*Research Article*

# Prognostic Correlation of an Autophagy-Related Gene Signature in Patients with Head and Neck Squamous Cell Carcinoma

**Cai Yang** [ID],[1] **Hongxiang Mei** [ID],[1] **Liang Peng** [ID],[2] **Fulin Jiang** [ID],[1] **Bingjie Xie** [ID],[1] **and Juan Li** [ID][1]

[1]*Department of Orthodontics, West China Hospital of Stomatology, West China School of Stomatology, Sichuan University, State Key Laboratory of Oral Diseases, Chengdu 610041, China*
[2]*Intelligent Information Technologies and Applications Lab (IIT), School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610041, China*

Correspondence should be addressed to Juan Li; lijuan@scu.edu.cn

Considerable evidence indicates that autophagy plays a vital role in the biological processes of various cancers. The aim of this study is to evaluate the prognostic value of autophagy-related genes in patients with head and neck squamous cell carcinoma (HNSCC). Transcriptome expression profiles and clinical data acquired from The Cancer Genome Atlas (TCGA) database were analyzed by Cox proportional hazards model and Kaplan–Meier survival analysis to screen autophagy-related prognostic genes that were significantly correlated with HNSCC patients' overall survival. Functional enrichment analyses were performed to explore biological functions of differentially expressed autophagy-related genes (ARGs) identified in HNSCC patients. Six ARGs (EGFR, HSPB8, PRKN, CDKN2A, FADD, and ITGA3) identified with significantly prognostic values for HNSCC were used to construct a risk signature that could stratify patients into the high-risk and low-risk groups. This signature demonstrated great value in predicting prognosis for HNSCC patients and was indicated as an independent prognostic factor in terms of clinicopathological characteristics (sex, age, clinical stage, histological grade, anatomic subdivision, alcohol history, smoking status, HPV status, and mutational status of the samples). The prognostic signature was also validated by data from the Gene Expression Omnibus (GEO) database and International Cancer Genome Consortium (ICGC). In conclusion, this study provides a novel autophagy-related gene signature for predicting prognosis of HNSCC patients and gives molecular insights of autophagy in HNSCC.

## 1. Introduction

Head and neck squamous cell carcinomas include several types of malignancies from the oral cavity, pharynx, and larynx. Each year, over 600,000 cases are diagnosed worldwide, making HNSCC the sixth most prevalent cancer [1, 2]. Studies have shown the risk factors related to HNSCC are tobacco use, alcohol consumption, and human papillomavirus (HPV) infection. The typical treatments include surgery, radiotherapy, and chemotherapy. However, the mortality rate of HNSCC remains high, which results in 380,000 death annually mainly because of local recurrence and distant metastasis

[3, 4]. The current plight is partly on account of the lack of accurate and reliable biomarkers for prognosis prediction in the early stage of this disease [5]. Therefore, it is quite essential to explore an effective and specific signature to help severity assessment and guide decision-making in clinical practice.

Autophagy is an important process which selects and degrades dysfunctional organelles, microbes, and proteins by lysosomes to maintain cellular homeostasis and sustain metabolism. Aberrant expression of autophagy-related genes (ARGs) is closely related to multiple diseases, especially neurodegenerative disorder, inflammatory abnormality, and
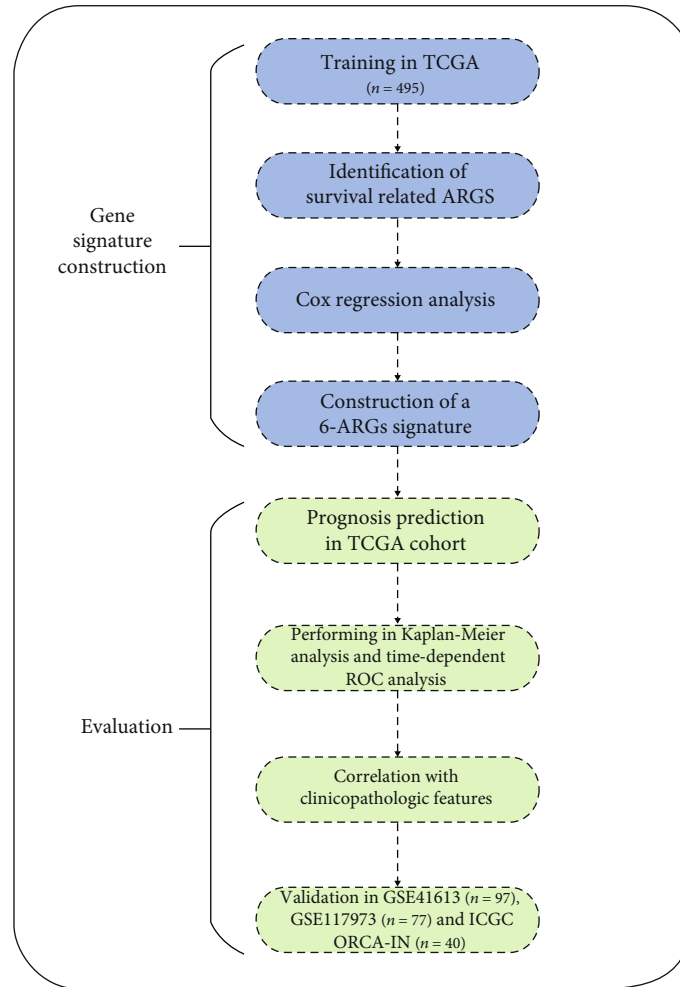
FIGURE 1: A schematic flowchart of this study.

