

Research article

TMErisk score: A tumor microenvironment-based model for predicting prognosis and immunotherapy in patients with head and neck squamous cell carcinoma

Yu Li ^{a,b,c,d,e,1}, Xiaozhou Pan ^{a,b,c,1}, Wenwei Luo ^{f,1}, Yaser Gamalla ^{g,h}, Zhan Ma ^{a,b,c}, Pei Zhou ^{g,***}, Chunfu Dai ^{e,**}, Dingding Han ^{a,b,c,g,i,*}

^a Department of Clinical Laboratory, Shanghai Children's Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, 200062, China

^b Institute of Pediatric Infection, Immunity, and Critical Care Medicine, School of Medicine, Shanghai Jiao Tong University, Shanghai, 200062, China

^c College of Health Science and Technology, School of Medicine, Shanghai Jiao Tong University, Shanghai, 200062, China

^d Department of Otolaryngology, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, 510000, China

^e Department of the Otolaryngology and Skull Base Surgery, Eye Ear Nose and Throat Hospital, Fudan University, Shanghai, 200031, China

^f Department of Otolaryngology-Head and Neck Surgery, Guangdong Provincial People's Hospital, Guang-dong Academy of Medical Sciences, Southern Medical University, Guangzhou, 510080, China

^g Laboratory of Medical Systems Biology, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, 510623, China

^h Department of Oncology, Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary, Alberta, T2N 4N1, Canada

ⁱ Medical School, Guangxi University, Nanning, 530004, China

ARTICLE INFO

Keywords:

HNSCC
TME
Tumor purity
Prognosis
Somatic mutation
Immunotherapy

ABSTRACT

Tumor microenvironment (TME) is closely associated with the progression and prognosis of head and neck squamous cell carcinoma (HNSCC). To investigate potential biomarkers for predicting therapeutic outcomes in HNSCC, we analyzed the immune and stromal status of HNSCC based on the genes associated with TME using the ESTIMATE algorithm. Immune and stromal genes were identified with differential gene expression and weighted gene co-expression network analysis (WGCNA). From these genes, 118 were initially selected through Cox univariate regression and then further input into least absolute shrinkage and selection operator (LASSO) regression analysis. As a result, 11 genes were screened out for the TME-related risk (TMErisk) score model which presented promising overall survival predictive potential. The TMErisk score was negatively associated with immune and stromal scores but positively associated with tumor purity. Individuals with high TMErisk scores exhibited decreased expression of most immune checkpoints and all human leukocyte antigen family genes, and reduced abundance of infiltrating immune cells. Divergent genes were mutated in HNSCC. In both high and low TMErisk score groups, the tumor protein P53 exhibited the highest mutation frequency. A higher TMErisk score was found to be associated with reduced overall survival probability and worse outcomes of

* Corresponding author. Department of Clinical Laboratory, Shanghai Children's Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, 200062, China.

** Corresponding author.

*** Corresponding author.

E-mail addresses: zhoupei0103@163.com (P. Zhou), cfdai66@126.com (C. Dai), handingding517@hotmail.com (D. Han).

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.heliyon.2024.e31877>

Received 8 May 2023; Received in revised form 22 May 2024; Accepted 23 May 2024

Available online 23 May 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

immunotherapy. Therefore, the TMErisk score could serve as a valuable model for the outcome prediction of HNSCC in clinic.

1. Introduction

HNSCC is the predominant cancer type, originating from the squamous epithelium of the oropharynx [1]. Annually, HNSCC affects more than 500,000 people worldwide, leading to about 380,000 deaths [2]. HNSCC can be categorized into four subtypes based on clinical, histological and molecular characteristics: atypical, basal, classical and mesenchymal [3]. One of the risk factors associated with HNSCC is HPV infection. On the basis of the tumor–node–metastasis (TNM) staging system, conventional treatments for HNSCC typically involve surgery, radiotherapy, and chemotherapy. In spite of advancements in multimodal intervention approaches, the 5-year overall survival (OS) rate for HNSCC patients remains limited to approximately 50 % [4]. Although recent genetic analyses have provided insights into novel gene markers [5–7], the identification of predictive or prognostic biomarkers in HNSCC is still an ongoing challenge.

The TME refers to the complex multicellular condition outside the developing tumor. This intricate environment is composed of immune cells, which play roles in both tumor suppression and promotion, stromal cells that provide structural support, and an extracellular matrix that contributes to tissue architecture. Additionally, the TME includes a range of secreted molecules that facilitate cell-to-cell communication and influence tumor progression [8]. Increasing evidence indicates that the TME is pivotal in regulating cancer progression and influencing prognosis [9]. Furthermore, molecules residing within the TME have been recognized as essential biomarkers for forecasting responses to treatments like radiotherapy and chemotherapy [10,11]. Therefore, assessing the TME condition of individuals is crucial for forecasting the prognosis of the illness and the effectiveness of treatments.

However, the quantitatively assessing the TME status in HNSCC is a challenging task. Fortunately, Yoshihara et al. have introduced a new algorithm called ESTIMATE (Estimation of STromal and Immune cells in Malignant Tumor tissues using Expression data). This algorithm enables the evaluation of the stromal and immune signatures, as well as the estimation of tumor purity [12]. In this study, leveraging large-scale HNSCC phenotype, survival, and gene mutation data from public databases, we utilized the ESTIMATE tool to construct a TMErisk score model for predicting prognosis in individuals with HNSCC.

2. Materials and methods

2.1. Data source

TCGA (The Cancer Genome Atlas) HNSCC data were downloaded from Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov/projects/TCGA-HNSC>). The subtype and HPV data of TCGA-HNSCC were available in the supplementary files of the associated study [3]. Gene expression data of melanoma (GSE91061) [13] were obtained from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). Gene expression data of IMvigor210 bladder cancer were extracted using “IMvigor210CoreBiologies” (version 1.0.0). Infiltrating immune cells were analyzed using the TIMER (<https://cistrome.shinyapps.io/timer/>) and xCell (<https://xcell.ucsf.edu>). All analyses were run in the R environment.

2.2. Estimate of immune and stromal scores

Using “estimate” (version 1.0.13) [12], the stromal and immune scores, as well as the tumor purity of the cancer samples in the training set were determined. Afterward, the samples were categorized into higher and lower score groups using the “surv_cutpoint” function from the “survminer” (version 0.4.8). The survival analysis was carried out utilizing the “survival” (version 3.2–7). The pan-cancer data for consensus measurement of purity estimations (CPE) were obtained from the supplementary data of Aran D et al.’s study [14].

2.3. Model construction with the intersections of DEGs and key modules

With the “limma” (version 3.42.2) [15], we screened differentially expressed genes (DEGs) based on the Benjamin–Hochberg adjusted P-value < 0.05 and $|\text{Log}_2(\text{Fold Change})| > 1$ as the threshold. The “WGCNA” (version 1.69) [16] was employed to identify key gene modules related to immune and stromal functions. In calculating its soft thresholding, $R^2 > 0.85$ was used as the screening threshold, and power = 7 was obtained. Furthermore, a one-step method was used to construct the network with power = 7, and modules with a height difference of less than 0.25 were combined. A total of 13 modules were finally selected. Gene modules with an absolute correlation coefficient greater than 0.3 were deemed to be significant association.

The intersection of DEGs and key module genes combined with OS data was analyzed using Cox regression to find genes significantly related to OS ($P < 0.05$). Furthermore, selected genes were served as input for the LASSO regression analysis to build the TMErisk score model, which was dependent on the “glmnet” (version 4.0–2) [17].

