

Identification and Clinical Validation of Genes Signatures with Grade and Survival in Head and Neck Carcinomas

Wei Ma

Nanjing Drum Tower Hospital: Nanjing University Medical School Affiliated Nanjing Drum Tower Hospital

Qing Cao

Yangzhou University

Wandong She (✉ shewandong@yeah.net)

Nanjing Drum Tower Hospital Clinical College of Nanjing Medical University

Research article

Keywords: head and neck carcinoma, malignancy, grade, prognosis, GEO

Posted Date: December 31st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-136855/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The mechanism of transition from low-grade to high-grade head and neck carcinomas (HNC) still remains unclear. This study aimed to explore the genes expression profiles that drive malignancy from low to high-grade HNC, as well as analyze their correlations with survival.

Methods: Gene expressions and clinical data of HNC were downloaded from the Gene Expression Omnibus (GEO) repository. The significantly different genes (SDGs) between low and high-grade HNC were screened by GEO2R tool and R software. Bioinformatics functions of SDGs were investigated by the enrichment analyses. Univariate and multivariate cox regressions were performed to identify prognostic SDGs of progression-free survival (PFS) and disease-specific survival (DSS). Receiver operating curve (ROC) was established to evaluate the ability to predict the prognosis. Then, the correlations between SDGs and clinical features were evaluated. The genes were experimentally validated by RT-PCR in clinical specimens.

Results: 35 SDGs were identified in 47 low-grade and 30 high-grade HNC samples. Enrichment analysis showed these SDGs were enriched in the DNA repair pathway and the regulation of I-kappaB kinase/NF-kappaB signaling pathway. Cox regression analyses showed that CXCL14, SLC44A1 and UBD were significantly associated with DSS, and PPP2R2C and SLC44A1 were associated with PFS. Patients at a high-risk or low-risk for prognosis were established based on genes signatures. High-risk patients had significantly shorter DSS and PFS than low-risk patients ($P=0.033$, 0.010 respectively). Multivariate cox regression showed human papillomavirus (HPV) ($P=0.033$), lymph node status ($P=0.032$) and residual status ($P=0.044$) were independent risk factors for PFS. ROC curves showed the risk score had better efficacy to predict survival both for DSS and PFS ($AUC=0.858$, 0.901 respectively). In addition, we found UBD, PPP2R2C and risk score were significantly associated with HPV status (all $P<0.05$). The experiment results showed CXCL14 and SLC44A1 were significantly overexpressed in the HNC grade I/II tissues and the UBD was overexpressed in the HNC grade III/IV tissues.

Conclusions: Our results suggested that SDGs had different expression profiles between the low-grade and high-grade HNC, and these genes may serve as prognostic biomarker to predict the survival.

Background

Head and neck carcinomas (HNC) are a group of heterogeneous tumors arising from the oral cavity, oropharynx, nasopharynx, hypopharynx, and larynx, ranking the sixth most prevalent cancer [1, 2]. Head and neck squamous cell carcinomas (HNSCC) account for 90% of all head and neck cancer. It's estimated that there are more than 600,000 new HNC cases and 350,000 deaths per year globally [3, 4]. HNC can be classified into different subgroups according to the human papillomavirus (HPV) status and histological grades [2, 4]. About 42% patients of HNC are diagnosed as advanced stage with extensive lymph nodes or distant metastasis at their initial visits [5]. Patients with HNC have benefited from the comprehensive treatment in recent years. However, the low-grade HNC have different treatment modalities from the

advanced ones, and the 5-year survival remains less than 50% despite the tremendous progress has been made in the multidisciplinary treatment, including surgery, radiotherapy and chemotherapy [6].

Selection of optimal management plans for HNC mainly is dependent on tailored risk evaluation [7-8]. Histological grade in HNC helps to assess the patients' risks to make therapeutic strategies and provide important clinical prognostic information. Despite the significances, relying solely on histological grade cannot provide a reference for clinical decision-making as a result of diagnostic inconsistency and classification discordance with different standard [9-10].

Additionally, the underlying mechanisms regulating the HNC progression from low to high-grade, such as NF - κ B pathways are still largely unknown. Therefore, it's imperative to identify new methods and biomarkers for increasing pathological grade values along with discovering new mechanisms about the transition from low-grade towards high-grade. The extensive applications of high-throughput sequencing technologies in cancer biology such as gene profile analysis, have revealed the relation between thousands of aberrant gene expressions associated with HNC patients [11-12]. Among them that have been functionally characterized, several has been linked to malignant progression [13-14]. Notably, many genes have key roles for diagnostic accuracy and predicting the prognosis [15-16].

In this study, we have comprehensively analyzed the significantly different genes (SDGs) and clinical information from the Gene Expression Omnibus (GEO) in order to explore whether different grade HNC have distinct gene expressions. To determine the clinical relevance, we also investigated the associations between genes and survival. Results were further verified in experiment using the clinical specimens' tissues.

Methods

Patient samples and data extraction

The gene expression data and clinical information of HNC were downloaded from the Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/gds/>). SDGs were obtained from the GSE117973 and were initially analyzed with GEO2R. R software (version 3.6.1) were used to identify SDGs by using the Wilcoxon test with the limma package. In this dataset, HNC was divided into low-grade (grade I/II) and high-grade (grade III/IV) groups. The SDGs with false discovery rate (FDR) < 0.05 and $|\log_2$ fold change (FC)| > 0.5 were considered to be differentially expressed.

Enrichment analysis

The functional analyses of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were conducted by using the SDGs with R package. GO analysis include the biological process (BP), cellular component (CC), and molecular function (MF). Top results with the false discovery rate (FDR) \leq 0.05 were considered noteworthy.

Survival analysis and ROC analysis

We evaluated the correlations between the disease-specific survival (DSS), progression-free survival (PFS) and SDGs by univariate and multivariate cox proportional hazards regression analysis. The prognostic factors ($P \leq 0.05$) were entered into multivariate cox regression to identify the independent prognostic risk factors.

The receiver operating curve (ROC) analysis was used to assess the sensitivity and specificity of the independent risk factors. The area under curve (AUC) of the ROC ranges from 0.5 to 1, with near 1 indicating perfect predictive ability and 0.5 indicating without predictive ability.

Experimental validation

To verify the prognostic genes expression levels in HNC tissues, we conducted the experimental validation in 45 HNC patients' specimens (25 grade I/II and 20 grade III/IV) who received surgery from January 2019 to August 2020 in the Clinical Medical College of Yangzhou University, Yangzhou, Jiangsu. This study was approved by the Internal Review Board of the Clinical Medical College of Yangzhou University, Yangzhou, Jiangsu.

Total RNA from 45 HNC tissues was purified using RNAiso plus (Takara, Dalian, China). Complementary DNA (cDNA) was synthesized from 1 μ g of total RNA using a PrimeScript[®] RT reagent Kit with gDNA (genomic DNA) Eraser (Takara). TB Green[®] Premix Ex Taq[®] II kit (Takara) was used to detect the indicated RNA levels on the QuantStudio Real-Time polymerase chain reaction (PCR) System (Applied Biosystems, USA). The RT reaction was performed for 1 cycle under the condition of 30°C for 10 min, 42°C for 30 min, 95°C for 5 min and 5°C for 5 min. Then the PCR reaction was performed using a Takara Shuzo PCR Amplification kit (cat. no. R011; Takara Bio, Inc.) with primer sets specific for different genes. The thermal conditions for the gene and GAPDH were denaturation for 30 sec at 95°C, annealing for 30 sec at 56°C, and extension for 30 sec at 72°C. The amplifications were performed using 25–28 cycles, respectively. The relative expression levels of the candidate genes were normalized to endogenous GAPDH (glyceraldehyde-3-phosphate dehydrogenase). The primers synthesized and by GENEWIZ company, Suzhou, China. The primers are listed in Supplementary Table S1.