cancer [6]. Autophagy plays a multifaceted role in tumor initiation and progression in respect of microenvironment stress, nutritional supplement, and immune status [7, 8]. Research suggests that autophagy-related mechanism is an optimistic target for future cancer therapy [9]; thus, exploring the valuable knowledge of autophagy is in urgent need and can offer huge potential in long term.

In this study, we comprehensively analyzed the correlation between expression profiles of ARGs and clinical features of HNSCC patients from The Cancer Genome Atlas (TCGA) database and constructed a 6-ARG risk signature. This model was identified to be an independent prognostic signature for HNSCC patients. Functional analysis was applied to reveal more information of ARGs. These findings could provide novel biomarkers for predicting the survival of HNSCC patients and give new insights in personalized therapy.

## 2. Materials and Methods

*2.1. Data Acquisition and Preparation.* RNA sequencing data of 545 HNSCC specimens, simple nucleotide variation data of 506 HNSCC specimens, and corresponding clinical infor-

mation of 528 patients were downloaded from TCGA database (https://cancergenome.nih.gov/). A total of 232 ARGs were obtained from the Human Autophagy Database (HADb; http://autophagy.lu/clustering/index.html). We extracted autophagy-related gene expression data using Practical Extraction and Report Language (Perl; https://www.perl.org/). The clinical data was combined with corresponding gene expression profiles by unique ID of the patients. Samples without complete survival information or with overall survival time less than 7 days were excluded. Consequently, the data of 495 patients were retained for gene signature construction and further analyses. Three independent microarray HNSCC cohorts obtained from the GEO database (GSE41613, $n = 97$; GSE117973, $n = 77$) and ICGC (ORCA-IN, $n = 40$) were used as the testing groups. Gene expression data from the GSE117973 dataset was normalized using the zero-mean normalization method implemented in the Sklearn library with Python programming language [10].