2.4. Function enrichment and immune signatures

The gene set enrichment analysis (GSEA) was carried out using “gseKEGG” and “gseGO” functions provided in the “cluster Profiler” (version 3.14.3). Gene lists of the immune checkpoint and HLA gene family were obtained from the supplementary data of Huang X et al.’s study [18] and Chen D et al.’s study [19], respectively. Furthermore, the enrichment scores were calculated using Gene Set Variation Analysis (GSVA) (version 1.34.0), and the proportion was acquired by using CIBERSORT (version 1.03).

2.5. Analysis of the tumor mutation status

Analysis of somatic mutations was performed with the “maftools” (version 1.0.2). Genes mutated in at least 20 samples were selected. Then, survival analysis was conducted to screen genes ($P < 0.05$) significantly associated with prognosis, and co-occurring mutations.

2.6. Evaluation of the response to therapy

The IC50 values of 138 drugs for the samples were inferred using the “pRRophetic” (version 0.5) [20]. The Wilcoxon test with the Bonferroni correction was employed to identify variations in IC50 levels between low- and high-TMERisk score categories. The

Table 1
The clinical characteristics of the dataset.

	TCGA_Train (382)		TCGA_Test (163)	
	Tumor (351)	Normal (31)	Tumor (150)	Normal(13)
Age (years)				
>60	181	19	75	8
≤60	170	12	75	5
Sex				
Male	259	19	108	11
Female	92	12	42	2
Pathologic_T				
T0	1	0	0	0
T1	34	2	11	1
T2	99	15	33	5
T3	64	7	32	2
T4	120	7	52	5
TX	22	0	11	0
NA	11	0	11	0
Stage				
stage i	18	1	7	1
stage ii	52	12	17	4
stage iii	53	7	25	1
stage iv	183	11	78	6
not reported	45	0	23	1
Grade				
G1	43	3	19	4
G2	127	16	82	7
G3	81	9	38	1
G4	1	0	1	1
GX	8	3	8	0
NA	1	0	2	0
Radiation_therapy				
YES	166	8	89	4
NO	106	5	32	6
NA	79	18	29	3
Smoking_history				
YES	264	22	116	10
NO	81	8	30	3
NA	6	1	4	0
Subtype				
Atypical	40	\	\	\
Basal	59	\	\	\
Classical	38	\	\	\
Mesenchymal	55	\	\	\
NA	159	\	\	\
HPV_Status				
Positive	23	\	\	\
Negative	169	\	\	\
NA	159	\	\	\

response to immunotherapy was deduced based on their tumor immune dysfunction and exclusion (TIDE) score (<http://tide.dfci.harvard.edu>) and immunophenoscore (IPS) (<https://tcia.at/home>). In general, superior responsiveness to immunotherapy is indicated by lower TIDE scores and elevated IPS values.

3. Results

3.1. Immune and stromal scores

To calculate immune and stromal scores using the ESTIMATE algorithm, our research included a total of 501 HNSCC patients alongside 44 normal individuals from the TCGA database. The samples were split in a 7:3 ratio for training and testing (Table 1). Analysis revealed that tumor samples exhibited higher immune and stromal scores compared to normal samples (Fig. 1A). Significant differences in both scores were also found across different tumor subtypes (Fig. 1B), and the immune score significantly increased in HPV-positive individuals compared with HPV-negative individuals (Fig. 1C). However, no difference could be detected when considering age, sex, tumor grade, radiation therapy, smoking history, pathologic_T, and tumor stage (Figure S1A–G). Kaplan-Meier curves showed that higher scores corresponded to better OS probabilities compared to those with lower scores (log-rank test) (Fig. 1D and E). Moreover, a strong inverse correlation was identified between both scores and tumor purity as inferred by ESTIMATE and CPE, respectively (Fig. 1F and G).

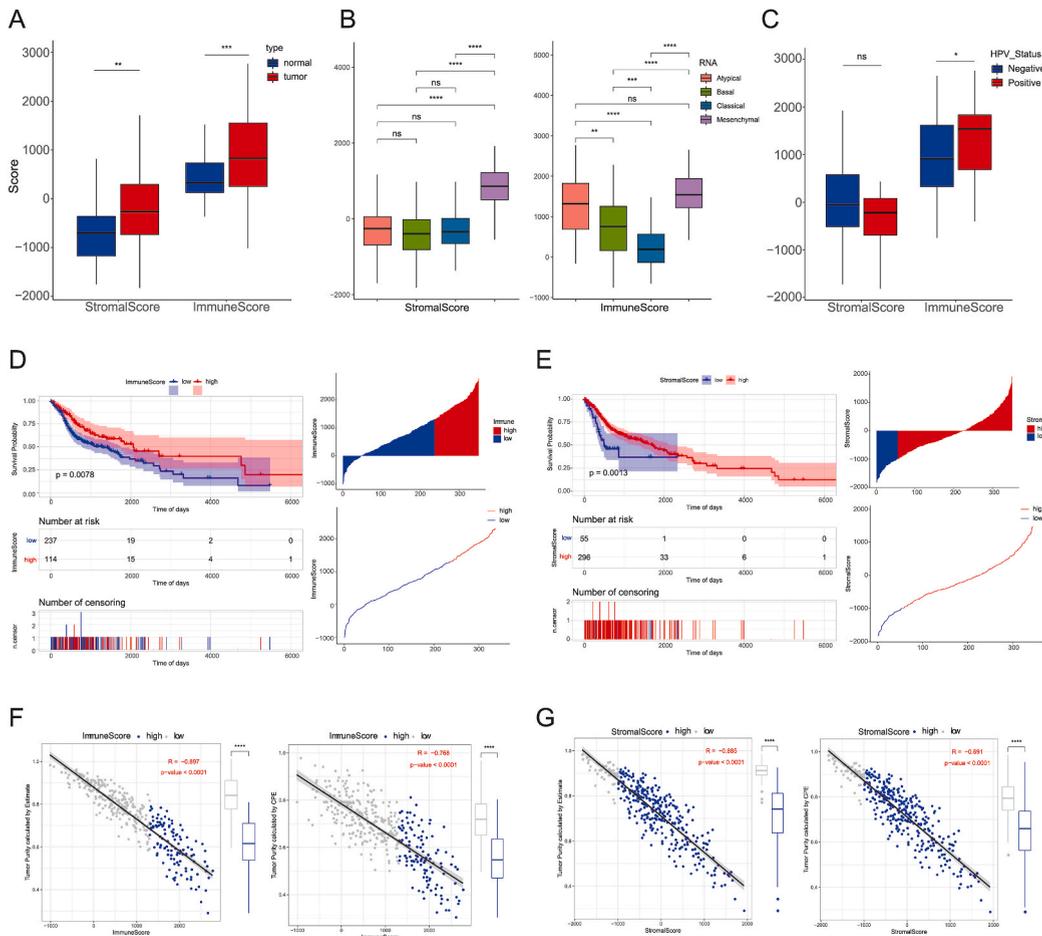


Fig. 1. Association of the immune and stromal scores and clinical characteristics in HNSCC. (A) Comparison of immune and stromal scores in tumor (n = 351) and normal (n = 31) samples. (Wilcoxon rank-sum test) (B) Immune and stromal scores in atypical (n = 40), basal (n = 59), classical (n = 38) and mesenchymal (n = 55) subtypes. (Wilcoxon rank-sum test) (C) Immune and stromal scores in HPV positive (n = 23) and negative (n = 169) individuals. (Wilcoxon rank-sum test) (D–E) Overall survival probability of patients with the high- and low-immune and stromal scores. (Log-rank test) (F–G) The Pearson’s correlation coefficient between immune and stromal scores and tumor purity calculated by ESTIMATE and CPE. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001; ns not significant.