Results

Distinct gene patterns in low-grade and high-grade HNC tissues

A total of 77 HNC samples with gene expressions and clinical data were obtained from GSE117973, including 47 low-grade and 30 high-grade samples. There were 35 significantly different genes (SDGs) between the two groups. Among these SDGs, 23 genes were down-regulated and 12 were up-regulated in the high-grade group compared with low-grade group (Table 1). The heatmap and volcano plots are

Enrichment analysis

Given the importance of the SDGs and further exploration about their functions, we performed the GO and KEGG analysis. GO results showed the SDGs were strongly associated with nucleotide-excision repair and DNA polymerase complex pathways. In the BP category, the SDGs were enriched in the nucleotide-excision repair pathway, as well as the regulation of I – kappaB kinase/NF – kappaB signaling pathway. In the CC category, the SDGs were enriched in DNA polymerase complex pathway. In the MF category, the SDGs were enriched in the structural constituent of cytoskeleton (Fig. 2A-B). In the KEGG analysis, the SDGs were involved in the base excision repair activity, which was similar in the GO analysis (Fig. 2C-D).

Prognostic SDGs in DSS and PFS

To explore whether the SDGs are associated with the DSS and PFS, we firstly performed the univariate cox regression was used to investigate SDGs with prognosis (Fig. 3A, 3B). Then, this was followed by the multivariate cox regression, and four genes (CXCL14, SLC44A1, UBD and PPP2R2C) were found to be linked with survival (shown in Table 2). We identified that CXCL14 and SLC44A1 were significantly associated with DSS, and the PPP2R2C, SLC44A1 were prognostic genes of PFS. Among these, CXCL14 and PPP2R2C were risk genes ($HR > 1$). The SLC44A1 and UBD genes were protective in survival ($HR < 1$). According to the prognostic genes expressions and their coefficient ^[17], we then calculated the risk score ($= \sum_{n=1}^j Coef_j * X_j$, with Coef j indicating the coefficient and X_j representing the relative expression levels of each gene standardized by z-score) of each patient and used the median risk score value as a cutoff point for classifying the 30 high-grade HNC patients into a high-risk group and a low-risk group (n = 15, respectively). The DSS and PFS time was much lower in the high-risk group (DSS, median time = 1.431 years vs. 2.625 years, $P=0.033$, Fig. 4A; PFS, median time = 1.361 years vs. 2.261 years, $P = 0.010$, Fig. 4B).

Prognostic hazard curves

We ranked the risk scores of patients for DSS and PFS and analyzed their distributions. As the heatmap of risk score shown, patients with high risk scores in the DSS group showed upregulation of CXCL14 and the downregulation of UBD (Fig. 5A). Moreover, patients with high risk scores in the PFS group showed the downregulation of PPP2R2C, implying it's a protective gene (Fig. 5B). The dot plots showed the survival status of DSS and PFS of HNC patients (Fig. 5C-F). When the risk score increased, the patients' risk increased, and the survival time decreased.

Independent risk factors of survival and ROC model

We combined the SDGs with clinical information in HNC patients. Then we performed univariate and multivariate cox regression analyses to investigate the independent risk factors for DSS and PFS. As shown in Fig. 6A, univariate cox regression showed the risk score was associated with DSS ($P = 0.003$).

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js independent risk factors for DSS (all $P < 0.05$)

(Fig. 6B). For the PFS, the risk score was the significant risk factor in the univariate cox regression ($P = 0.003$) (Fig. 6C). Multivariate cox regression showed HPV ($P = 0.033$), lymph node status ($P = 0.032$) and residual status ($P = 0.044$) were independent risk factors for survival (Fig. 6D).

In order to provide a model to predict the survival, we constructed the ROC curves using the possible risk factors associated with DSS and PFS respectively. In addition, we assess the feasibility using the area under curve (AUC) values. Risk score, HPV, R and tumor cell content were selected to establish the ROC, and the results showed the risk score had better ability to predict the DSS (AUC = 0.858) (Fig. 7A). In the PFS analysis, five prognostic parameters, including the risk score, HPV, T, N, R and tumor cell content were recruited. The risk score performance showed more excellent predictive ability than other factors (AUC = 0.901) (Fig. 7B).

Clinical correlation analysis

We further explored the relationships between the prognostic SDGs and clinical features, we calculated the correlations using the t-test or Kruskal-Wallis test. We found UBD, PPP2R2C and risk score were significantly associated with HPV status (all P values ≤ 0.05). UBD expressions were higher in the HPV (+) patients, and the PPP2R2C expression were higher in the HPV (-) patients (Fig. 8A-B). We also found risk score was significantly associated HPV status (Fig. 8C-D).

Experimental validation

According to the screening and validation steps as described above, we performed experimental validation using the four prognostic genes (CXCL14, PPP2R2C, SLC44A1, UBD), and GAPDH was set as an internal reference. The experiment results showed the CXCL14 and SLC44A1 were significantly overexpressed in the HNC grade I/II tissues and the UBD was overexpressed in the HNC grade III/IV tissues. There was no significant difference in the expression levels of PPP2R2C between the two groups. The results are shown in Fig. 9A-D.

In addition, we divided the 45 HNC patients into 21 HPV (+) and 24 HPV (-) groups according their HPV test results in clinic. Then, we further explored and verified the relations between the four prognostic genes (CXCL14, PPP2R2C, SLC44A1, UBD) and HPV status. The results were in agreement to our bioinformatics analysis and that the UBD was significantly higher in the HPV (+) group and PPP2R2C was significantly higher in the HPV (-) group. The results are shown in Fig. 9E-F.

Discussion

Cancers are primary caused by genetic alterations that result in the dysregulation of gene networks which are responsible for malignancy. Myriad studies have now used high-throughput sequencing technology to profile different types of cancer samples. Current molecular landscapes of head and neck cancer primarily focus on the biological differences between the HPV-negative and positive populations. It's

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js /ell accepted that genes with frequent and

significant genetic alterations are involved in various HNC cell functions, including the tumor development and progression [18-19]. However, few provides definitive evidence for elucidating the gene distinctions between the low-grade and high-grade HNC. In this study, we found that low-grade and high-grade HNC had different gene expression profiles, which directly linked to the DNA repair that may drive malignancy transformation from low-grade to high-grade. We also investigated the gene associations with clinical implications and discovered that SDGs were significantly related with the DSS and PFS. To increase the reliability of the research, we verified these findings using our clinical specimens' tissues, and the results were consistent with our findings.

We discovered that the SDGs were mainly enriched in the NF - kappaB signaling pathway and DNA repair by GO and KEGG analyses. Several studies have strongly supported the associations between NF - kappaB signaling pathway and HNC [20-22]. Qin Y et.al found that CCL18 (chemokine (C-C motif) ligand 18) could promote the HNSCC malignancy and its level was significantly associated with histological grade through regulating the NF - κB signal pathway [20]. Yu B et al provided evidence that NF - κB pathway can be activated by CD147, which was correlated positively with HNSCC grade. Furthermore, the NF - κB inhibitor could reduce the invasion of HNSCC cells [21]. In addition, the NF - κB pathway induced by XPR1 was related to many aspects of TSCC (tongue squamous cell carcinoma), including the tumor grade and patients' prognosis [22]. These studies illuminated the functions of NF - κB signal pathway in the HNSCC malignancy in terms of histological grade.

Continuous and chronic exposure to the tobacco, alcohol and infection with human papillomavirus (HPV) are the predominant risk factors for HNC, which induce DNA damage [23]. DNA repair mechanisms, such as excision repair (ER), mismatch repair (MMR), NHEJ (non-homologous end joining) and HR (homologous recombination) protect genome against damage and provide stability for gene and chromosome [23]. Any low DNA repair efficacy is recognized as one of the mechanisms for HNC initiations and progressions. Besides these, gene mutations and polymorphisms associated with DNA repair that HNC cells undergo are also determining factors promoting the HNC malignancy [23]. Much of the evidence of this comes from whole exome sequencing studies. For example, exonic and intronic variants of several genes work together during the process of DNA repair, especially in the DSBR (double strand break repair) and FA (Fanconi anemia) pathways [24]. Moreover, it's reported that certain gene involving the DNA repair pathways were correlated with HNC tumor size and clinical stage [25].