*2.2. Identification of Differentially Expressed ARGs and Functional Annotation.* R programming language (version 3.6.2) was utilized as data analysis and plotting tool
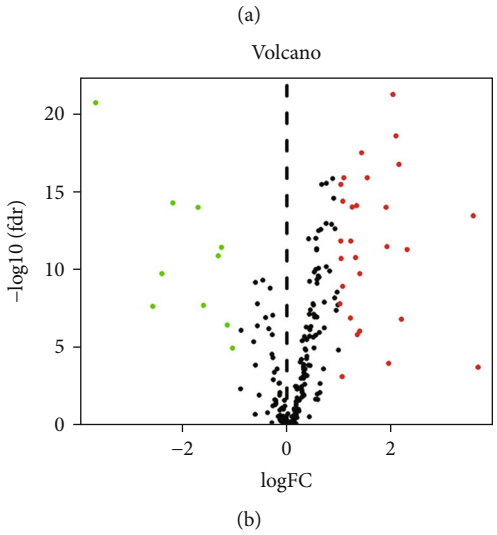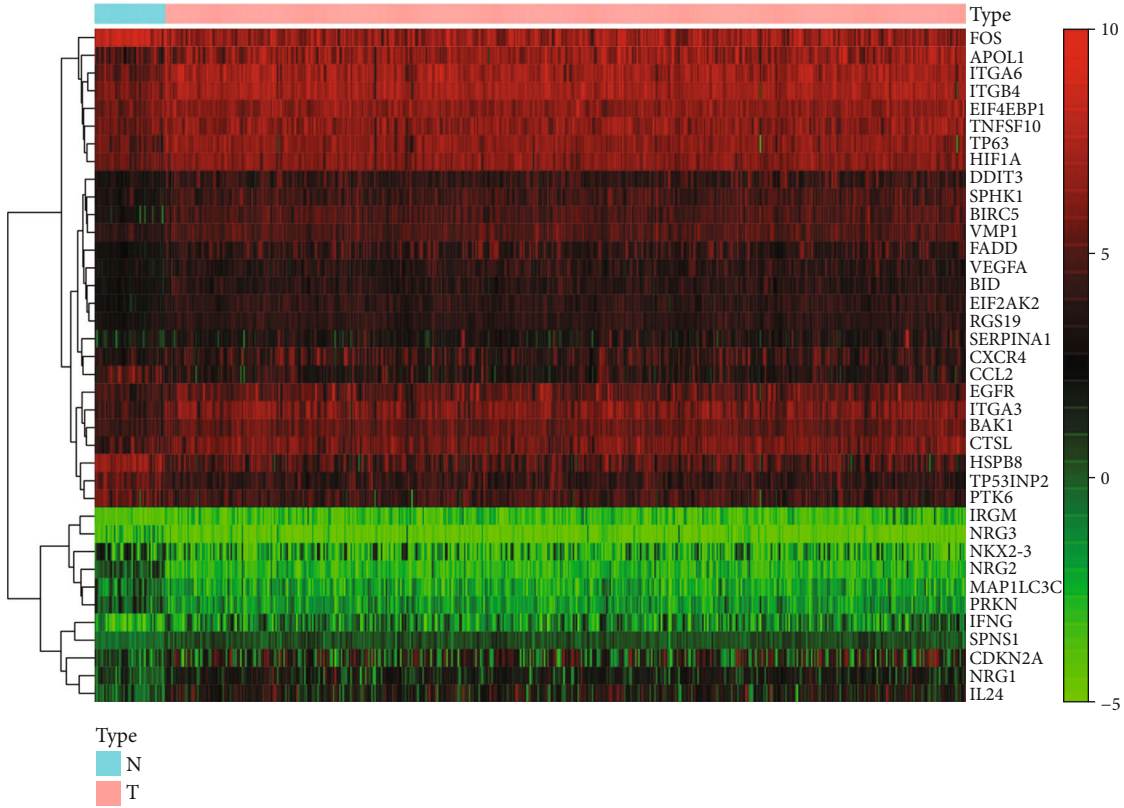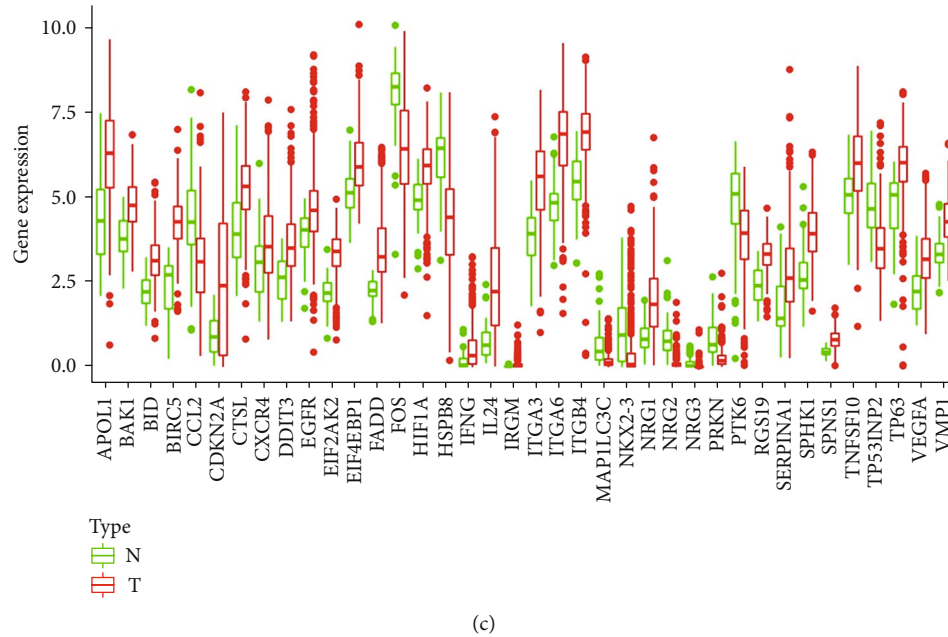
(a)