3.2. Gene intersection of DEGs and key modules for model construction

A total of 316 DEGs related to immune scores and 355 DEGs related to stromal scores were identified (Benjamin–Hochberg adjusted P-value <0.05, |Log₂ (Fold Change)| > 1; Figure S2A, B, Table S1). In the WGCNA analysis, 13 modules were finally selected (Figure S3A–C, Table S2) from a pool of 19,035 protein-coding genes. The yellow, brown, pink, and blue modules were strongly correlated with the stromal score, while the tan and brown modules were strongly correlated with the immune score (|correlation coefficient| > 0.3; Figure S3D). Then, we intersected the DEGs and the module genes (Fig. 2A, Table S3). Using the univariate Cox regression, 118 genes that showed a strong correlation with OS were screened out (P < 0.05; Fig. 2B, Table S4). Subsequently, these genes underwent LASSO regression for dimensionality reduction. The lambda with the minimal standard error was selected with 10-fold cross validation, which was then used to estimate the coefficient of the genes (Figure S4). After excluding genes with a coefficient of zero, a final set of 11 genes (CCR7, WIPF1, CLEC4A, S1PR4, TENT5C, SFRP2, CALML5, CXCL9, DES, JCHAIN, and CKM) were finally

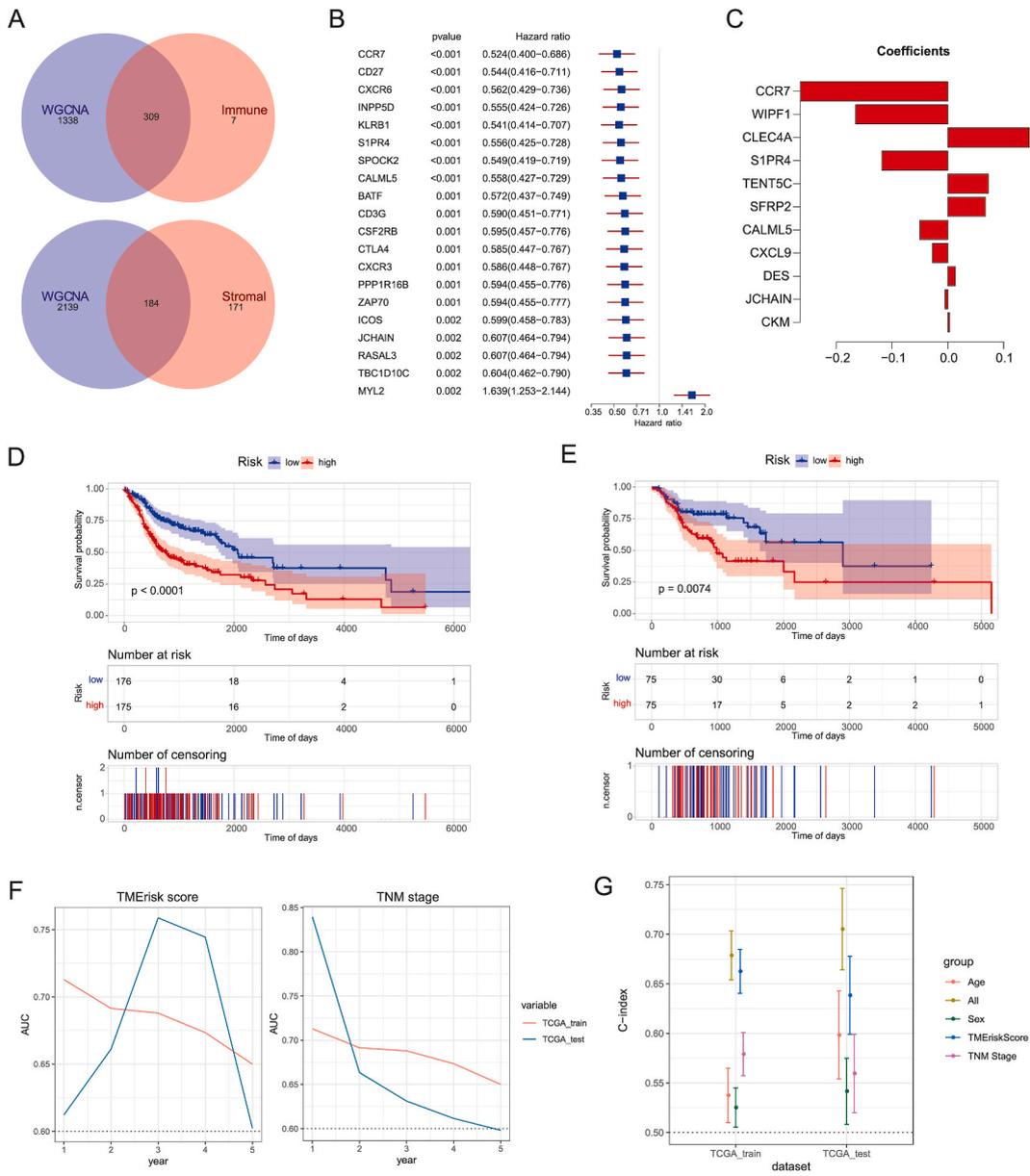
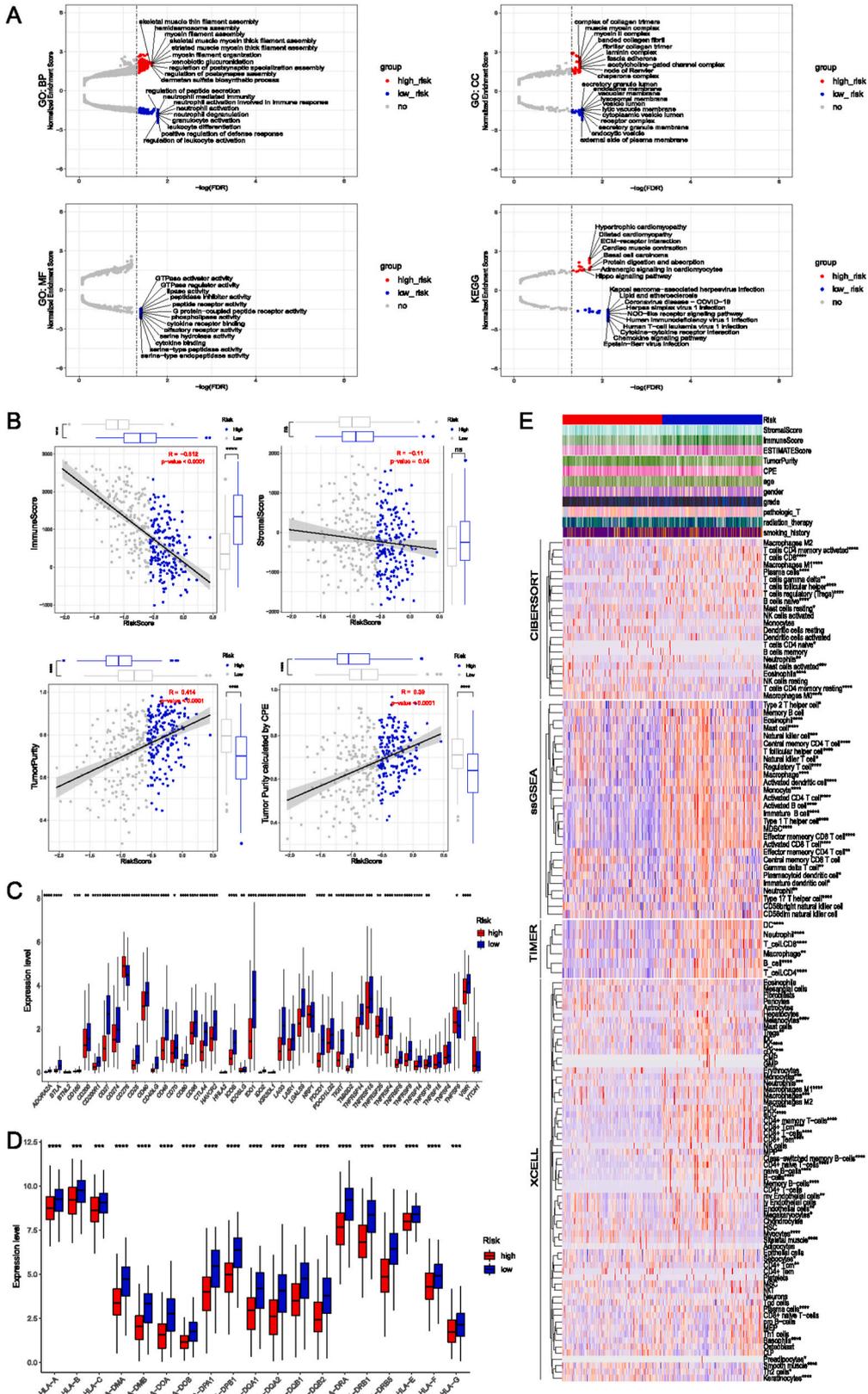


