Comprehensive analysis of ubiquitin-proteasome system genes related to prognosis and immunosuppression in head and neck squamous cell carcinoma

Juncheng Wang1,5,*, Jianing Li2, Luan Zhang2, Yuexiang Qin3, Fengyu Zhang1,5, Rulong Hu1,5, Huihong Chen1,5, Yongquan Tian1,5, Zhifeng Liu4,5,*, Yuxi Tian6, Xin Zhang1,4

1Department of Otolaryngology Head and Neck Surgery, Xiangya Hospital, Central South University, Changsha 410008, Hunan, P.R. China
2Department of Clinical Research, Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, Guangdong, P.R. China
3Department of Health Management, The Third Xiangya Hospital, Central South University, Changsha 410013, Hunan, P.R. China
4Department of Otorhinolaryngology, The First Affiliated Hospital of University of South China, Hengyang 421001, Hunan Province, P.R. China
5Otolaryngology Major Disease Research, Key Laboratory of Hunan Province, Changsha 410008, Hunan, P.R. China
6Department of Geriatrics, Respiratory Medicine, Xiangya Hospital, Central South University, Changsha 410008, Hunan, P.R. China
*Equal contribution

Correspondence to: Yuxi Tian, Xin Zhang; email: tianyuxi@csu.edu.cn, xinzhang@csu.edu.cn
Keywords: head and neck squamous cell carcinoma (HNSCC), ubiquitin proteasome system (UPS), the cancer genome atlas (TCGA) database, gene expression omnibus (GEO) database, immunosuppression
Received: February 11, 2021 Accepted: July 30, 2021 Published: August 16, 2021

Copyright: © 2021 Wang et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

The ubiquitin-proteasome system (UPS) with a capacity of degrading multiple intracellular proteins is an essential regulator in tumor immunosurveillance. Tumor cells that escape from recognition and destruction of immune system have been consistently characterized an important hallmark in the setting of tumor progression. Little know about the exact functions of UPS-related genes (UPSGs) and their relationships with antitumor immunity in head and neck squamous cell carcinoma (HNSCC) patients. In this study, for the first time, we comprehensively identified 114 differentially expressed UPSGs (DEUPSGs) and constructed a prognostic risk model based on the eight DEUPSGs (BRCA1, OSTM1, PCGF2, PSMD2, SOCS1, UCHL1, UHRF1, and USP54) in the TCGA-HNSCC database. This risk model was validated using multiple data sets (all P < 0.05). The high-risk score was found to be an independently prognostic factor in HNSCC patients and was significantly correlated with T cells suppression. Accordingly, our risk model can act as a prognostic signature and provide a novel concept for improving the precise immunotherapy for patients with HNSCC.

INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) affects more than 800,000 people each year worldwide, leading patients with HNSCC about sixth in global incidence [1, 2]. HNSCC is a collection of malignancies with complex local autonomy, given that it originates from the squamous epithelium of the upper aero-
digestive tract [3], attributing to this malignancy with high morbidity and mortality. Although advances in therapeutic strategies on local control have been improved, the 5-year overall survival (OS) rate remains 40%-50% [4, 5]. In this regard, there is a potential to identify novel biomarkers to broaden current treatment and enhance survival for patients with HNSCC.

The ubiquitin-proteasome system (UPS) is responsible for the degradation of more than 80% of intracellular proteins, such as short-lived and misfolded or non-essential proteins [6, 7]. UPS is involved in multiple biological functions, including gene transcription, translation and repair, cell cycle, cell proliferation, apoptosis, and antigen presentation [8, 9]. As UPS is involved in multitudinous critical functions, dysfunction of UPS may result in various diseases, such as nervous system diseases and cancers [10–14]. The UPS is a complex system encompassing a huge number of genes, including those encoding 10 ubiquitin (Ub)-activating enzymes (E1s), approximately 40 Ub-conjugating enzymes (E2s), over 600 validated Ub-protein ligases (E3s), approximately 100 deubiquitinating enzymes (DUBs), and nearly 50 proteasome subunits [15–18]. Accumulating studies have demonstrated that aberrant expression of UPS members altered proteolysis, facilitating the tumorigenesis and progression of HNSCC [19]. For instance, UBE2C (ubiquitin-conjugating enzyme E2 C), an E3s that overexpressed in various cancers and served as an inhibitor of p53 by facilitating its ubiquitination-mediated degradation. It is an agonist of ZEB1/2, ABCG2, and ERCC1 by promoting their transcript, which played an essential role in increasing cell proliferation and invasion, inducing epithelial-mesenchymal transitions (EMT), and chemotherapeutic resistance [20–23]. In addition, abnormal expression of PSMD2 (proteasome 26S subunit ubiquitin receptor, non-ATPase 2), a subunit of the 19S regulatory particle (RP) of the proteasome, has been reported in lung adenocarcinoma, breast cancer, and hepatocellular carcinoma [24–26]. The clinical significance of UPS genes has not been systematically investigated in patients with HNSCC. Therefore, the mechanism underlying the relationship between UPS-related genes (UPSGs) and the prognosis of HNSCC patients still needs to be further determined.

The Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) provide potent resources for directly obtaining gene expression profiles from patient tissues. Therefore, this study aims to provide a systematic investigation of the expression patterns via bioinformatics analysis and a prognostic risk model based on the UPSGs. We proposed that UPSGs are associated with the prognosis of patients with HNSCC. Importantly, our finding demonstrated that HNSCC patients with poor prognoses have high-risk scores that are associated with impaired T cell antitumor responses.

RESULTS

Identification of differentially expressed UPSGs in HNSCC tissues

We analyzed the expression of 804 UPSGs (Supplementary Table 1) that were distributed through all chromosomes in 498 head and neck tumor tissues and 44 adjacent tissues. 114 differentially expressed UPSGs (DEUPSGs) were identified, as shown in Supplementary Figure 1, where 97 upregulated and 17 downregulated DEUPSGs were observed (FDR < 0.05 and |Fold change| > 1.5, Figure 1A, 1B). The genomic information of DEUPSGs was shown in Supplementary Table 2.

Functional analysis of DEUPSGs in HNSCC

To assess the potential functions of DEUPSGs, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were analyzed using the TCGA database. The most significant GO terms were enriched for determining DEUPSGs functions, with respect to biological process (BP), cellular component (CC), and molecular function (MF), were shown in Figure 2A–2C. Enriched terms were significantly correlated with the proteasomal protein catabolic process, proteasome complex, and ubiquitin-like protein transferase activity. The KEGG pathway enrichment analysis indicated that ubiquitin-mediated proteolysis and proteasome pathways may play essential roles in the formation and development of HNSCC (Figure 2D).

Construction and identification of the prognostic risk model

To identify the prognostic value of DEUPSGs in HNSCC patients, Univariate Cox regression analysis was used to confirm the expression patterns of 114 DEUPSGs in the TCGA training set. The forest plots displayed the eight prognosis-related DEUPSGs in HNSCC (Figure 1C). The prognostic risk model was established based on the eight prognosis-related DEUPSGs using LASSO regression analysis (Supplementary Figure 2). The coefficient values of the eight DEUPSGs were shown in Table 1. The risk score was calculated for each sample as the following equation:

$$\text{Risk score} = BRCA1 \times (-0.037) + OSTM1 \times 0.0294 + PCGF2 \times 0.0502 + PSMD2 \times 0.0058 + SOCS1 \times (-0.0157) + UCHL1 \times 0.0022 + UHRF1 \times (-0.0496) + USP54 \times (-0.0887)$$
Samples were classified into two groups based on the median risk score (0.0622) of the training set. OS analysis indicated that the low-risk group has a markedly better prognosis than the high-risk group ($P < 0.0001$, Figure 3A). The receiver operating characteristic (ROC) curve indicated that the area under the ROC (AUC) value was 0.664 (Figure 3B). The risk scores and OS were performed by risk plots and scatter plots.

Figure 1. Differential expression of UPS-related genes (UPSGs) and identification of 8 UPSGs with prognostic value in HNSCC samples. (A) 114 differentially expressed UPSGs (DEUPSGs) are depicted as a heat map. (B) 97 upregulated and 17 downregulated DEUPSGs are shown as a volcano plot (FDR < 0.05 and |Fold change| > 1.5). (C) The eight risk DEUPSGs in the prognostic risk model are shown using a forest plot.
Expression patterns of risk genes in the high- and low-risk groups were depicted in Figure 3E, showing that high expression levels of PSMD2, OSTM1 (Osteoclastogenesis associated transmembrane protein 1), PCGF2 (polycomb group ring finger 2), and UCHL1 (ubiquitin C-terminal hydrolase L1) can be considered as risk factors were correlated with a high-risk score. Furthermore, high expression levels of BRCA1 (BRCA1 DNA repair associated), SOCS1 (suppressor of cytokine signaling 1), UHRF1 (ubiquitin-like with PHD and ring finger domains 1), and USP54 (ubiquitin specific peptidase 54) were associated with a low-risk score.