Instead of distinguishing genes directly through their associations with survival, we screened genes from different grades and then identified the prognostic genes. SLC44A1, also known as CTL1 (choline transporter-like 1), encodes an intermediate-affinity choline transporter protein. Choline is essential for all cells to synthesize the membrane phospholipids phosphatidylcholine (PC) and sphingomyelin and its uptake through SLC44A1 is strongly associated with cell viability, apoptosis and the malignant progression [26-27]. SLC44A1 may be involved in the tumorigenesis and the metastasis of colon cancer, and currently used as a prognostic biomarker [26]. However, the field is still in its early stages and only a

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js es of SLC44A1 in HNC. Our experimental results

demonstrated SLC44A1 was upregulated in the low grade (I/II) HNC, implying it may play a protective role in HNC. Nishiyama R et al. found the functional inhibition of CTL1(SLC44A1) by cationic drugs could significantly increase caspase-3/7 activity and promote tongue cancer cell death [28]. Identification of the CTL1-mediated choline transport system could provide a potential new target for tongue cancer therapy.

Also contained in the prognosis is the PPP2R2C. This gene has been confirmed to be linked to gliomas, lung cancer and prostate cancer, and it is thought to be a potential tumor-suppressor gene [29–31]. Nonetheless, the role of PPP2R2C in our results has been questioned, which showed the lower expression in the high-grade group, implying a tumor-suppressor role. However, survival analysis showed the contradictory risk role (HR > 1) in the prognosis. The discrepancy may be explained by the small number samples. The determination of the role of PPP2R2C in HNC necessitates experimental analysis that will delineate the contribution of PPP2R2C in the function of HNC cells. Our experimental result confirmed the slightly higher level of PPP2R2C in high-grade HNC, implying it may serve as a tumor-promoting gene, but without significant statistical difference (P > 0.05). The previous researches demonstrated that PPP2R2C is subjected to transcriptional regulation by factors such as miRNAs that involved in the HNC cell proliferation, invasion and recurrence [32–33]. However, the mechanisms of miRNAs on cancer cell activities through regulating the expression of PPP2R2C are needed to be further investigated.

In order to verify whether the prognostic risk factors could predict the survival, we further established ROC model using the factors selected from multivariate cox regression. The results demonstrated the risk score manifested excellent predictive ability, implying it could serve as an accurate survival indicator both for DSS and PFS (AUC = 0.0858, 0.901 respectively). To gain a better estimate how the genes influence the clinical characteristics, we assessed the relationships between the SDGs and clinical features in HNC patients. Our study found the UBD, PPP2R2C and risk score were strongly associated with HPV status (all P < 0.05). As previous study reported by Wang J et.al, UBD expression level was much higher in the HPV (+) oropharyngeal squamous cell carcinoma (OSCC) compared with the HPV (-) OSCC [34]. Their results are congruent with our findings. We also found the PPP2R2C level was higher in the HPV (-) group relative to HPV (+) group in the experiment results. However, no more literatures have studied the correlation between PPP2R2C and HPV until now. Further researches are needed to explore the associations. In addition, these may be critical to understand the cause of differences in clinical presentation and molecular landscapes and develop tailored therapy between the HPV (+) and HPV (-) HNC.

The strength of our study is that we performed a systematic analysis to identify SDGs in different HNC grades using a public database, with experimental validation. This work may help shed light on the HNC malignant progression and develop new targeted drugs. Several limitations should not be ignored in this study. Firstly, the sample number is too small to reach a robust conclusion that applies to all tumors in the head and neck. Secondly, the exact mechanisms how these SDGs drive malignancy transformation from low-grade to high-grade are still unknown. Lastly, we failed to examine the significance of SDGs for all clinical implications, such as therapy modality. Notwithstanding its limitations, this study does provide a preliminary overview of SDGs profile in HNC and these limitations can be solved if there are more

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js tions in the future.

Conclusions

In conclusion, we identified different SDGs expression profiles between the high and low-grade HNC grades by analyzing the public database and the experiment. This study explains the prognostic genes and survival of HNC from the perspectives of bioinformatics. However, there still needs further validations to confirm the findings of our study.

Abbreviations

HNC; head and neck carcinomas; HNSCC:head and neck squamous cell carcinomas; SDGs:significantly different genes; GEO:Gene Expression Omnibus; FDR:false discovery rate; FC:fold change; PFS:progression-free survival; DSS:disease-specific survival; GO:Gene Ontology; KEGG:Kyoto Encyclopedia of Genes and Genomes; ROC:receiver operating curve; HPV:human papillomavirus; AUC:area under curve; cDNA:complementary DNA; gDNA:genomic DNA; PCR:Real-Time polymerase chain reaction; GAPDH:glyceraldehyde-3-phosphate dehydrogenase. HR:hazard ratio. CI:confidence interval. TSCC:tongue squamous cell carcinoma

Declarations

Acknowledgments

We would like to thank the authors of this studies that were included in our study.

Ethics approval and consent to participate

This study was approved by the Internal Review Board of Clinical Medical College, Yangzhou University, Yangzhou, Jiangsu, China.

Consent for publication

Not applicable.

Availability of data and materials

All data were available in GEO database (<https://www.ncbi.nlm.nih.gov/gds/>). And all data analyzed and displayed in the present manuscript are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Not applicable.

Authors' contributions

WM contributed to conception, design, data acquisition and analysis. QC performed RNA extraction and interpreted of PCR results. WDS conceived and designed the experimental protocol, and reviewed and approved the final version of the manuscript. All authors have read and approved the manuscript.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics. 2018. *CA Cancer J Clin.* 2018 Jan;68(1):7–30. doi: 10.3322/caac.21442.
2. Kawakita D, Matsuo K. Alcohol and head and neck cancer. *Cancer Metastasis Rev.* 2017 Sep;36(3):425–34. doi:10.1007/s10555-017-9690-0.
3. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer.* 2015 Mar 1;136(5):E359-86. doi: 10.1002/ijc.29210.
4. Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. *Nat Rev Cancer.* 2011 Jan;11(1):9–22. doi:10.1038/nrc2982.
5. Liu C, Yu Z, Huang S, Zhao Q, Sun Z, Fletcher C, Jiang Y, Zhang D. Combined identification of three miRNAs in serum as effective diagnostic biomarkers for HNSCC. *EBioMedicine.* 2019 Dec;50:135–143. doi: 10.1016/j.ebiom.2019.11.016.
6. Bonner JA, Harari PM, Giralt J, Cohen RB, Jones CU, Sur RK, Raben D, Baselga J, Spencer SA, Zhu J, et al. Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. *Lancet Oncol.* 2010 Jan;11(1):21–8. doi:10.1016/S1470-2045(09)70311-0.
7. Caudell JJ, Torres-Roca JF, Gillies RJ, Enderling H, Kim S, Rishi A, Moros EG, Harrison LB. The future of personalised radiotherapy for head and neck cancer. *Lancet Oncol.* 2017 May;18(5):e266–73. doi:10.1016/S1470-2045(17)30252-8.
8. Ferris RL. Immunology and Immunotherapy of Head and Neck Cancer. *J Clin Oncol.* 2015 Oct 10;33(29):3293 – 304. doi: 10.1200/JCO.2015.61.1509.
9. Huang SH, O'Sullivan B. Overview of the 8th Edition TNM Classification for Head and Neck Cancer. *Curr Treat Options Oncol.* 2017 Jul;18(7):40. doi:10.1007/s11864-017-0484-y.
10. Haughey BH, Sinha P, Kallogjeri D, Goldberg RL, Lewis JS Jr, Piccirillo JF, Jackson RS, Moore EJ, Brandwein-Gensler M, Magnuson SJ, et al. Pathology-based staging for HPV-positive squamous carcinoma of the oropharynx. *Oral Oncol.* 2016 Nov;62:11–19. doi: 10.1016/j.oraloncology.2016.09.004.
11. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sawyer C, McKenna A, et al. The mutational landscape of head and neck squamous cell carcinoma.