Fig. 2. TMErisk score model construction and validation. (A) The Venn diagram showing the overlap of DEGs associated with immune and stromal scores, along with TME-related module genes. (B) The top 20 genes exhibiting strong correlation with OS after the univariate Cox regression. (C) Regression coefficients of the 11 selected genes used for modeling. (D–E) OS prediction performance of the TMErisk score model in the training (left) and testing set (right). (Log-rank test) (F) Time dependent ROC curve of TMErisk score (left) and TNM stage (right) showing 5-year AUC values on the basis of OS data. (G) The C-index indicating the prediction accuracy of the TMErisk score, age, sex, and TNM stage.



(caption on next page)

Fig. 3. The association between the TMErisk score and the immune response. (A) GO and KEGG enrichment analysis including biological process (BP), cellular component (CC), molecular function (MF) and KEGG. Dashed lines represent FDR 0.05. The dot sizes correspond to the ratio of enriched genes. (B) The correlation of the TMErisk score with the immune and stromal score, tumor purity, as well as CPE, respectively. (C) The expression level of 45 immune checkpoints in the group with low and high TMErisk scores. (D) The expression level of HLA family genes in the group with low and high TMErisk scores. (E) The landscape of infiltrating immune cells in the low- and high-TMErisk score groups estimated by GSVA, CIBERSORT, TIMER, and xCell. (Wilcoxon rank-sum test, *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001).

screened out for modeling (Fig. 2C): TMErisk score = $-0.2647 \times \text{CCR7} - 0.1654 \times \text{WIPF1} - 0.1184 \times \text{S1PR4} - 0.0503 \times \text{CALML5} - 0.0276 \times \text{CXCL9} - 0.0049 \times \text{JCHAIN} + 0.0032 \times \text{CKM} + 0.0134 \times \text{DES} + 0.0675 \times \text{SFRP2} + 0.0726 \times \text{TENT5C} + 0.1462 \times \text{CLEC4A}$.

3.3. Prognostic performance of the TMErisk score model

The TMErisk score was significantly higher in HPV-negative patients. Differences were also observed among tumor subtypes, pathologic T stages, and TNM stages (Figure S5). Moreover, both univariate and multivariate cox regression analyses in the training and testing sets showed that the TMErisk score was an independent prognostic factor for OS (Figure S6A and S6B). To assess the model's effectiveness, participants were categorized according to the high or low-TMErisk score by the median. In the training set, individuals within the high-TMErisk score group exhibited notably worse OS, a result that was consistent in the testing set (Fig. 2D and E). Time-dependent ROC (receiver operating characteristic) curves can use additional information on onset time for each person to generate ROC curves at various time intervals. This allows us to compare the model's predictive ability and determine its validity over time. Thus, the predictive efficiency of the TMErisk scores model and TNM stage was further assessed based on the 5-year AUC (area under the ROC curve) values from OS data (Fig. 2F). The concordance index (C-index) is widely used for the global evaluation of discrimination power in prediction models. The accuracy of the TMErisk score model surpassed that of the TNM stage, which was inferred by the C-index. Moreover, the amalgamation of TMErisk score, age, sex, and TNM stage exhibited improved predictive performance (Fig. 2G). A stratified analysis of the TMErisk score was then conducted in subgroups of patients with different clinical indicators, such as sex, smoking history, radiation therapy choice, tumor subtype, HPV status, and TNM stage. We found that in most subgroups except for classical and mesenchymal subtypes, as well as HPV positive individuals, the TMErisk score made a substantial contribution to prognostic prediction (Figure S6C).

3.4. Association of the TMErisk score and the immune response in HNSCC

To investigate the underlying mechanisms accounting for the different TMErisk scores between groups, our study conducted GSEA analysis (Table S5). The high-TME risk score group exhibited enrichment of genes associated with collagen trimer complexes, interactions between extracellular matrix (ECM) and receptor interaction, laminin complexes, and basal cell carcinoma. Conversely, the low-TMErisk score group showed enrichment of genes involved in regulating leukocyte activation, granulocyte activation, and the chemokine signaling pathway (Fig. 3A).

Furthermore, the immune and stromal scores showed an inverse relationship with the TMErisk score, while tumor purity and the CPE exhibited a positive association with the TMErisk score (Fig. 3B). Immune checkpoints are vital for triggering immune responses against tumors, which are often modulated in the TME of various malignancies. The expression level of a total of 45 immune checkpoints in HNSCC was investigated, and significant expression difference of 39 genes were found between the TMErisk score groups. For example, CD276 and TNFSF9 were elevated in the high-TMErisk score category, whereas CD274, CTLA4, PDCD1, and LAG3 had significantly increased levels in the low-TMErisk score category (Fig. 3C). Moreover, our evaluation of 19 genes from the human leukocyte antigen (HLA) family indicated differential expression between these two groups, with most HLA family genes showing decreased expression in the group with high risk (Fig. 3D). Furthermore, the TMErisk score was negatively correlated with all HLA family genes as well as a most immune checkpoints. A majority of HLA family genes were positively correlated with immune checkpoints, except for HHLA2, TNFSF18, and VTCN1, which were negatively correlated with a subset of HLA family genes (Figure S7A).

Next, the proportion of infiltrating immune cells was estimated using tools, including GSVA, CIBERSORT, TIMER, and xCell. The analysis revealed that most immune cells were dramatically reduced when TMErisk scores were high (Fig. 3E). Then, the samples were divided based on the proportion of these cells estimated by CIBERSORT. In combination with OS data, analysis revealed that individuals with a higher proportion of CD8⁺ T cells, activated memory CD4⁺ T cells, and T follicular helper cells experienced improved outcomes. On the other hand, a higher proportion of eosinophils and activated mast cells was associated with worse OS probability (Figure S7B). Moreover, the distribution of these immune cell proportions in in groups with low and high TMErisk scores was consistent with the prognosis of individuals in these groups (Figure S7C).