An independent validation data set was involved in the evaluation of the risk model. Based on the prognostic risk model, patients with HNSCC were divided into low- and high-risk groups. Survival analysis performed by Kaplan-Meier displayed significant prognostic differences between the two groups ($P < 0.001$, Supplementary Figure 3A). The relationship between the eight DEUPSGs and risk scores was shown in Supplementary Figure 3B. At the same time, similar results were found in the TCGA training set and TCGA all data set (Supplementary Figure 3C, 3D).

The GEO (GSE65858) database was used as an external data set for validating the classification performance of the risk model. Samples in the GSE65858 were segregated into low- and high-risk groups. The classification performance of the risk model and the expression pattern of the risk genes were consistent with the training set (Supplementary Figure 4).

**Clinical independence of the risk model**

To assess the independence of the risk model in clinical application, univariate and multivariate Cox regression analyses were subjected to clinical parameters from the TCGA training set, TCGA test set, GSE65858 database,
Table 1. List of the eight prognostic genes of the risk model in the TCGA training set.

<table>
<thead>
<tr>
<th>ENSG ID</th>
<th>Symbol</th>
<th>Location</th>
<th>Expression status</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSG00000012048</td>
<td>BRCA1</td>
<td>Chromosome 17</td>
<td>Upregulated</td>
<td>-0.0370</td>
</tr>
<tr>
<td>ENSG00000081087</td>
<td>OSTM1</td>
<td>Chromosome 10</td>
<td>Upregulated</td>
<td>0.0294</td>
</tr>
<tr>
<td>ENSG00000277258</td>
<td>PCGF2</td>
<td>Chromosome 17</td>
<td>Upregulated</td>
<td>0.0502</td>
</tr>
<tr>
<td>ENSG00000175166</td>
<td>PSMD2</td>
<td>Chromosome 3</td>
<td>Upregulated</td>
<td>0.0058</td>
</tr>
<tr>
<td>ENSG00000185338</td>
<td>SOCS1</td>
<td>Chromosome 16</td>
<td>Upregulated</td>
<td>-0.0157</td>
</tr>
<tr>
<td>ENSG00000154277</td>
<td>UCHL1</td>
<td>Chromosome 4</td>
<td>Upregulated</td>
<td>0.0022</td>
</tr>
<tr>
<td>ENSG00000276043</td>
<td>UHRF1</td>
<td>Chromosome 19</td>
<td>Upregulated</td>
<td>-0.0496</td>
</tr>
<tr>
<td>ENSG00000166348</td>
<td>USP54</td>
<td>Chromosome 10</td>
<td>Downregulated</td>
<td>-0.0887</td>
</tr>
</tbody>
</table>

and TCGA all data set. Univariate and multivariate Cox regression analyses suggested that the risk score had a significant association with prognosis in the TCGA training set (HR = 4.182, 95% CI = 2.540-6.887, P < 0.001; HR = 4.513, 95% CI = 2.732-7.375, P < 0.001, respectively, Figure 4A, 4B), the TCGA test set (HR = 1.993, 95% CI = 1.173-3.385, P < 0.05; HR = 1.954, 95% CI = 1.144-3.338, P < 0.05, respectively, Figure 4C, 4D), the TCGA all data set (HR = 2.947, 95% CI = 2.048-4.241, P < 0.01; HR = 2.961, 95% CI = 2.051-4.277, Figure 3.

Figure 3. Identification of the prognostic risk model in HNSCC patients. (A) Kaplan-Meier survival curve with overall survival (OS) in the high- and low-risk HNSCC patients in the TCGA training set. (B) ROC curve showing AUC for the risk score and other clinical factors of HNSCC patients in the TCGA training set. (C) The risk plot distribution of the high- and low-risk HNSCC patients. (D) Scatter plot showing the survival status of HNSCC patients. (E) The expression of risk genes of HNSCC samples in the TCGA training set.
Figure 4. Prognostic predictive value of risk score in HNSCC patients. (A, B) Univariate and multivariate Cox regression analyses of the clinical factors of the patients in the TCGA training set, respectively. (C, D) Univariate and multivariate Cox regression analyses of the clinical factors of the patients in TCGA test set, respectively. (E) Nomogram for OS in HNSCC patients.
$P < 0.01$, respectively, (Supplementary Figure 5A, 5B), and the GEO data set (HR = 1.754, 95% CI = 0.917-3.352, $P < 0.01$; HR = 1.598, 95% CI = 0.814-3.135, $P < 0.05$, respectively, Supplementary Figure 5C, 5D). These data demonstrated that the risk model has an effectively prognostic power and exhibits an independent predictive value. A nomogram contained both risk score and clinical features was shown in Figure 4E.