(b)

Figure 2: Continued.

(c)

Figure 2: Differentially expressed autophagy-related genes (ARGs). (a) The heat maps of the 38 differently expressed ARGs. The red color indicates high gene expression while the green color indicates low gene expression. N indicates nontumor tissues; T indicates tumor tissues. (b) The volcano plot for the 232 ARGs from TCGA database. Red indicates high expression, and green indicates low expression. Black indicates that those genes show no difference between HNSCC and paired nontumor tissues. (c) The boxplot of the differentially expressed ARGs.

throughout this research. The R package limma was used to select differentially expressed ARGs between HNSCC and nontumor samples. After normalization by the formula log $2(x + 1)$ transformed, 38 differentially expressed ARGs were identified with log fold change (FC) > 1 or <-1 and adjusted $p$ value < 0.05. To investigate biological functions of the selected ARGs, gene ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were performed by R software and $p < 0.05$ was considered to be of statistically significant difference. The GOplot package was applied to visualize the enrichment terms.

2.3. Construction of the Prognostic Signature. Univariate Cox regression analysis was performed in the training set (TCGA cohort) to evaluate the association between differentially expressed ARGs and overall survival (OS) of HNSCC patients. ARGs with $p$ value < 0.05 were considered to be statistically significant for OS and were identified as candidate genes. In order to minimize partial likelihood deviance and prevent overfitting of the model, we performed the least absolute shrinkage and selection operator (LASSO) Cox regression analysis and selected the most effective prognostic ARGs. The selected ARGs were subjected to the multivariate Cox regression analysis to further screen genes which were capable of predicting the prognosis independently. Subsequently, a risk model composed of several ARGs was established using glmnet and survival R packages.

2.4. Validation of ARG-Based Risk Model. The risk score of each patient was developed by a linear combination of

the expression level of genes multiplied regression coefficients calculated by the Cox regression model. HNSCC patients were assigned into the high-risk or low-risk groups according to corresponding risk scores, using the median score as a cutting-off point. A Kaplan–Meier survival curve was plotted to estimate the difference between the two groups using the log-rank test. Meanwhile, the time-dependent receiver operating characteristic (ROC) analysis was employed to evaluate the efficiency of the prognostic signature using the survivalROC package, and the respective area under the ROC curve (AUC) was assessed for predicting accuracy.

To verify if the prognostic ARGs signature could be an independent indicator for predicting the OS of patients with HNSCC, the multivariate Cox regression analysis was performed. Patients were separated into different groups according to sex, age, clinical stage, histological grade, anatomic subdivision, alcohol history, smoking status, HPV status, and tumor mutation burden (TMB) value. These clinicopathological characteristics and the risk score served as covariates and ROC curves were also plotted to evaluate the value in predicting prognosis. Kaplan–Meier survival analysis was also performed in the testing sets from GEO and ICGC cohorts to validate the efficacy of our risk signature. $P$ value < 0.05 was regarded as significant.