3.5. Somatic mutation in groups with low and high TMErisk scores

To explore the genetic mechanisms driving the progression of HNSCC, we analyzed somatic mutations in samples from the low- and high-TMErisk score groups. Tumor mutation burden (TMB) represents total number of substitution, insertion or deletion mutations per megabase in the coding regions of tumor cell genomes. Higher TMB may lead to more neo-antigens for T-cell recognition, and associates with better outcomes in immunotherapy. Our findings revealed that the frequency of variations was marginally higher in the

category with low TMErisk score compared to that with high score (Fig. 4A). Not surprisingly, the gene with highest mutation frequency in HNSCC was the tumor protein P53 (TP53). Its mutation rates were 81 % and 56 % in the group with high and low TMErisk score, respectively. It is also noteworthy that Titin (TTN) gene ranked second in terms of mutation frequency, with a mutation rate around 36 %–37 % in both score groups. However, significant correlation was not found between the number of somatic mutations and TMErisk score (Figure S8A). Nonetheless, we identified 10 genes with mutations that were expected to have a significant negative impact on prognoses (Fig. 4B). Moreover, we found a higher co-mutation frequency between DMD and PAPP2 (Fig. 4C). Next, we analyzed the most significant risk genes, namely PNCT, DMD, and PAPP2. No mutation of these top 3 genes could be detected in more than half of the HNSCC samples, and the co-mutation frequency among them was low (< 5 %). The occurrence of single gene mutations and co-mutations did not differ significantly between groups with diverse TMErisk scores (Figure S8B). Additionally, PNCT exhibited a higher mutation frequency within the high-TMErisk score group. Although DMD had slightly more mutation sites in group with low TMErisk score, the number of mutation sites for PAPP2 remained nearly identical in both groups (Figure S8C).

3.6. Association of the TMErisk score and therapeutics

To evaluate the effectiveness of drug therapy in HNSCC patients across varying TMErisk scores, the IC50 for 138 drugs in TCGA-HNSCC patients were inferred, and the top 20 drugs with notable differences were identified. Among them, ABT-888, AICAR, ATRA,

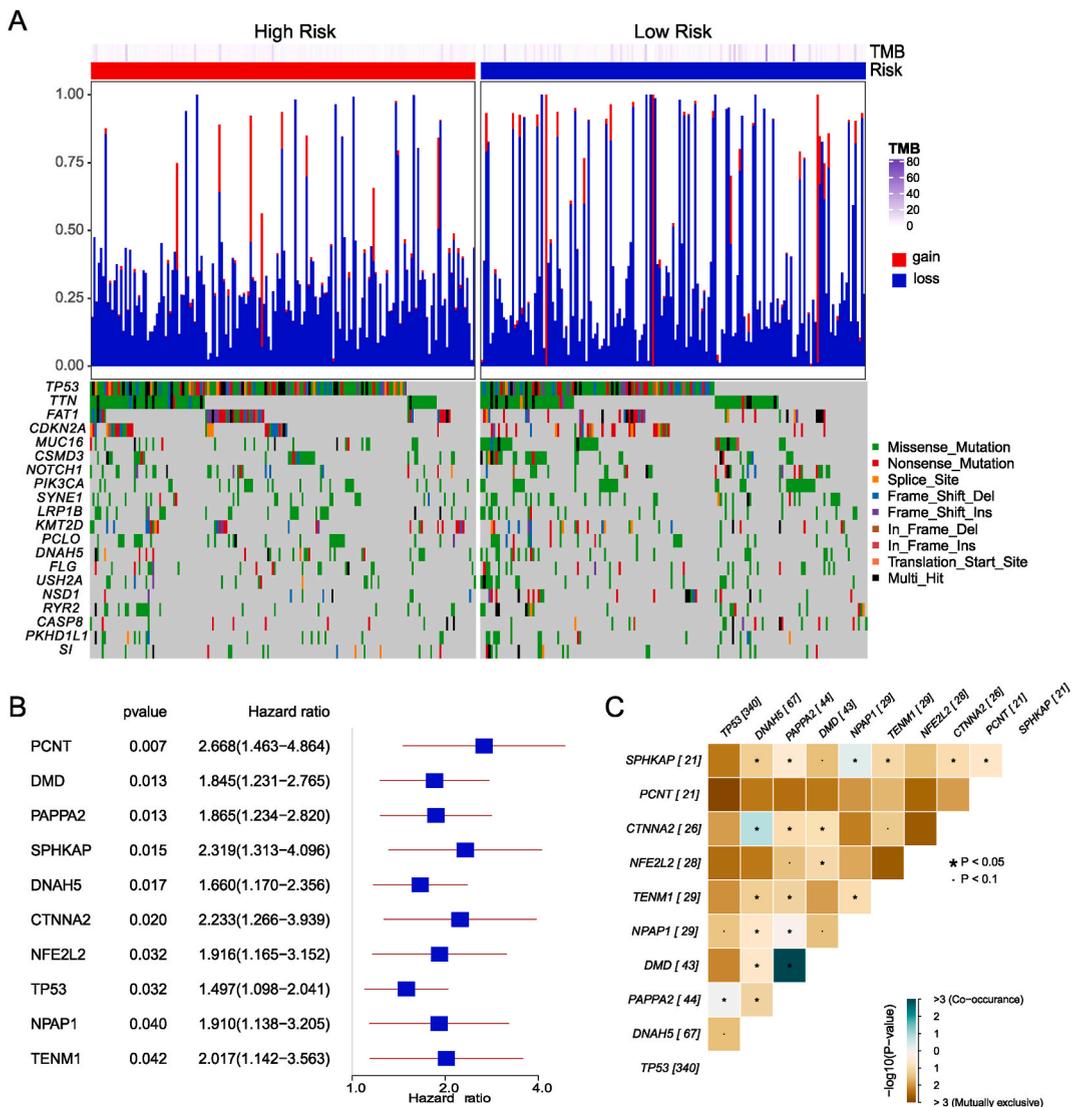


Fig. 4. The somatic mutation in the group with low and high TMErisk scores. (A) The TMB in the group with low and high TMErisk scores. (B) Cox regression analysis showing significant correlation of 10 mutant genes with prognosis. (C) Co-mutation situation of the 10 mutant genes. (Fisher’s exact test).

AZD6244, BIRB-0796, Gefitinib, JNK inhibitor VIII, lenalidomide, metformin, methotrexate, MK-2206, nilotinib, PLX4720, SL-0101-1, vorinostat, and VX-702 showed better efficacy for individuals with high-TMERisk scores, while the remaining drugs performed better in the group with low-TMERisk scores (Fig. 5A, Table S6). Moreover, the TIDE score was found to be positively correlated with the TMERisk score in both training and testing sets (Fig. 5B). Conversely, no substantial correlation existed between the IPS and the TMERisk score (Figure S9A).

To ensure the specificity of the TMERisk score model to HNSCC patients, we applied it to the bladder cancer and melanoma data. There were no remarkable variations in the immunotherapy response associated with groups with high and low TMERisk scores in bladder cancer (Figure S9B, C). In melanoma, the TMERisk scores were notably elevated in samples with progressive disease, as opposed to samples with stable disease and partial response/complete response, despite the absence of a significant alternation in immunotherapy response between two risk groups (Figure S9D, E).

4. Discussion

HNSCC is an extraordinarily heterogeneous malignancy caused by various etiologies including classical risk factors of smoking and excessive alcohol uptake, as well as the more recently identified risk of HPV infection [21]. As a result, early identification and comprehensive forecasting of outcomes are critical due to the elevated death rates associated with HNSCC. Additionally, researchers have noticed that intense therapies may cause a higher incidence of severe adverse reactions [22]. Thus, a personalized therapeutic response is needed to be considered to minimize toxicity without compromising treatment outcomes. Our findings revealed that the TMERisk score model serves as an effective predictor for both OS and therapeutic response in HNSCC patients.