**Gene set enrichment analysis in the high-risk and low-risk groups**

Gene set enrichment analysis (GSEA) was employed for pathways identification in the TCGA database for detecting twenty pathways from the high- and low-risk groups (Supplementary Table 3). The selected signaling pathways showed significant differences between the two groups (FDR < 0.25, NOM $p < 0.05$) (Table 2), in which glycogen metabolism, extracellular matrix, and proteolysis were significantly enriched in the high-risk group (Figure 5A, 5B). In contrast, DNA repair, fatty acid metabolism, other metabolisms, and immune-related pathways were significantly enriched in the low-risk group (Figure 5C–5E). Interestingly, the T cell receptor (TCR) signaling pathway was enriched in the low-risk group (Figure 5F), which may indicate a potential relationship of the high-risk core with impaired immunosurveillance and T cell antitumor response in HNSCC. The rest of GSEA graphs were shown in Supplementary Figure 6.

**Exploration of the relationship between the risk score and antitumor immunity**

To determine whether the high-risk score was correlated with impaired T cell activity in HNSCC, we employed the ESTIMATE algorithm to estimate immune and stromal scores based on the TCGA database. The immune score was higher in the low-risk group than the high-risk group ($P < 0.001$, Figure 6A), and the risk score had a negative correlation with the immune score in HNSCC patients ($R = -0.19$, $P < 0.0001$, Figure 6B). However, the stromal score was higher in the high-risk group than the low-risk group ($P < 0.01$, Figure 6C), and the risk score had a positive correlation with the stromal score in HNSCC samples ($R = 0.17$, $P < 0.001$, Figure 6D).

We subsequently analyzed the fraction of both tumor-infiltrated innate and adaptive immune cells in HNSCC samples by CIBERSORT. The relationships of risk score with those immune cells were shown in Figure 6E. The results showed that the fraction of CD8 T cells ($P < 0.001$), CD4 memory activated T cells ($P < 0.001$), and follicular helper T cells ($P < 0.01$) in samples from the high-risk group were lower than the low-risk group, indicating that high-risk score was associated with immunosuppressive phenotypes that may result from impaired T cell response and activation.

**Correlation between the genes of the risk model and the four T cell subpopulations**

According to the potential relationship between the risk model and the four T cell subpopulations (Figure 7A), we further tested the correlation between the eight genes and the four T cell subpopulations (Figure 7B–7I). Consistent with the profile of these genes in the risk model, decreased CD8 T cells were associated with the upregulated OSTM1 ($P < 0.01$), PCGF2 ($P < 0.05$), PSMD2 ($P < 0.01$), UCHL1 ($P < 0.05$) and the downregulated SOCS1 ($P < 0.01$). Additionally, decreased CD4 memory-activated T cells were related to the upregulated UCHL1 ($P < 0.01$) and the downregulated SOCS1 ($P < 0.01$). Similarly, the asthenia of follicular helper T cells may be caused by high expressed OSTM1 ($P < 0.05$), PSMD2 ($P < 0.001$) and low expressed USP54 ($P < 0.001$). Moreover, the increase of CD4 memory resting T cells was associated with the high expressed OSTM1 ($P < 0.01$). Thus, the UPS-related genes OSTM1, PCGF2, PSMD2, UCHL1, SOCS1, and USP54 were associated with immunosuppression in HNSCC.

**DISCUSSION**

Tumor cells that escape from immunosurveillance and attack have been the characteristics of immunosuppression in the tumor progression, which is largely associated with the impaired or dysregulated host immune system [27]. By managing protein degradation and abundance, the UPS pathway has been reported as an essential regulator in immunosurveillance of the tumor microenvironment [28]. Accumulating evidence has demonstrated that E3 dysfunction perturbs antitumor immune responses [29–31]. Current knowledge on the functions of the UPS pathway in the development of HNSCC remains obscure. In this regard, an improved understanding of the UPSGs functions in the text of HNSCC can provide a therapeutic value.

This study analytically conferred prognostic evidence by identification of 114 DEUPGs and establishment of the eight prognosis-related DEUPGs. A prognosis risk model was established based on the eight DEUPGs. Previous studies have reported the abnormal expression of UPSGs in renal cancer and breast cancer [32, 33]. Analyses performed at both mRNA and protein levels can improve the reliability of the results, whereas the limited sample size of these analyses may lead to a bias.
Table 2. Gene sets enriched in high-risk and low-risk groups.