## 3. Results

3.1. Differentially Expressed Autophagy-Related Genes. A flowchart was provided to demonstrate the study design
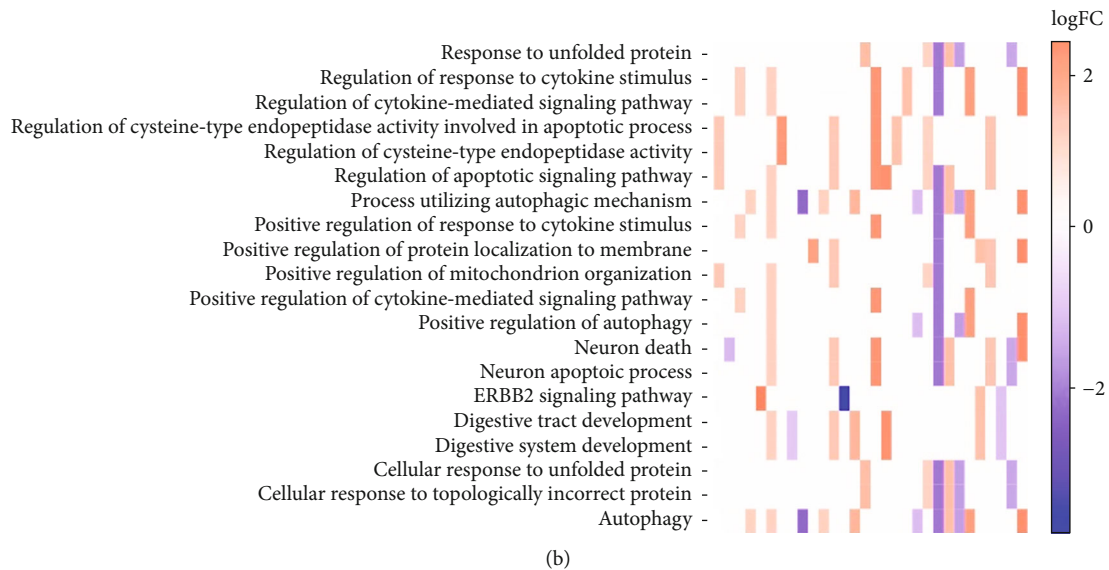
(a)



(b)
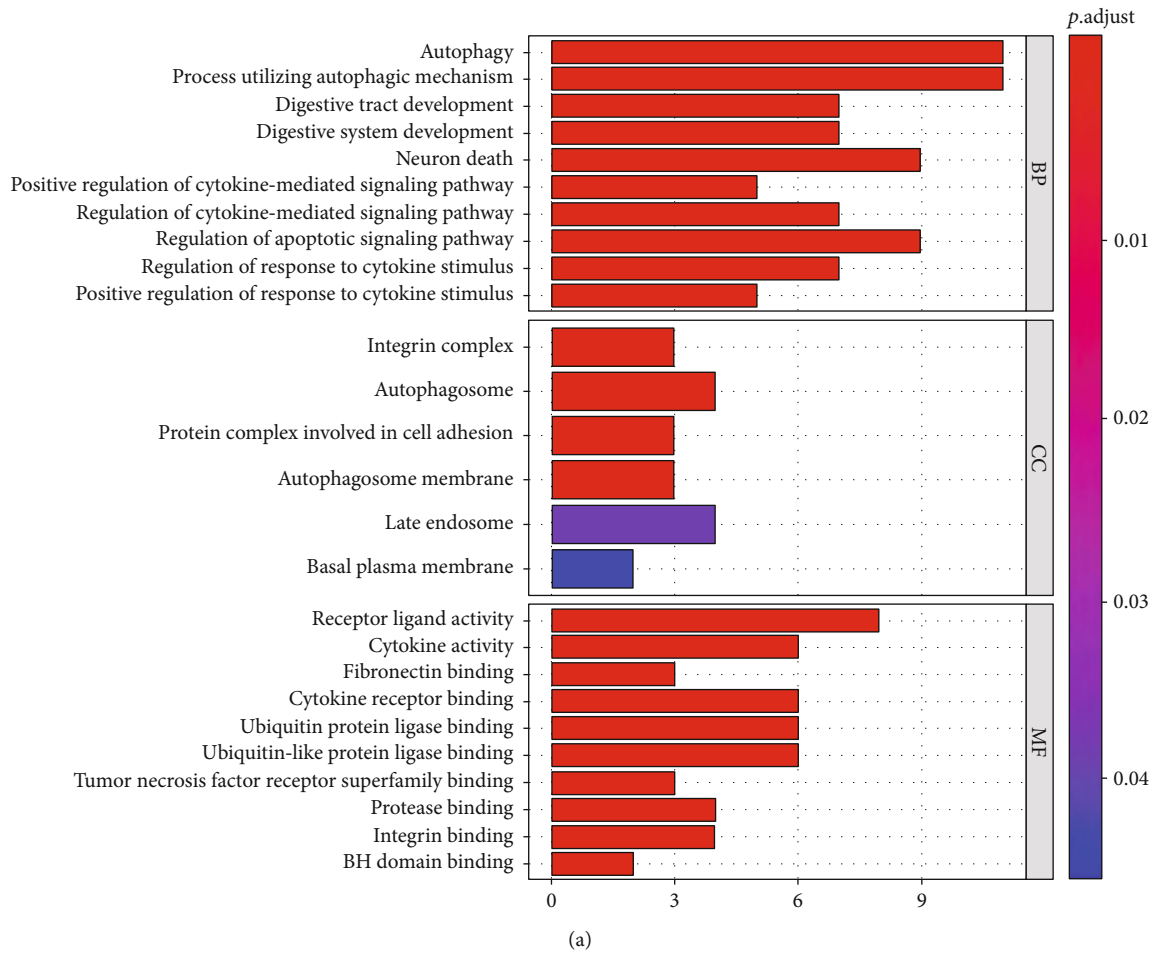
FIGURE 3: The barplot and heat map of gene ontology (GO) enrichment analysis. (a) The relationship between enriched GO terms and differentially expressed autophagy-related genes. BP indicates biological process; CC indicates cellular component; MF indicates molecular function. (b) The color of each block depends on the logFC value of the specified gene.