- Science. 2011 Aug 26;333(6046):1157-60. doi: 10.1126/science.1208130.
12. Qi Z, Barrett T, Parikh AS, Tirosh I, Puram SV. Single-cell sequencing and its applications in head and neck cancer. *Oral Oncol.* 2019 Dec;99:104441. doi:10.1016/j.oraloncology.2019.104441.
 13. Bunbanjerdasuk S, Vorasan N, Saethang T, Pongrujikorn T, Pangpunyakulchai D, Mongkonsiri N, Arsa L, Thokanit N, Pongsapich W, Anekpuritanang T, et al. Oncoproteomic and gene expression analyses identify prognostic biomarkers for second primary malignancy in patients with head and neck squamous cell carcinoma. *Mod Pathol.* 2019 Jul;32(7):943–956. doi: 10.1038/s41379-019-0211-2.
 14. Jin Z, Zhao X, Cui L, Xu X, Zhao Y, Younai F, Messadi D, Hu S. UBE2C promotes the progression of head and neck squamous cell carcinoma. *Biochem Biophys Res Commun.* 2020 Mar 5;523(2):389–397. doi: 10.1016/j.bbrc.2019.12.064.
 15. Shen Y, Liu J, Zhang L, Dong S, Zhang J, Liu Y, Zhou H, Dong W. Identification of Potential Biomarkers and Survival Analysis for Head and Neck Squamous Cell Carcinoma Using Bioinformatics Strategy: A Study Based on TCGA and GEO Datasets. *Biomed Res Int.* 2019 Aug 7;2019:7376034. doi: 10.1155/2019/7376034.
 16. Wintergerst L, Selmansberger M, Maihoefer C, Schüttrumpf L, Walch A, Wilke C, Pitea A, Woischke C, Baumeister P, Kirchner T, et al. A prognostic mRNA expression signature of four 16q24.3 genes in radio(chemo)therapy-treated head and neck squamous cell carcinoma (HNSCC). *Mol Oncol.* 2018 Dec;12(12):2085–101. doi:10.1002/1878-0261.12388.
 17. Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D, Levy R. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N Engl J Med.* 2004 Apr;29(18):1828–37. doi:10.1056/NEJMoa032520. 350) .
 18. Foy JP, Bazire L, Ortiz-Cuaran S, Deneuve S, Kielbassa J, Thomas E, Viari A, Puisieux A, Goudot P, Bertolus C, et al. A 13-gene expression-based radioresistance score highlights the heterogeneity in the response to radiation therapy across HPV-negative HNSCC molecular subtypes. *BMC Med.* 2017 Sep 1;15(1):165. doi: 10.1186/s12916-017-0929-y.
 19. Kostareli E, Holzinger D, Bogatyrova O, Hielscher T, Wichmann G, Keck M, Lahrmann B, Grabe N, Flechtenmacher C, Schmidt CR, et al. HPV-related methylation signature predicts survival in oropharyngeal squamous cell carcinomas. *J Clin Invest.* 2013 Jun;123(6):2488–501. doi:10.1172/JCI67010.
 20. Qin Y, Wang J, Zhu G, Li G, Tan H, Chen C, Pi L, She L, Chen X, Wei M, et al. CCL18 promotes the metastasis of squamous cell carcinoma of the head and neck through MTDH-NF- κ B signalling pathway. *J Cell Mol Med.* 2019 Apr;23(4):2689–701. doi:10.1111/jcmm.14168.
 21. Yu B, Zhang Y, Wu K, Wang L, Jiang Y, Chen W, Yan M. CD147 promotes progression of head and neck squamous cell carcinoma via NF-kappa B signaling. *J Cell Mol Med.* 2019 Feb;23(2):954–66. doi:10.1111/jcmm.13996.
 22. Chen WC, Li QL, Pan Q, Zhang HY, Fu XY, Yao F, Wang JN, Yang AK. Xenotropic and polytropic retrovirus receptor 1 (XPR1) promotes progression of tongue squamous cell carcinoma (TSCC) via

- activation of NF- κ B signaling. *J Exp Clin Cancer Res.* 2019 Apr 17;38(1):167. doi: 10.1186/s13046-019-1155-6.
23. Dylawerska A, Barczak W, Wegner A, Golusinski W, Suchorska WM. Association of DNA repair genes polymorphisms and mutations with increased risk of head and neck cancer: a review. *Med Oncol.* 2017 Nov 15;34(12):197. doi: 10.1007/s12032-017-1057-4.
 24. Das R, Kundu S, Laskar S, Choudhury Y, Ghosh SK. Assessment of DNA repair susceptibility genes identified by whole exome sequencing in head and neck cancer. *DNA Repair (Amst).* 2018 Jun-Jul;66–67:50–63. doi: 10.1016/j.dnarep.2018.04.005.
 25. Mutlu P, Mutlu M, Yalcin S, Unsoy G, Yaylaci A, Saylam G, Akin I, Korkmaz H, Gunduz U. Detection of XRCC1 gene polymorphisms in Turkish head and neck squamous cell carcinoma patients: a comparative analysis with different populations. *J BUON.* 2015 Mar-Apr;20(2):540–7.
 26. Gao P, He M, Zhang C, Geng C. Integrated analysis of gene expression signatures associated with colon cancer from three datasets. *Gene.* 2018 May 15;654:95–102. doi: 10.1016/j.gene.2018.02.007.
 27. Inazu M. Choline transporter-like proteins CTLs/SLC44 family as a novel molecular target for cancer therapy. *Biopharm Drug Dispos.* 2014 Nov;35(8):431–49. doi:10.1002/bdd.1892.
 28. Nishiyama R, Nagashima F, Iwao B, Kawai Y, Inoue K, Midori A, Yamanaka T, Uchino H, Inazu M. Identification and functional analysis of choline transporter in tongue cancer: A novel molecular target for tongue cancer therapy. *J Pharmacol Sci.* 2016 Jun;131(2):101–9. doi:10.1016/j.jphs.2016.04.022.
 29. Fan YL, Chen L, Wang J, Yao Q, Wan JQ. Over expression of PPP2R2C inhibits human glioma cells growth through the suppression of mTOR pathway. *FEBS Lett.* 2013 Dec 11;587(24):3892-7. doi: 10.1016/j.febslet.2013.09.029.
 30. Banerjee AK, Read CA, Griffiths MH, George PJ, Rabbitts PH. Clonal divergence in lung cancer development is associated with allelic loss on chromosome 4. *Genes Chromosomes Cancer.* 2007 Sep;46(9):852–60. doi:10.1002/gcc.20472.
 31. Bluemn EG, Spencer ES, Mecham B, Gordon RR, Coleman I, Lewinshtein D, Mostaghel E, Zhang X, Annis J, Grandori C, et al. PPP2R2C loss promotes castration-resistance and is associated with increased prostate cancer-specific mortality. *Mol Cancer Res.* 2013 Jun;11(6):568–78. doi:10.1158/1541-7786.MCR-12-0710.
 32. Yan L, Cai K, Liang J, Liu H, Liu Y, Gui J. Interaction between miR-572 and PPP2R2C, and their effects on the proliferation, migration, and invasion of nasopharyngeal carcinoma (NPC) cells. *Biochem Cell Biol.* 2017 Oct;95(5):578–84. doi:10.1139/bcb-2016-0237.
 33. Wu AH, Huang YL, Zhang LZ, Tian G, Liao QZ, Chen SL. MiR-572 prompted cell proliferation of human ovarian cancer cells by suppressing PPP2R2C expression. *Biomed Pharmacother.* 2016 Feb;77:92–7. doi: 10.1016/j.biopha.2015.12.005.
 34. Wang J, Xi X, Shang W, Acharya A, Li S, Savkovic V, Li H, Haak R, Schmidt J, Liu X, et al. The molecular differences between human papillomavirus-positive and -negative oropharyngeal