<table>
<thead>
<tr>
<th>MSigDB collection</th>
<th>Name</th>
<th>NES</th>
<th>ES</th>
<th>NOM p-val</th>
<th>FDR q-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>c2.cp.kegg.v7.1.symbols.gmt</td>
<td>KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_CHONDROITIN_SULFATE</td>
<td>2.099</td>
<td>0.813</td>
<td>0.000</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>KEGG_ECM_RECEPTOR_INTERACTION</td>
<td>1.903</td>
<td>0.676</td>
<td>0.010</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>KEGG_GLYCOSPHINGOLIPID_BIOSYNTHESIS_GANGLIO_SERIES</td>
<td>1.889</td>
<td>0.743</td>
<td>0.002</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>KEGG_OTHER_GLYCAN_DEGRADATION</td>
<td>1.854</td>
<td>0.705</td>
<td>0.006</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>KEGG_FOCAL_ADHESION</td>
<td>1.813</td>
<td>0.556</td>
<td>0.018</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_KERATAN_SULFATE</td>
<td>1.801</td>
<td>0.683</td>
<td>0.004</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_HEPARAN_SULFATE</td>
<td>1.719</td>
<td>0.563</td>
<td>0.010</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>KEGG_PROTEASOME</td>
<td>1.746</td>
<td>0.495</td>
<td>0.018</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>KEGG_DNA_REPLICATION</td>
<td>1.634</td>
<td>0.585</td>
<td>0.047</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td>KEGG_MISMATCH_REPAIR</td>
<td>-1.971</td>
<td>-0.810</td>
<td>0.000</td>
<td>0.143</td>
</tr>
<tr>
<td></td>
<td>KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY</td>
<td>-1.961</td>
<td>-0.796</td>
<td>0.004</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>KEGG_PANTOTHENATE_AND_COA_BIOSYNTHESIS</td>
<td>-1.841</td>
<td>-0.669</td>
<td>0.002</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>KEGG_HOMOLOGOUS_RECOMBINATION</td>
<td>-1.803</td>
<td>-0.690</td>
<td>0.016</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>KEGG_CYSSTEINE_AND_METHIONINE_METABOLISM</td>
<td>-1.792</td>
<td>-0.552</td>
<td>0.012</td>
<td>0.137</td>
</tr>
<tr>
<td></td>
<td>KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY</td>
<td>-1.773</td>
<td>-0.488</td>
<td>0.008</td>
<td>0.141</td>
</tr>
<tr>
<td></td>
<td>KEGG_ALPHALINOLENIC_ACID_METABOLISM</td>
<td>-1.726</td>
<td>-0.610</td>
<td>0.010</td>
<td>0.153</td>
</tr>
<tr>
<td></td>
<td>KEGG_LINOLEIC_ACID_METABOLISM</td>
<td>-1.712</td>
<td>-0.565</td>
<td>0.018</td>
<td>0.155</td>
</tr>
<tr>
<td></td>
<td>KEGG_BASE_EXCISION_REPAIR</td>
<td>-1.676</td>
<td>-0.620</td>
<td>0.045</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>KEGG_ARACHIDONIC_ACID_METABOLISM</td>
<td>-1.667</td>
<td>-0.476</td>
<td>0.014</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>KEGG_FATTY_ACID_METABOLISM</td>
<td>-1.660</td>
<td>-0.527</td>
<td>0.028</td>
<td>0.151</td>
</tr>
</tbody>
</table>

in the results, because only abnormal gene expression is analyzed, and verification of clinical significance or potential functional analysis is lacking.

In this risk model, high expression levels of PSMD2, OSTM1, PCGF2, and UCHL1 were risk factors, while high expression levels of BRCA1, SOCS1, UHRF1, and USP54 acted as protective factors. Overexpression of PSMD2 has been previously shown to be involved in the development of lung adenocarcinoma, breast cancer, and hepatocellular carcinoma [24–26]. UCHL1 has been reported a potential oncoprotein in colorectal cancer, breast cancer, and uterine serous cancer. It promotes proliferation and metastasis of cancer cells and leads to a radioresistant phenotype via regulation of β-catenin/TCF and HIF-1α pathway [34–38]. In line with those previous studies, our results also suggested that PSMD2 and UCHL1 may have therapeutic potential as targets against cancer. Accumulating evidence has shown that BRCA1 mutation is a crucial risk factor in multiple cancers, including breast cancer, skin cancer, ovarian cancer, and colorectal cancer [39–43]. SOCS1 acts as an antioncogene in various tumors, arresting the cell cycle, inhibiting cancer cell migration and invasion, and attenuating tumor growth [44–47]. Nevertheless, non-coding RNA (ncRNA) regulating SOCS1 promotes the occurrence and development of cancers [48–50]. Similarly, UHRF1 also acts as an antioncogene in multiple cancers and inhibits tumor development and progression [51–54]. Therefore, the abovementioned studies suggest that BRCA1, SOCS1, and UHRF1 serve as antioncogenes, consistent with our results.