and analysis (Figure 1). The RNA-seq profiles obtained from public database TCGA consisted of 501 HNSCC tissue samples and 40 nontumor samples. After extracting the expression data of 232 autophagy-related genes in HNSCC patients,10 downregulated genes (NRG2, NRG3, MAP1LC3C, PRKN, HSPB8, CCL2, FOS, TP53INP2, PTK6, and NKX2-3) and 28 upregulated genes (EIF4EBP1, BAK1, RGS19, HIF1A, CXCR4, CTSL, VMP1, SPNS1,

(a)
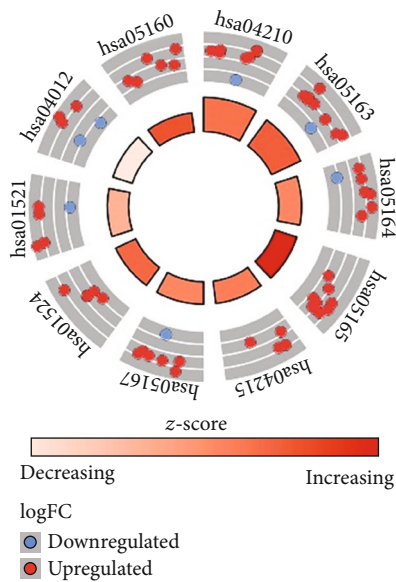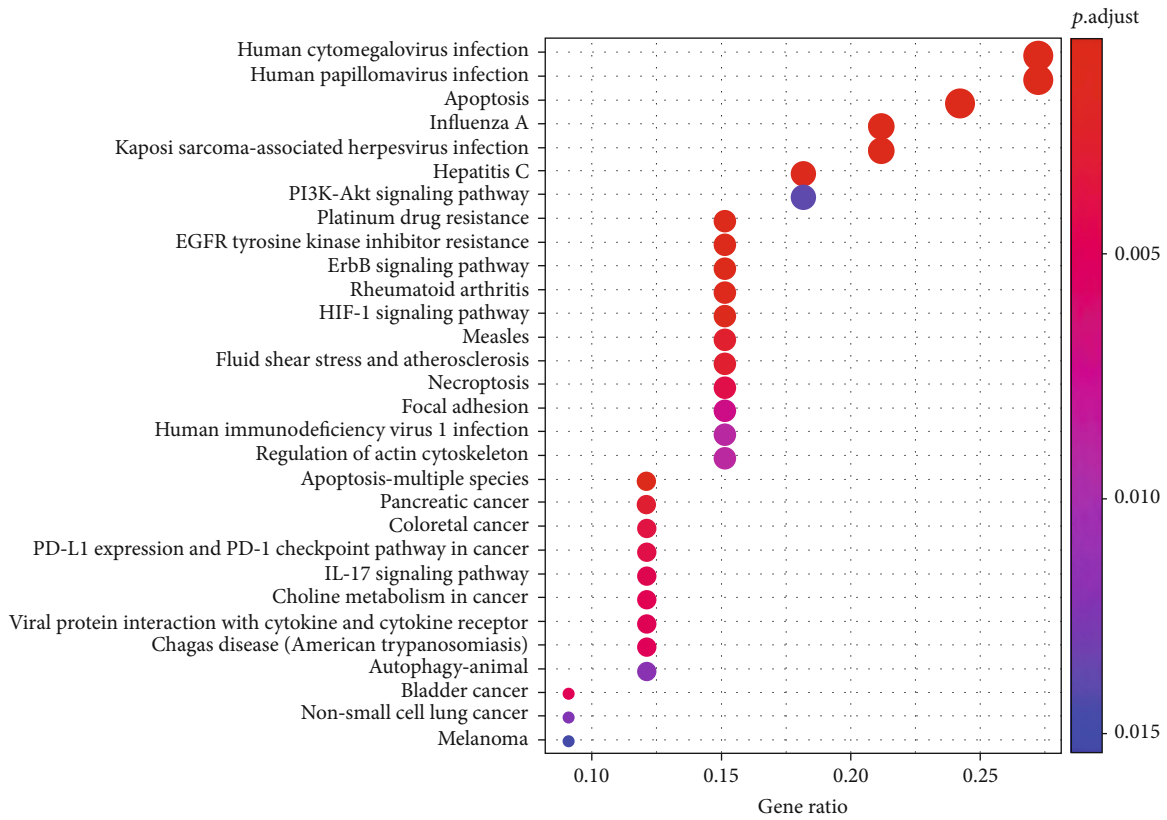


