/intimm/dxv034.
[58] Mahmoud SM, Lee AH, Paish EC, Macmillan RD, Ellis IO, Green AR. The prognostic significance of B lymphocytes in invasive carcinoma of the breast. Breast Cancer Res Treat 2012;132(2):545–53. https://doi.org/10.1007/s10549-011-1620-1.
[59] Pretscher D, Distel LV, Grabenbauer GG, Wittlinger M, Buettner M, Niedobitek G. 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. https://doi.org/10.1186/1471-2407-9-292.
[60] Teillaud JL, Dieu-Nosjean MC. Tertiary lymphoid structures: an anti-tumor school for adaptive immune cells and an antibody factory to fight cancer? Front Immunol 2017;8:830. https://doi.org/10.3389/fimmu.2017.00830.
[61] Wood O, Woo J, Seumois G, Savelyeva N, McCann KJ, Singh D, et al. Gene expression analysis of Til Rich Hpv-driven head and neck tumors reveals a distinct B-Cell signature when compared to Hpv independent tumors. Oncotarget 2016;7(35):56781–97. https://doi.org/10.18632/oncotarget.10788.
[62] Bron L, Jandus C, Andrejevic-Blant S, Speiser DE, Monnier P, Romero P, et al. Prognostic value of arginase-Ii expression and regulatory T-cell infiltration in head and neck squamous cell carcinoma. Int J Cancer 2013;132(3):E85–93. https://doi.org/10.1002/ijc.27728.
[63] Lukesova E, Boucek J, Rotnaglova E, Salakova M, Koslabova E, Grega M, et al. High level of tregs is a positive prognostic marker in patients with Hpv-positive oral and oropharyngeal squamous cell carcinomas. BioMed Res Int 2014;2014:. https://doi.org/10.1155/2014/303929303929.
[64] Lim KP, Chun NA, Ismail SM, Abraham MT, Yusoff MN, Zain RB, et al. Cd4 +Cd25hicd127low regulatory T cells are increased in oral squamous cell carcinoma patients. PLoS ONE 2014;9(8):e103975.
[65] Song D, Lyu H, Feng Q, Luo J, Li L, Wang X. Subtyping of head and neck squamous cell cancers based on immune signatures. Int Immunopharmacol 2021;99:. https://doi.org/10.1016/j.intimp.2021.108007108007.
[66] Woodford D, Johnson SD, De Costa AM, Young MR. An inflammatory cytokine milieu is prominent in premalignant oral lesions, but subsides when lesions progress to squamous cell carcinoma. J Clin Cell Immunol 2014;5(3). https://doi.org/10.4172/2155-9899.1000230.
[67] de Jong EC, Smits HH, Kapsenberg ML. Dendritic cell-mediated T cell polarization. Springer Semin Immunopathol 2005;26(3):289–307. https://doi.org/10.1007/s00281-004-0167-1.
[68] So T, Lee SW, Croft M. Tumor necrosis factor/tumor necrosis factor receptor family members that positively regulate immunity. Int J Hematol 2006;83 (1):1–11. https://doi.org/10.1532/ijh97.05120.
[69] Slootweg PJ, de Pagter M, de Weger RA, de Wilde PC. Lymphocytes at Tumor Margins in Patients with Head and Neck Cancer. Relationship with Tumor Size, Human Lymphocyte Antigen Molecules, and Metastasis. Int J Oral Maxillofac Surg 1994;23(5):286–9. https://doi.org/10.1016/s0901-5027(05)80110-9.
[70] Zhang S, Wang B, Ma F, Tong F, Yan B, Liu T, et al. Characteristics of B lymphocyte infiltration in Hpv(+) head and neck squamous cell carcinoma. Cancer Sci 2021;112(4):1402–16. https://doi.org/10.1111/cas.14834.
[71] Reyes-Gibby CC, Wang J, Silvas MR, Yu R, Yeung SC, Shete S. Mapk1/Erk2 as novel target genes for pain in head and neck cancer patients. BMC Genet 2016;17:40. https://doi.org/10.1186/s12863-016-0348-7.
[72] Sharif Siam MK, Sarker A, Sayeem MMS. In silico drug design and molecular docking studies targeting Akt1 (rac-alpha serine/threonine-protein kinase) and Akt2 (Rac-beta serine/threonine-protein kinase) proteins and investigation of Cyp (cytochrome P450) inhibitors against maob (monoamine oxidase B) for oscc (oral squamous cell carcinoma) treatment. J Biomol Struct Dyn 2021;39 (17):6467–79. https://doi.org/10.1080/07391102.2020.1802335.
[73] Chen YP, Wang YQ, Lv JW, Li YQ, Chua MLK, Le QT, et al. Identification and validation of novel microenvironment-based immune molecular subgroups of head and neck squamous cell carcinoma: implications for immunotherapy. Ann Oncol 2019;30(1):68–75. https://doi.org/10.1093/annonc/mdy470.