The genes in the TMERisk score model are reported to involve in positive or negative regulation of the TME. For example, the

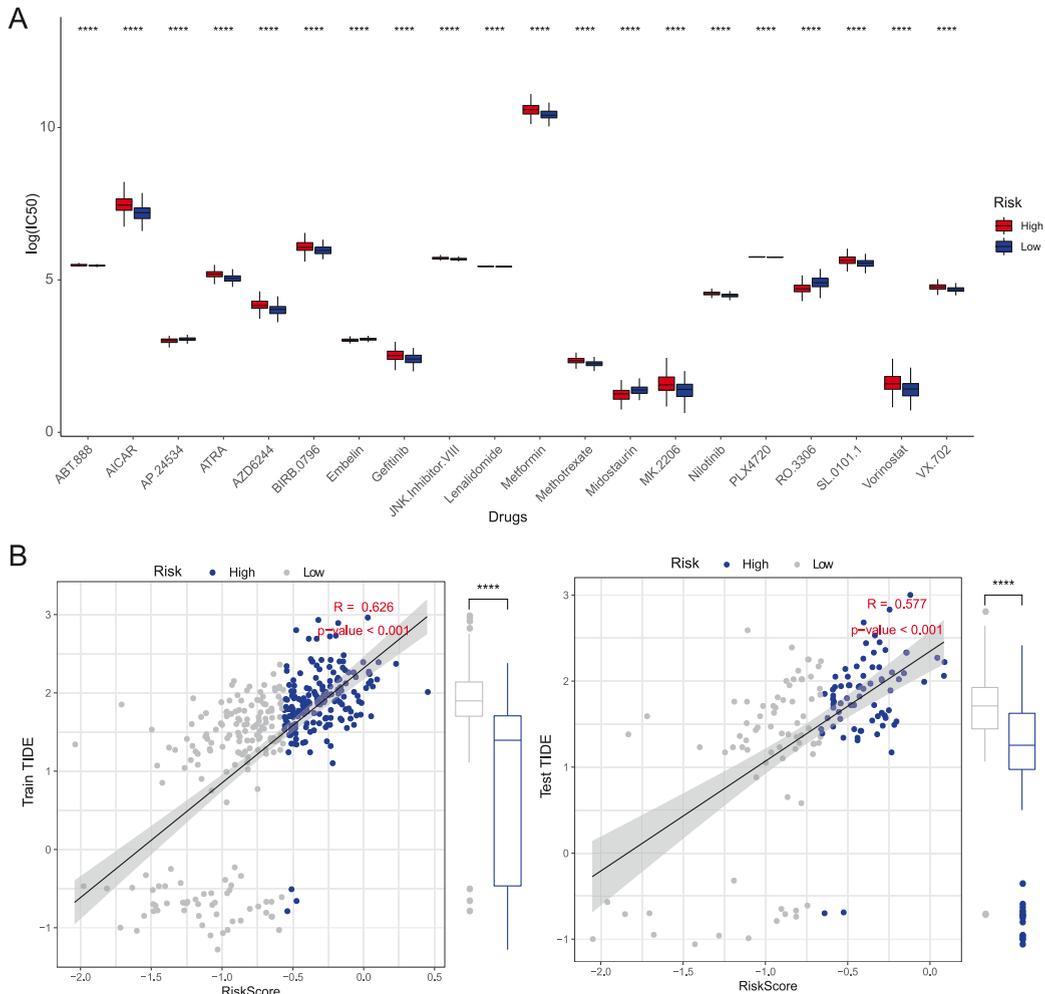


Fig. 5. Predicting therapeutic response with the TMERisk score. (A) The therapy efficacy of the top 20 medications showing notable variances between the low and highTMERisk score categories. (Wilcoxon rank-sum test) (B) Correlation between the TMERisk score and the TIDE score in the training (left) and testing set (right).

upregulation of CCR7, a chemokine receptor, in various cancers promotes the migration of cancer cell towards lymphatic nodes. However, it can also reactivate the immune response against tumors and is associated with survival in cervical cancer [23], making its role in therapeutic approaches a topic of ongoing discussion [24]. S1PR4 is associated with modulating myeloid cell activation, potentially impacting lymphocyte reactions [25]. A recent study exhibited that S1PR4 signaling might impede the response to chemotherapy through CD8⁺ T cells [26]. Another included gene is WIPF1, which crucially influences the organization and polymerization of the actin cytoskeleton, linked to the transition from epithelia to mesenchymal states [27]. Yu Pan et al. pinpointed that the inhibition of WIPF1 halts tumor progression and spread, suggesting its oncogenic properties [28]. CALML5, containing a calcium-binding motif, is expressed in the epidermis. Methylation of CALML5 is associated with the recurrence of HPV-associated oropharyngeal cancer [29]. The chemokine CXCL9 is involved in immune infiltration, whose predictive potential in the outcome and PD-1 therapy has been indicated in colorectal cancer [30], breast cancer [31] and uterine corpus endometrioid carcinoma [32]. Similarly, CLEC4A is also critical in tumorigenesis. In lung cancer and hepatocellular carcinoma, CLEC4 regulates the infiltration of various immune cells, making it a potential target for novel immunotherapy [33,34]. Recent reports have indicated that TENT5C, an unconventional poly(A) RNA polymerase known for boosting mRNA stability and gene expression, interferes with Plk4 function [35]. Although there is evidence suggesting that TENT5C could inhibit tumor growth and invasion, the precise mechanisms are not yet fully elucidated. Moreover, SFRP2, a secreted protein, was implicated in the advancement and unfavorable prognosis in breast cancer [36]. Nevertheless, roles of the other three genes (JCHAIN, CKM, and DES) in tumorigenesis have been limitedly investigated so far, indicating the need for further elucidation in future studies.

The TP53 gene is known to have a high mutation frequency in human cancers. Nearly all cancer types, including HNSCC, exhibit somatic TP53 mutations. In HNSCC, it reaches a rate of approximately 40.6 % [37]. In our study, the TP53 mutation rate is elevated in the high-TMERisk score group, suggesting its potential role in the prognosis of HNSCC. Interestingly, TTN ranked second among the genes with a high mutation frequency in our study. TTN is closely involved in all striated muscle cells and its mutations is primarily associated with cardiac phenotypes or skeletal muscle diseases [38]. Thus, further investigation is necessary to elucidate the association between TTN and HNSCC. Additionally, we found that the mutation of poor outcome-related genes was only detected in a small proportion of HNSCC samples and co-mutation was very rare. It may reflect the diversity of HNSCC pathogenesis, in which the role of divergent gene is inconspicuous.

Oropharyngeal cancers, in contrast to oral cavity and laryngeal cancers that are often linked to tobacco and alcohol use, are now increasingly linked to HPV infection [1]. In our study, a noticeable difference in the immune score was identified between the HPV-negative and HPV-positive categories. Moreover, our scoring model only showed the significant potential of prognostic prediction in HPV-negative patients, suggesting that HNSCC induced by HPV infection may occur in a specific way. The E6 and E7 genes are two essential oncogenes in HPV. They have a critical function in interacting with and controlling p53 and pRb, two major intracellular tumor suppressor proteins, thereby significantly affecting the cell proliferation and the repair of DNA [39]. In contrast, in HPV-negative HNSCC, TP53 is frequently deleted or mutated [3,40]. Currently, with global vaccination campaigns, especially the nine-valent vaccine, morbidity caused by HPV-positive HNSCC is expected to decline in the near future. Moreover, long-term exposure to Epstein-Barr virus (EBV) has been recognized to increase the risk of HNSCC originating from the nasopharynx [41]. However, because of the lack of information on EBV infection in our data, further stratified studies are needed to analyze the EBV-infected population using our model.