| ID | Description |
|----------|---------------------------------------------------|
| hsa04210 | Apoptosis |
| hsa05163 | Human cytomegalovirus infection |
| hsa05164 | Influenza A |
| hsa05165 | Human papillomavirus infection |
| hsa04215 | Apoptosis-multiple species |
| hsa05167 | Kaposi sarcoma-associated herpesvirus infection |
| hsa01524 | Platinum drug resistance |
| hsa01521 | EGFR tyrosine kinase inhibitor resistance |
| hsa04012 | ErbB signaling pathway |
| hsa05160 | Hepatitis C |

(b)

FIGURE 4: The bubble plot and circle plot of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. (a) The relationship between KEGG pathways and differentially expressed autophagy-related genes. (b) The outer circle shows a scatter plot for each term of the logFC value of the specified gene. Red circles display upregulation, and blue means downregulation. The higher the $Z$-score indicates, the higher expression of the enriched pathway.

TABLE 1: Clinical characteristics of head and neck squamous cell carcinoma patients from TCGA in this study.

| Features | Alive (n = 281) | Dead (n = 214) | Total (n = 495) |
|---|---|---|---|
| Gender | | | |
| Female | 64 | 68 | 132 |
| Male | 217 | 146 | 363 |
| Age | | | |
| Mean (SD) | 59.44 (11.2) | 62.97 (12.4) | 60.97 (11.9) |
| Median [min, max] | 60 [19, 85] | 63 [24, 88] | 61 [19, 88] |
| Anatomic subd | | | |
| Oral cavity | 179 | 149 | 328 |
| Other | 102 | 65 | 167 |
| Clinical stage | | | |
| Stage I | 21 | 4 | 25 |
| Stage II | 50 | 30 | 80 |
| Stage III | 55 | 35 | 90 |
| Stage IVA | 150 | 134 | 284 |
| Stage IVB | 5 | 8 | 13 |
| Stage IVC | 0 | 3 | 3 |
| Grade | | | |
| G1 | 39 | 22 | 61 |
| G2 | 166 | 131 | 297 |
| G3 | 63 | 53 | 116 |
| G4 | 2 | 0 | 2 |
| GX | 9 | 7 | 16 |
| NA | 2 | 1 | 3 |
| Alcohol history | | | |
| Yes | 192 | 138 | 330 |
| No | 83 | 71 | 154 |
| NA | 6 | 5 | 11 |
| Smoking status | | | |
| Never | 67 | 42 | 109 |
| Former/current | 212 | 164 | 376 |
| NA | 2 | 8 | 10 |

SD: standard deviation; sub: subdivision.