Figures

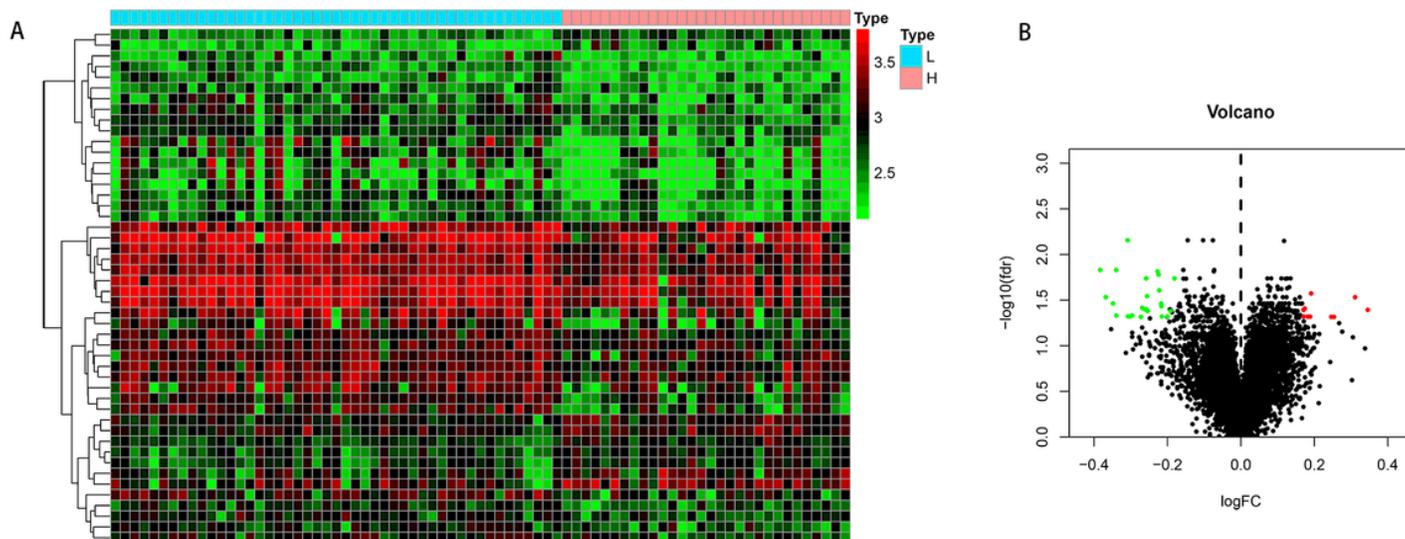


Figure 1

Differential expression of SDGs in HNC samples. (A) heatmap of SDGs expression profiles in low-grade and high-grade HNC samples. The red color indicates the up-regulated genes and green indicates down-regulated genes. L: low-grade; H: high-grade. (B) Volcano plot of SDGs in low-grade and high-grade HNC samples. The red dots represent upregulated genes, and the green dots represent downregulated genes.

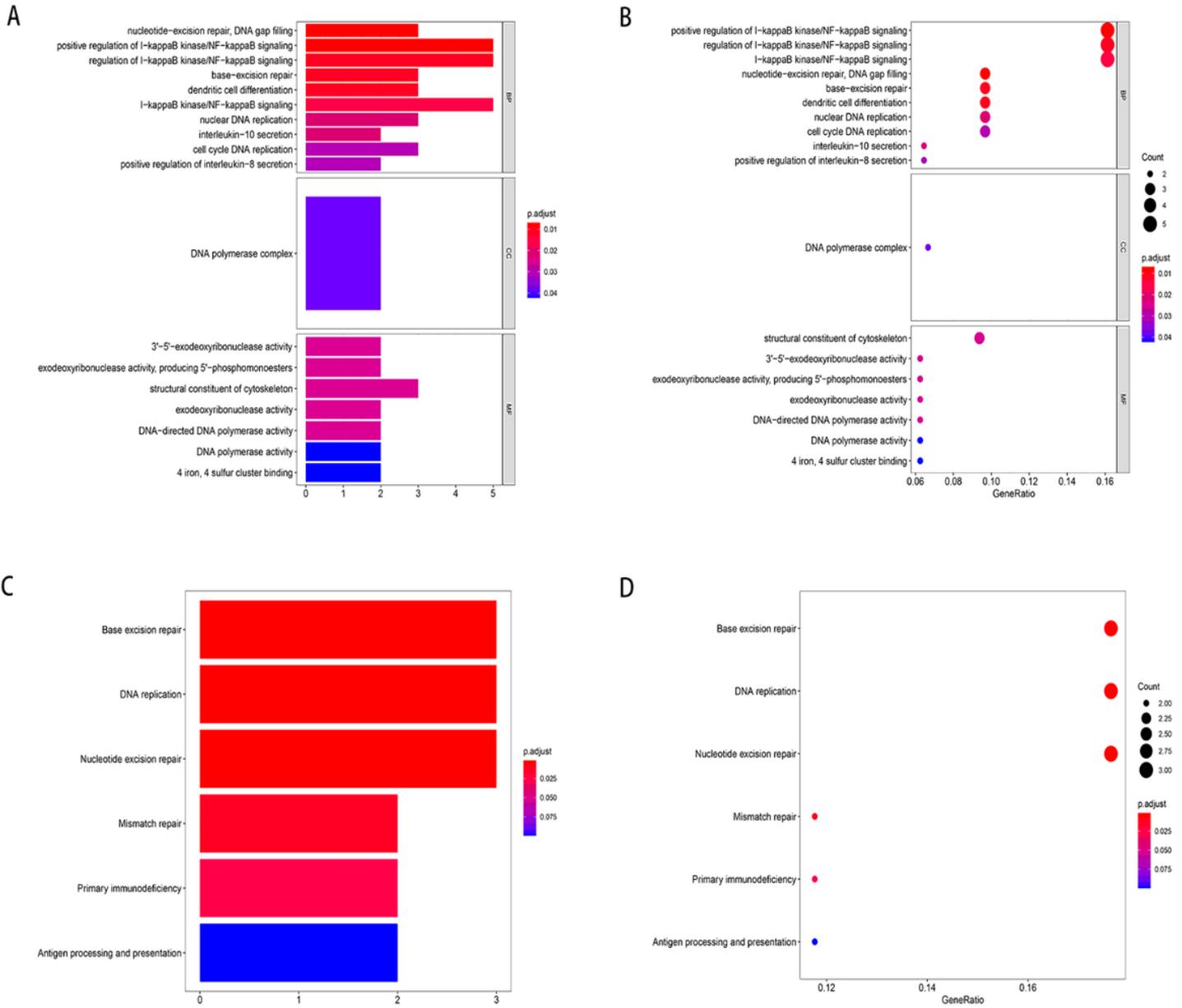


Figure 2

Enrichment analysis of SDGs. (A) bar plot of GO analysis, including the BP, CC and MF analysis. (B) bubble diagram of GO analysis. The larger bubble and darker color indicated the more significant enrichment process. (C) bar plot of KEGG analysis. (D) bubble diagram of KEGG analysis.

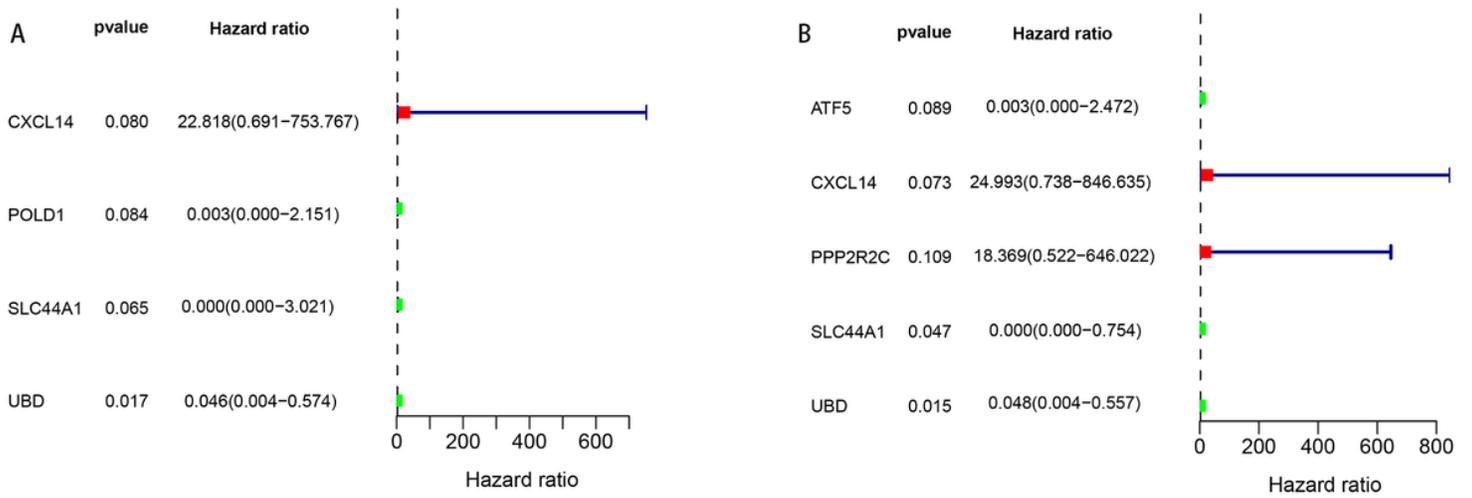


Figure 3

Characteristics of risk SDGs in the prognostic forest plot models. P values and hazard ratios of the risk SDGs for the (A) DSS and (B) PFS models are shown. Survival analysis shows UBD is significantly associated with DSS in the univariate cox regression. SLC44A1 and UBD are significantly associated with PFS.