Interestingly, our GSEA results showed that the TCR signaling pathway was highly enriched in the low-risk group, which suggested the impaired T cell activation presented in the high-risk group. TCR engagement initiates a central signaling pathway that is crucial for T cell proliferation, survival, and differentiation into killer cells for adaptive immunity. Abnormalities of TCR signaling could result in immunodeficiency [55–58]. Consequently, according to the relationship between the risk score and the distribution of T cell subpopulations, we found that the high-risk score has a decreased immune score, which may suggest that the high-risk score is critical for evaluating whether tumor-infiltrated T cells have an immunosuppressive status and dysregulated antitumor response. Moreover, the high expression of OSTM1, PCGF2, PSMD2 combined with the low expression of UCHL1, SOCS1, and USP54 were engaged in the suppression of T cell proliferation and activation in the HNSCC microenvironment. It has been previously shown that differential expression of UCHL1 was majorly involved in the proliferation and differentiation of T cells, including CD8 memory CTLp, circulating TCR-gamma/delta+ lymphocytes, and mature T cells [59]. SOCS1 has been identified as
Figure 5. GSEA analysis showing the enriched pathways of the high- and low-risk groups. (A) Multiple GSEA showing glycan-related metabolism pathways in the high-risk group. (B) Multiple GSEA showing extracellular matrix and proteolysis related pathways in the high-risk group. (C) Multiple GSEA showing DNA repair in the low-risk group. (D) Multiple GSEA showing fatty acid metabolism pathways in the low-risk group. (E) Multiple GSEA showing other metabolism- and immune-related pathways in the low-risk group. (F) Single GSEA showing the T cell receptor signaling pathway.
Figure 6. Association between risk score and tumor immunity. (A) Distribution of immune scores in high- and low-risk HNSCC patients. (B) Association between the risk score and immune score in HNSCC samples. (C) Distribution of stromal scores in high- and low-risk HNSCC patients. (D) Association between the risk score and stromal score in HNSCC samples. (E) Comparison of immune cell fractions between the high-risk and low-risk HNSCC patients.
Figure 7. Correlation of the genes of the risk model with the four subpopulations of T cells. (A) Comparison of the four subpopulations of T cells (CD8 T cells, CD4 memory activated T cells, and follicular helper T cells) between the high- and low-risk groups. (B–I) Distribution of the four T cell subpopulations based on the high and low expression of BRCA1, OSTM1, PCGF2, PSMD2, SOCS1, UCHL1, UHRF1, and USP54, respectively.
an inhibitor against cytokine release and a growth factor receptor, it plays a crucial role in regulating T cell homeostasis, development, and homeostasis activation [60–64]. However, the function of OSTM1, PCGF2, PSMD2, and USP54 have not been investigated in the antitumoral immune response. The present studies suggest that the high expression of OSTM1, PCGF2, and PSMD2 and low expression of UCHL1 were associated with immunosuppressive phenotypes.

Generally, we identified an eight -UPSGs based risk model according to the TCGA-HNSCC database and analyzed its biological functions, demonstrating that the risk score obtained from this model was significantly correlated with immunosuppressive status. However, some limitations in the context of the study should be acknowledged. First, this study is performed by bioinformatics analysis alone, the predictive results could be insufficient. Second, this is a retrospective study rather than a prospective one, further validations with large clinical cohort and actual experiments are needed.

CONCLUSIONS

In summary, we present a risk model constructed by DEUPSGs that can be considered as potential prognostic biomarkers and associated with an immunosuppressive status and impaired antitumor response of T cells in the HNSCC microenvironment. This systematic analysis on the interaction of UPSGs based risk model and immune profile provides a novel understanding of the precise immunotherapy for HNSCC patients.

MATERIALS AND METHODS

Data downloading and processing

The most recent transcriptional data and clinical features on 500 HNSCC samples and 44 adjacent samples were obtained from the TCGA database (https://portal.gdc.cancer.gov/) and the cBio Cancer Genomics portal (https://www.cbioportal.org/) (Table 3). We chose 498 HNSCC samples with follow-up data and randomly divided them into two groups: a training set (n=298) and a test set (n=298), as shown in Supplementary Tables 4, 5. In addition, we obtained the GSE65858 data set with 270 HNSCC samples as an external set from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/).