Using survival data, we computed the AUC values of the model for varying time periods ranging from 1 to 5 years. All the time-dependent AUC values were found to be greater than 0.6, demonstrating the reliable efficiency of the model. Moreover, the stratified analysis indicated that the scoring model exhibited predictive potential for survival outcomes across diverse subgroups, suggesting its broad utility. The C-index also demonstrated that the TMERisk score, either independently or in conjunction with age, sex, and TNM stage enhanced the prediction efficacy for overall prognosis when compared with the TNM stage. In general, these results indicate that TMERisk score is advantageous in clinical applications for evaluating the prognosis and OS of patients.

Immune status is believed to be critical in tumorigenesis. Consistent with this, the immune and stromal scores showed strong negative correlation with tumor purity, CPE, and the TMERisk score. Genes classified under the low-TMERisk score category were functionally enriched in pathways related to immune activation, which was not observed in the group with high risk score. Furthermore, the expression of immune checkpoints and HLA family genes dramatically changed, alongside a decrease of infiltrating immune cells in the group with high TMERisk score. These findings suggest a strong interplay between immune response and tumor cells in HNSCC, highlighting the potential of the TMERisk score for guiding precision medicine. Reports have shown that different responses to therapies are linked to diverse TME characteristics. Our study revealed specific reactions to medications in individuals with varying TMERisk levels. Furthermore, the TMERisk score indicated a significant correlation with TIDE, highlighting its potential for predicting the outcomes of immunotherapy. However, it should be noted that the predictive capacity of the TMERisk score, compared to TIDE, may be restricted to specific tumor types of HNSCC.

Nonetheless, our study has several limitations that need to be acknowledged. Due to the lack of published HNSCC datasets or the missing clinical data on the expression dataset, we had to divide our data into the testing set for model validation and the training set for model training and functional analysis. We attempted to use external data sources including E-MTAB-5793, GSE26549, GSE42743, GSE75538, and ICGC_ORCA-IN for validation. However, the analysis based on these data is not robust because of their small sample size. Furthermore, although the genes included in our model have been examined in various cancers, additional experimental research may reveal their significance in HNSCC. Nevertheless, in addition to exploring the functions and mechanisms of specific genes, a single gene did not show the outstanding ability to predict the prognosis as the combination of genes did. Hence, future validations of the TMERisk score model as a predictive tool on larger cohorts with comprehensive clinical data are expected to provide more contributions to HNSCC treatment and interventions.

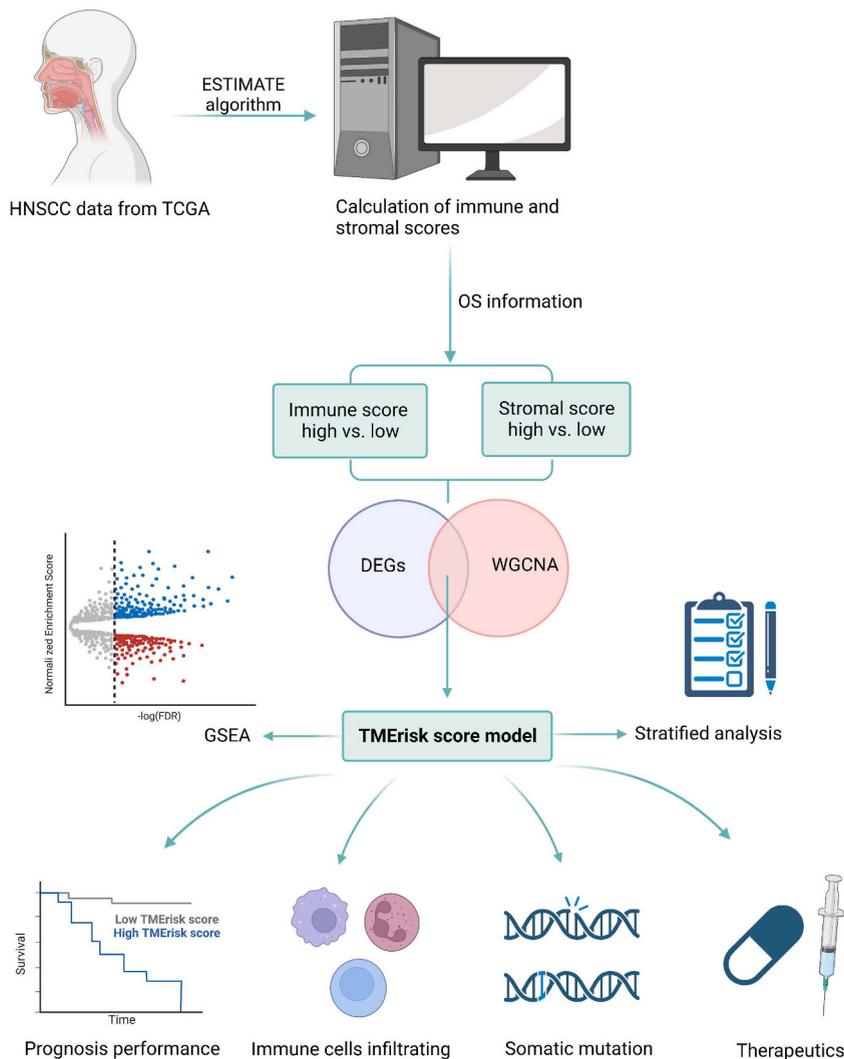


Fig. 6. Schematic diagram of TMErisk score model construction.

5. Conclusions

The study presented here developed a scoring model named TMErisk using TME gene expression data sourced from HNSCC individuals within the TCGA dataset. This scoring model, consisting of 11 genes, demonstrated strong predictive potential for overall survival. Additionally, we observed that individuals classified in the high-TMErisk score category demonstrated decreased levels of immune and stromal cell infiltration, altered genes expression patterns of HLA family and immune checkpoints, elevated tumor purity and somatic mutation burden when compared to counterparts classified in the group with low TMErisk scores (Fig. 6). The TMErisk score model we developed holds promise for predicting disease prognosis and guiding therapy for HNSCC patients.

Ethics statement

Review or approval by an ethics committee and informed consent were not needed for this study because all data used were acquired from public database which is provided without any identifier which would allow attribution of private information to an individual.

Funding

This work was supported by Natural Science Foundation of Shanghai (Project No. 21ZR1452900), National Natural Science Foundation of China (Project No. 82001121), National Natural Science Foundation of China (Project No. 82101208), Shanghai Municipal Key Clinical Specialty (Project No. shslczdk00801), and Guangzhou Basic and Applied Basic Research Foundation (Project

No. 202201011110 and 2022A1515220215).

Data availability statement

The data analyzed in this study are downloaded from the TCGA and GEO databases.

CRedit authorship contribution statement

Yu Li: Writing – original draft, Funding acquisition, Formal analysis. **Xiaozhou Pan:** Writing – original draft, Formal analysis. **Wenwei Luo:** Writing – review & editing, Data curation. **Yaser Gamalla:** Data curation. **Zhan Ma:** Writing – review & editing. **Pei Zhou:** Funding acquisition, Conceptualization. **Chunfu Dai:** Conceptualization. **Dingding Han:** Writing – original draft, Supervision, Investigation, Conceptualization.

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.