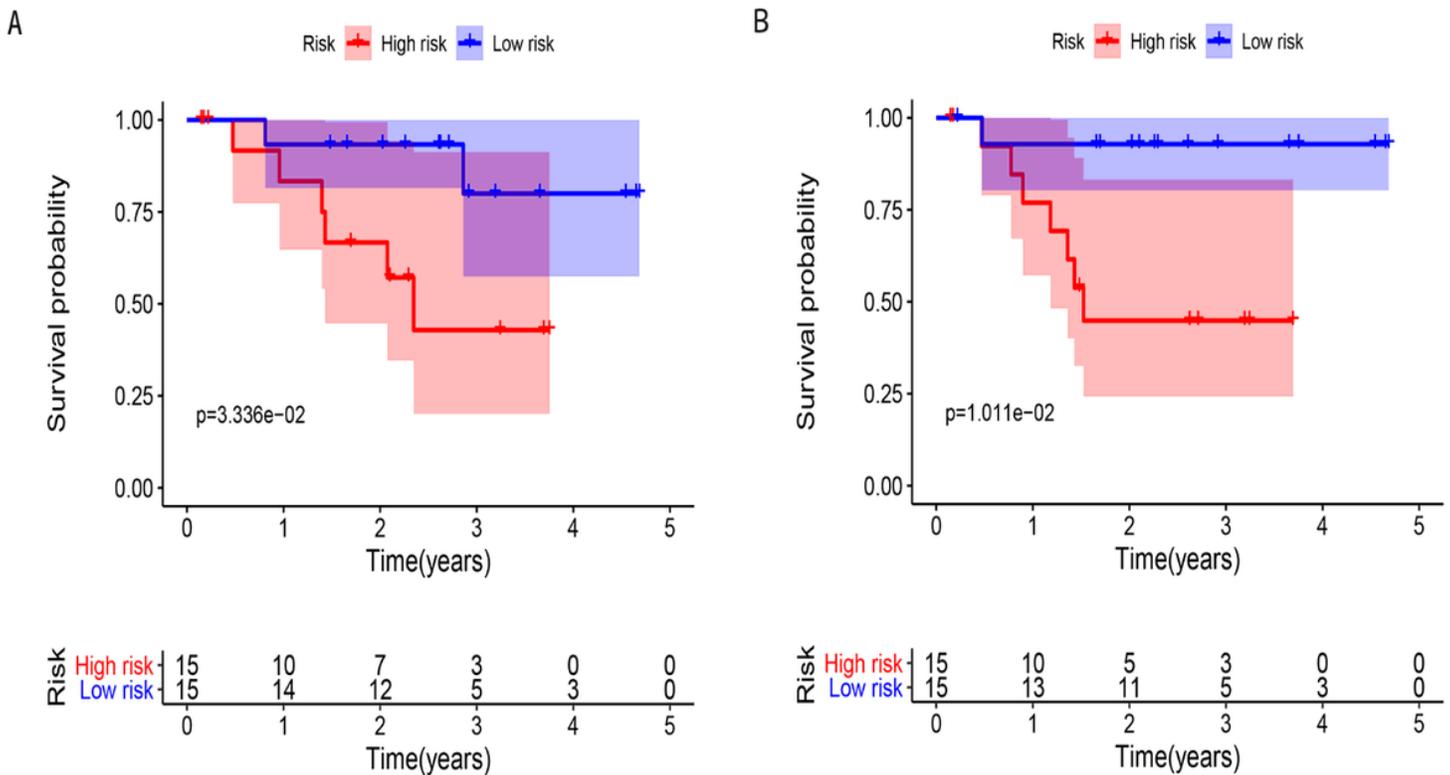


Figure 4

K-M curves of HNC patients. Kaplan-Meier curve for (A) DSS and (B) PFS in the high-risk and low-risk groups when stratified by the risk score. Low-risk group patients have higher survival probabilities than high-risk group patients (DSS: $p=3.336 \times 10^{-2}$, PFS: $p=1.011 \times 10^{-2}$ respectively).

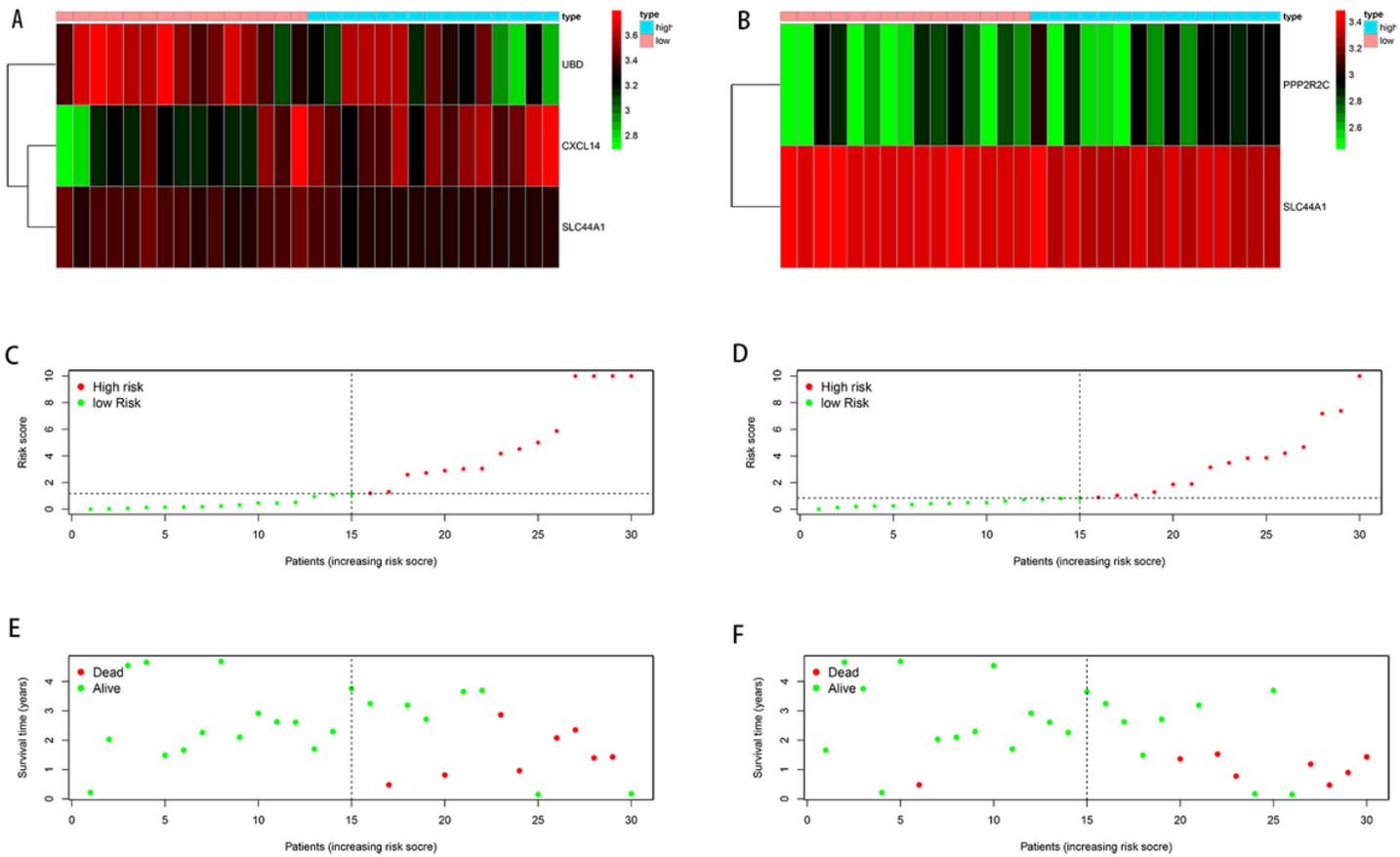


Figure 5

Risk score analysis based on the gene signature in the HNC group. DSS: (A), (C), (E); PFS: (B), (D), (F). Upper panel: heatmap of UBD, CXCL14, SLC44A1 and PPP2R2C expression in HNC samples. The colors from green to red indicate the expression level from low to high. Middle panel: patient survival status and time distributed by risk score. The dotted line indicates the individual inflection point of the risk score curve, by which the patients were categorized into low-risk and high-risk groups. Bottom panel: risk score curve of the autophagy signature. The green dots represent patients alive and the red dots represent the dead.