Identification of differentially expressed genes

The differentially expressed UPS-related genes (DEUPSGs) between head and neck tumor tissues and adjacent tissues were screened out through the Wilcoxon signed-rank test using the limma R package, according to the following cut-off values: FDR < 0.05 and |Fold change| > 1.5. Subsequently, a heat map was generated using the “heatmap” package, and volcano dot plots were created to present the DEUPSGs. We used the OmicCircos R package to show the distribution of the DEUPSGs [65].

Functional enrichment analyses

The enrichplot R package was built for the GO and KEGG pathway enrichment analyses to identify the functions of the abovementioned DEUPSGs [66], containing BP, CC, and MF terms and pathways. Statistical significance was defined as both P-value and FDR < 0.05.

Prognostic risk model establishment

We used univariate Cox regression analysis to screen out DEUPSGs significantly associated with OS. P < 0.05 was chosen as the threshold. We further used Lasso regression analysis to establish a multi-gene prognostic model. The risk score of each patient was calculated, employing the regression coefficient of each selected gene based on the following equation:

\[ \text{Risk score} = \sum_{i=1}^{n} \text{Coef}_i \times \text{Exp}_i, \]

where \( n \) is the number of prognostic genes, \( \text{gene}_i \) is the \( i \)th prognostic gene, \( \text{Coef}_i \) is the regression coefficient of genes, and \( \text{Exp}_i \) is the expression value of the prognostic genes. Then, the HNSCC patients were divided into high-risk and low-risk groups according to the median risk score. Moreover, we used rms package to generate a nomogram.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA, https://www.gsea-msigdb.org/) is a powerful computational tool used to determine statistical differences of specific functional gene sets between two biological states [67]. Here, we performed the GSEA analysis using GSEA v4.0.3 to obtain the differentially expressed genes from the high- and low-risk groups. After 1000 repeats for each analysis, gene sets with \( p \)-value < 0.05 and FDR < 0.25 were identified as enriched sets.

Identification of immune scores and tumor-infiltrating immune cells

To identify infiltrated immune cell numbers and stromal cell numbers, the R software package “ESTIMATE algorithm” was used for calculating immune and
Table 3. Clinical characteristics of HNSCC patients in the TCGA and GEO databases.

<table>
<thead>
<tr>
<th>Clinical characteristics</th>
<th>TCGA (n=500)</th>
<th>%</th>
<th>GEO (GSE65858) (n=270)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 60</td>
<td>220</td>
<td>44.0</td>
<td>153</td>
<td>56.7</td>
</tr>
<tr>
<td>≥ 60</td>
<td>280</td>
<td>56.0</td>
<td>117</td>
<td>43.3</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>133</td>
<td>26.6</td>
<td>47</td>
<td>17.4</td>
</tr>
<tr>
<td>Male</td>
<td>367</td>
<td>73.4</td>
<td>223</td>
<td>82.6</td>
</tr>
<tr>
<td>Histologic grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>61</td>
<td>12.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>299</td>
<td>59.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G3</td>
<td>119</td>
<td>23.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G4</td>
<td>2</td>
<td>0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gx</td>
<td>16</td>
<td>3.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>3</td>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>19</td>
<td>3.8</td>
<td>18</td>
<td>6.7</td>
</tr>
<tr>
<td>II</td>
<td>95</td>
<td>19.0</td>
<td>37</td>
<td>13.7</td>
</tr>
<tr>
<td>III</td>
<td>102</td>
<td>20.4</td>
<td>37</td>
<td>13.7</td>
</tr>
<tr>
<td>IV</td>
<td>270</td>
<td>54.0</td>
<td>178</td>
<td>65.9</td>
</tr>
<tr>
<td>NA</td>
<td>14</td>
<td>2.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>33</td>
<td>6.6</td>
<td>35</td>
<td>13</td>
</tr>
<tr>
<td>T2</td>
<td>143</td>
<td>28.6</td>
<td>80</td>
<td>29.6</td>
</tr>
<tr>
<td>T3</td>
<td>130</td>
<td>26.0</td>
<td>58</td>
<td>21.5</td>
</tr>
<tr>
<td>T4</td>
<td>179</td>
<td>35.8</td>
<td>97</td>
<td>35.9</td>
</tr>
<tr>
<td>Tx</td>
<td>11</td>
<td>2.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>4</td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>239</td>
<td>47.8</td>
<td>94</td>
<td>34.8</td>
</tr>
<tr>
<td>N+</td>
<td>239</td>
<td>47.8</td>
<td>176</td>
<td>65.2</td>
</tr>
<tr>
<td>Nx</td>
<td>18</td>
<td>3.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>4</td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>470</td>
<td>94.0</td>
<td>263</td>
<td>97.4</td>
</tr>
<tr>
<td>M1</td>
<td>5</td>
<td>1.0</td>
<td>7</td>
<td>2.6</td>
</tr>
<tr>
<td>Mx</td>
<td>20</td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>5</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceased</td>
<td>218</td>
<td>43.6</td>
<td>94</td>
<td>34.8</td>
</tr>
<tr>
<td>Living</td>
<td>282</td>
<td>56.4</td>
<td>176</td>
<td>65.2</td>
</tr>
</tbody>
</table>

stromal scores of HNSCC samples in the TCGA database [68]. Additionally, CIBERSORT was applied to acquire the composition of the infiltrated immune cells in HNSCC samples [69].