Acknowledgements

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e31877>.

References

- [1] D.E. Johnson, et al., Head and neck squamous cell carcinoma, *Nat. Rev. Dis. Prim.* 6 (1) (2020) 92.
- [2] F. Bray, et al., Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J Clin* 68 (6) (2018) 394–424.
- [3] N. Cancer Genome Atlas, Comprehensive genomic characterization of head and neck squamous cell carcinomas, *Nature* 517 (7536) (2015) 576–582.
- [4] C.R. Leemans, B.J. Braakhuis, R.H. Brakenhoff, The molecular biology of head and neck cancer, *Nat. Rev. Cancer* 11 (1) (2011) 9–22.
- [5] J. Wang, et al., Six-gene signature for predicting survival in patients with head and neck squamous cell carcinoma, *Aging (Albany NY)* 12 (1) (2020) 767–783.
- [6] B. Solomon, R.J. Young, D. Rischin, Head and neck squamous cell carcinoma: genomics and emerging biomarkers for immunomodulatory cancer treatments, *Semin. Cancer Biol.* 52 (Pt 2) (2018) 228–240.
- [7] Y. Chen, et al., An immune-related gene prognostic index for head and neck squamous cell carcinoma, *Clin. Cancer Res.* 27 (1) (2021) 330–341.
- [8] M.R. Junttila, F.J. de Sauvage, Influence of tumour micro-environment heterogeneity on therapeutic response, *Nature* 501 (7467) (2013) 346–354.
- [9] T. Wu, Y. Dai, Tumor microenvironment and therapeutic response, *Cancer Lett.* 387 (2017) 61–68.
- [10] Y. Sun, et al., Treatment-induced damage to the tumor microenvironment promotes prostate cancer therapy resistance through WNT16B, *Nat. Med.* 18 (9) (2012) 1359–1368.
- [11] B.J. Moeller, et al., Radiation activates HIF-1 to regulate vascular radiosensitivity in tumors: role of reoxygenation, free radicals, and stress granules, *Cancer Cell* 5 (5) (2004) 429–441.
- [12] K. Yoshihara, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, *Nat. Commun.* 4 (2013) 2612.
- [13] N. Riaz, et al., Tumor and microenvironment evolution during immunotherapy with nivolumab, *Cell* 171 (4) (2017) 934–949 e16.
- [14] D. Aran, M. Sirota, A.J. Butte, Systematic pan-cancer analysis of tumour purity, *Nat. Commun.* 6 (2015) 8971.
- [15] M.E. Ritchie, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res.* 43 (7) (2015) e47.
- [16] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinf.* 9 (2008) 559.
- [17] S. Engebretsen, J. Bohlin, Statistical predictions with glmnet, *Clin. Epigenet.* 11 (1) (2019) 123.
- [18] X. Huang, et al., Identification of tumor antigens and immune subtypes of pancreatic adenocarcinoma for mRNA vaccine development, *Mol. Cancer* 20 (1) (2021) 44.
- [19] D. Chen, et al., Identification and characterization of robust hepatocellular carcinoma prognostic subtypes based on an integrative metabolite-protein interaction network, *Adv. Sci.* 8 (17) (2021) e2100311.
- [20] P. Geeleher, N. Cox, R.S. Huang, pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels, *PLoS One* 9 (9) (2014) e107468.
- [21] C.R. Leemans, P.J.F. Snijders, R.H. Brakenhoff, The molecular landscape of head and neck cancer, *Nat. Rev. Cancer* 18 (5) (2018) 269–282.
- [22] M. Machtay, et al., Factors associated with severe late toxicity after concurrent chemoradiation for locally advanced head and neck cancer: an RTOG analysis, *J. Clin. Oncol.* 26 (21) (2008) 3582–3589.
- [23] W.J. Tian, et al., CCR7 has potential to be a prognosis marker for cervical squamous cell carcinoma and an index for tumor microenvironment Change, *Front. Mol. Biosci.* 8 (2021) 583028.
- [24] A. Salem, et al., CCR7 as a therapeutic target in Cancer, *Biochim. Biophys. Acta Rev. Canc* 1875 (1) (2021) 188499.
- [25] C. Dillmann, et al., S1PR4 signaling attenuates ILT 7 internalization to limit IFN-alpha production by human plasmacytoid dendritic cells, *J. Immunol.* 196 (4) (2016) 1579–1590.
- [26] C. Olesch, et al., S1PR4 ablation reduces tumor growth and improves chemotherapy via CD8+ T cell expansion, *J. Clin. Invest.* 130 (10) (2020) 5461–5476.
- [27] S.K. Donnelly, et al., WIP provides an essential link between Nck and N-WASP during Arp2/3-dependent actin polymerization, *Curr. Biol.* 23 (11) (2013) 999–1006.
- [28] Y. Pan, et al., WIPF1 antagonizes the tumor suppressive effect of miR-141/200c and is associated with poor survival in patients with PDAC, *J. Exp. Clin. Cancer Res.* 37 (1) (2018) 167.

- [29] K. Misawa, et al., Identification of novel methylation markers in HPV-associated oropharyngeal cancer: genome-wide discovery, tissue verification and validation testing in ctDNA, *Oncogene* 39 (24) (2020) 4741–4755.
- [30] Z. Wu, et al., The chemokine CXCL9 expression is associated with better prognosis for colorectal carcinoma patients, *Biomed. Pharmacother.* 78 (2016) 8–13.
- [31] Y.K. Liang, et al., CXCL9 is a potential biomarker of immune infiltration associated with favorable prognosis in ER-negative breast cancer, *Front. Oncol.* 11 (2021) 710286.
- [32] S. Xue, et al., CXCL9 correlates with antitumor immunity and is predictive of a favorable prognosis in uterine corpus endometrial carcinoma, *Front. Oncol.* 13 (2023) 1077780.
- [33] Y. Zhang, et al., CLEC4s as potential therapeutic targets in hepatocellular carcinoma microenvironment, *Front. Cell Dev. Biol.* 9 (2021) 681372.
- [34] N. Pakvisal, et al., Differential expression of immune-regulatory proteins C5AR1, CLEC4A and NLRP3 on peripheral blood mononuclear cells in early-stage non-small cell lung cancer patients, *Sci. Rep.* 12 (1) (2022) 18439.
- [35] K. Kazazian, et al., FAM46C/TENT5C functions as a tumor suppressor through inhibition of Plk4 activity, *Commun. Biol.* 3 (1) (2020) 448.
- [36] C. Huang, et al., Secreted frizzled-related protein 2 is associated with disease progression and poor prognosis in breast cancer, *Dis. Markers* 2019 (2019) 6149381.
- [37] M. Olivier, M. Hollstein, P. Hainaut, TP53 mutations in human cancers: origins, consequences, and clinical use, *Cold Spring Harbor Perspect. Biol.* 2 (1) (2010) a001008.
- [38] C. Chauveau, J. Rowell, A. Ferreiro, A rising titan: TTN review and mutation update, *Hum. Mutat.* 35 (9) (2014) 1046–1059.
- [39] C.B. Woodman, S.I. Collins, L.S. Young, The natural history of cervical HPV infection: unresolved issues, *Nat. Rev. Cancer* 7 (1) (2007) 11–22.
- [40] X. Castellsague, et al., HPV involvement in head and neck cancers: comprehensive assessment of biomarkers in 3680 patients, *J. Natl. Cancer Inst.* 108 (6) (2016) djv403.
- [41] C.M. Tsang, et al., Translational genomics of nasopharyngeal cancer, *Semin. Cancer Biol.* 61 (2020) 84–100.