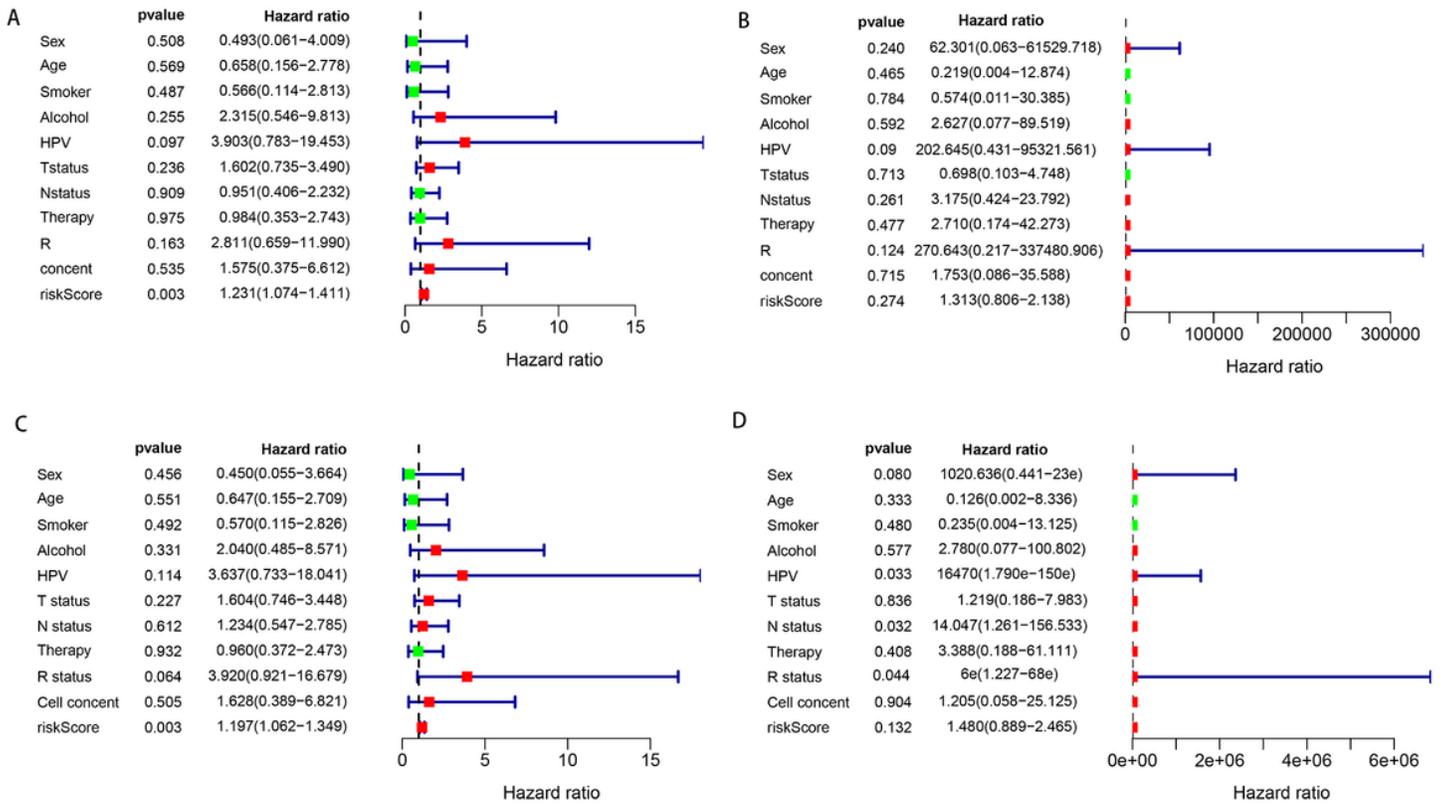


Figure 6

Forest plots of prognostic risk factors. Univariate cox regression forest plots of (A) DSS and (C) PFS. Multivariate cox regression forest plots of (B) DSS and (D) PFS.

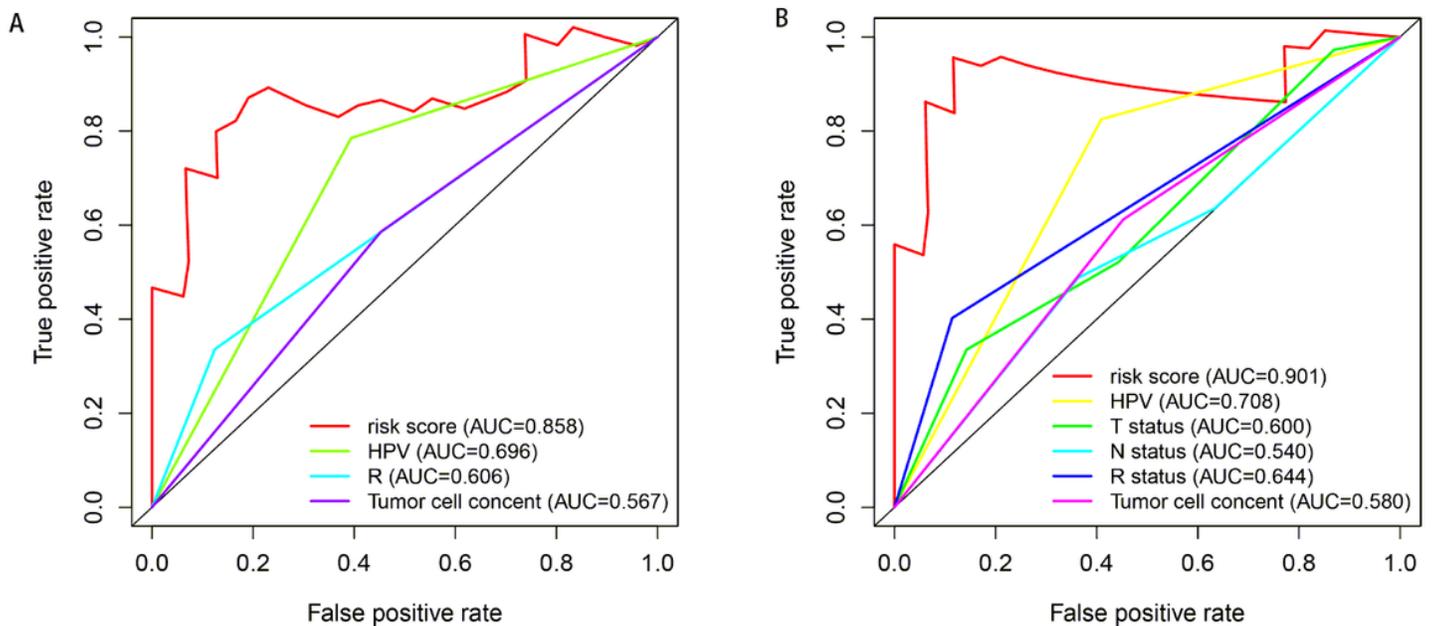


Figure 7

Prognostic performance of the risk factors models. ROC curves demonstrated the predictive abilities for (A) DSS and (B) PFS. The AUC ranges from 0.5 to 1.0. The AUC closes to 1.0, implying the more accurate it predicts. AUC: area under curve.