Statistical analysis
R package (v4.0.1) was subjected to all analyses. Kaplan-Meier curve was drawn with the log-rank test.
Univariate and multivariate Cox regression was used to determine the independence of gene markers. Spearman’s rank correlation test was utilized to identify the variables. Differences in the distributions of the variables were analyzed by the Chi-square test or Fisher’s exact test. $P < 0.05$ was considered statistically significant.

**Data availability statement**


**AUTHOR CONTRIBUTIONS**

YT and XZ designed this project and approved the manuscript. JW and ZL contributed to the data analysis and carried out the manuscript. JL, LZ, YQ, FZ, RH, HC and YT co-contributed to revising and polishing the manuscript and figures. All authors have read and approved the final submitted manuscript.

**CONFLICTS OF INTEREST**

The authors declare that they have no conflicts of interest.

**FUNDING**

This work was supported by grants from the Hunan provincial Health and Family Planning Commission (20201947, B20180186) and the Hunan Provincial Natural Science Foundation of China (2019JJ50547).

**REFERENCES**


33. Deng S, Zhou H, Xiong R, Lu Y, Yan D, Xing T, Dong L, Tang E, Yang H. Over-expression of genes and proteins of ubiquitin specific peptidases (USPs) and proteasome


49. Liu X, Wang J, Zhang G. miR-4458 regulates cell proliferation and apoptosis through targeting SOCS1 in


Supplementary Figure 1. The distribution of UPSGs on chromosomes. A total of 804 human UPSGs are distributed across all chromosomes, including the sex chromosome X and Y. The position of the line and dot indicates the position of the gene on the chromosome, and the height of the column indicates the abundance of gene expression.
Supplementary Figure 2. Construction of a prognostic risk model using LASSO regression analysis. (A) LASSO algorithms used to identify prognosis-related DEUPSGs. (B) The prognostic risk model constructed using LASSO coefficient values in the TCGA training set.
Supplementary Figure 3. Verification of the prognostic risk model in HNSCC patients in the TCGA database. (A) Kaplan-Meier survival curve for HNSCC patients in the TCGA test set. (B) Kaplan-Meier survival curve for HNSCC patients in the TCGA all set. (C) The risk plot distribution, survival status, and the expression of risk genes of HNSCC patients in the TCGA test set. (D) The risk plot distribution, survival status, and expression of risk genes in HNSCC patients in the TCGA all set.
Supplementary Figure 4. Performance validation of the risk model in HNSCC patients using the GEO database. (A) Kaplan-Meier survival curve with OS in the high- and low-risk HNSCC patients in the GSE65858 data set. (B) ROC curve showing the AUC for the risk score and other clinical factors of HNSCC patients in the GEO database. (C) The risk score distribution of the high- and low-risk HNSCC patients. (D) Scatter plot showing the survival statuses of HNSCC patients. (E) The expression of risk genes of HNSCC samples in the GSE65858 data set.
Supplementary Figure 5. Univariate and multivariate Cox regression analyses of OS in the GEO and TCGA data sets. (A, B) The clinical factors of the patients in the TCGA all set were assessed using univariate and multivariate Cox regression analysis, respectively. (C, D) The clinical factors of the patients in GEO (GSE65858) data set were assessed using univariate and multivariate Cox regression analysis, respectively.
Supplementary Figure 6. Other single GSEA figures of the high-risk and low-risk groups. Single GSEA plots showing enriched pathways in the high- and low-risk groups displayed in Table 3.
Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1–5.

Supplementary Table 1. List of ubiquitin proteasome system (UPS)-related genes analyzed in this study.

Supplementary Table 2. The 114 differentially expressed UPS-related genes (UPSGs) identified in HNSCC in the present study.

Supplementary Table 3. The top 20 enriched pathways of GSEA in the high-risk and low-risk groups.

Supplementary Table 4. The HNSCC patients of the TCGA training set.

Supplementary Table 5. The HNSCC patients of the TCGA test set.