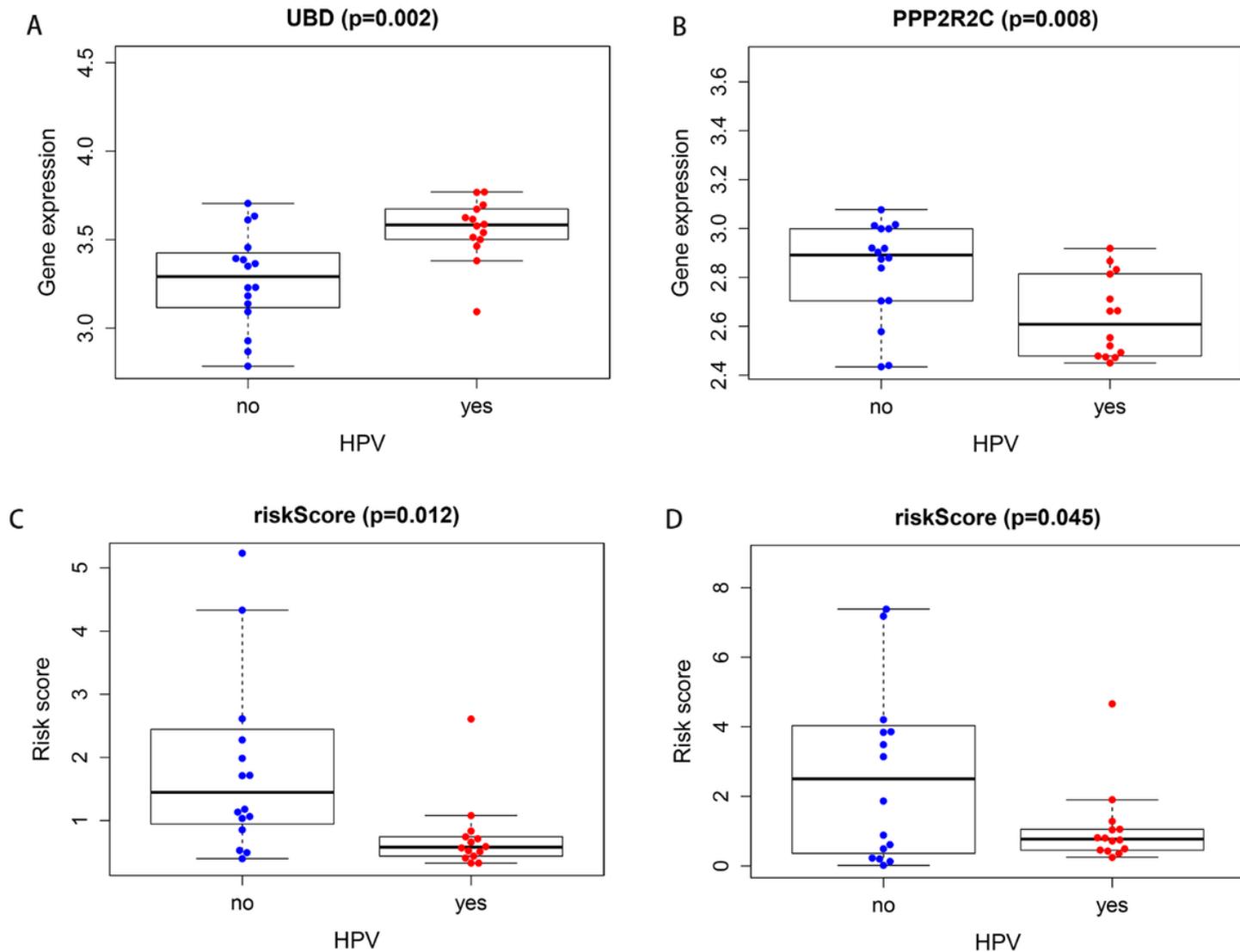


Figure 8

correlations between SDGs and clinical features. (A) UBD expression level and HPV status. (B) PPP2R2C expression level and HPV status. Risk score level and the HPV status in the DSS (C) and PFS (D) groups.

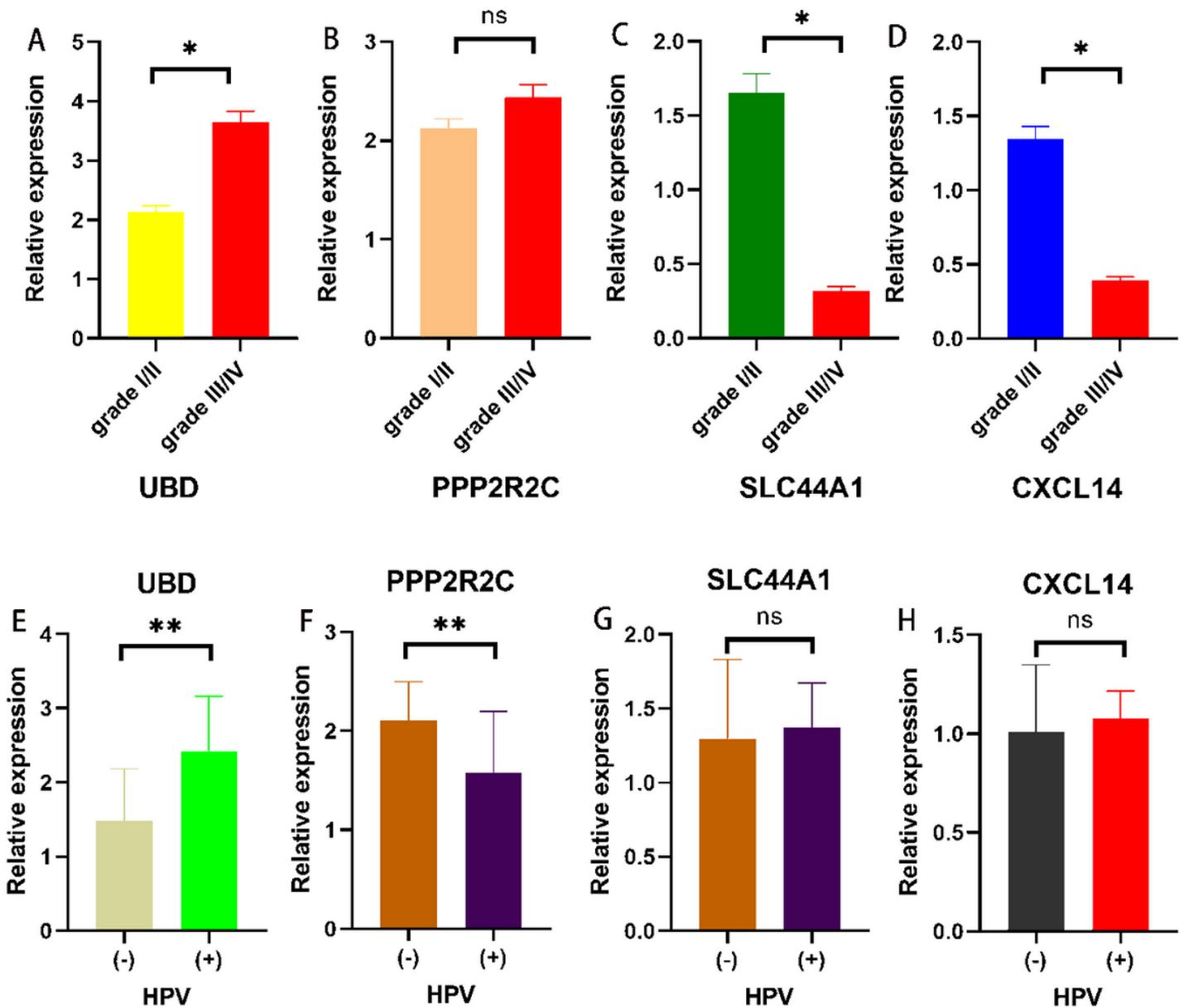


Figure 9

Four prognostic genes expression profiles and their correlations with HPV status in the clinical specimens' tissues. (A)UBD, (B) PPP2R2C, (C) SLC44A1 and (D)CXCL14 expression levels in the high grade (III/IV) and low grade (I/II) HNC. (E)UBD, (F) PPP2R2C, (G) SLC44A1 and (H) CXCL14 expression levels in the HPV (+) and HPV (-) HNC patients. *: significantly different. ns: not significant.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTableS1.docx](#)