TABLE 2: The results of univariate Cox analysis.

| Gene | HR | HR.95L | HR.95H | p value |
|---|---|---|---|---|
| CXCR4 | 0.990886 | 0.98339 | 0.998439 | 0.018111 |
| EGFR | 1.001677 | 1.000173 | 1.003184 | 0.028858 |
| HSPB8 | 0.993644 | 0.987676 | 0.999648 | 0.03803 |
| PRKN | 1.496081 | 1.114563 | 2.008194 | 0.007317 |
| NKX2-3 | 0.889201 | 0.808232 | 0.978281 | 0.01592 |
| CDKN2A | 0.982237 | 0.97227 | 0.992306 | 0.000573 |
| CTSL | 1.004316 | 1.000584 | 1.008062 | 0.023384 |
| FADD | 1.011174 | 1.004614 | 1.017777 | 0.000819 |
| ITGA3 | 1.004492 | 1.001521 | 1.007472 | 0.00302 |

HR: hazard ratio; L: low confidence interval; H: high confidence interval.

TABLE 3: The results of multivariate Cox analysis.

| Gene | Coef | HR | HR.95L | HR.95H | p value |
|---|---|---|---|---|---|
| EGFR | 0.0017 | 1.0017 | 1.0001 | 1.0033 | 0.0397 |
| HSPB8 | -0.0067 | 0.9933 | 0.9876 | 0.9991 | 0.0241 |
| PRKN | 0.6109 | 1.8421 | 1.4074 | 2.4109 | 0.0214 |
| CDKN2A | -0.0144 | 0.9857 | 0.9754 | 0.9960 | 0.0066 |
| FADD | 0.0075 | 1.0075 | 1.0008 | 1.0142 | 0.0277 |
| ITGA3 | 0.0046 | 1.0047 | 1.0016 | 1.0077 | 0.0027 |

Coef: coefficient; HR: hazard ratio; L: low confidence interval; H, high confidence interval.

TNFSF10, TP63, BID, VEGFA, SPHK1, EGFR, SERPINA1, DDIT3, EIF2AK2, ITGB4, ITGA3, APOL1, IRGM, BIRC5, FADD, ITGA6, IFNG, NRG1, IL24, and CDKN2A) were determined. Scatter plots displayed the expression patterns of differentially expressed ARGs between tumor and nontumor tissues (Figure 2).

3.2. Functional Enrichment Analysis of the Differentially Expressed ARGs. Functional enrichment analysis was conducted to provide an in-depth understanding of biological functions of the 38 differentially expressed ARGs. The GO term functional enrichment and the KEGG pathway enrichment of these genes are summarized in Figures 3 and 4. The top enriched GO terms for biological processes were autophagy, process utilizing autophagic mechanism, neuron death, and regulation of apoptotic signaling pathway. Cellular components had integrin complex, autophagosome, protein complex involved in cell adhesion, and autophagosome membrane. Based on molecular function, genes were mostly enriched in terms of receptor ligand activity, cytokine activity, cytokine receptor binding, ubiquitin protein ligase binding, and ubiquitin-like protein ligase binding (Figure 3). The KEGG pathway enrichment analysis revealed that these genes were notably associated with pathways in apoptosis, human cytomegalovirus infection, and human papillomavirus infection. Most of the $Z$-scores of enriched pathways were more than zero, indicating that most of the pathways were more likely to be enhanced (Figure 4).

3.3. Identification of an Autophagy-Related Gene Signature for the Prognosis of HNSCC. A total of 495 eligible HNSCC patients based on TCGA database were involved in this study, and the baseline of clinical characteristics was presented (Table 1). To investigate differentially expressed ARGs for clinical prognosis, univariate Cox regression analysis was performed and 9 ARGs with $p$ value $< 0.05$ were selected from the training set (Table 2). The candidate genes were passed on for LASSO regression. However, none of them were removed, suggesting that all 9 ARGs may provide specific information for prognosis (Figure S1). Subsequently, a multivariate Cox analysis was conducted, and six genes were identified to develop the prognostic signature (Table 3). The prognostic model was established based on 6 ARG expression levels and coefficients using the following formula: prognosis index (PI) = (0.0017 * EGFR) + (−0.0067
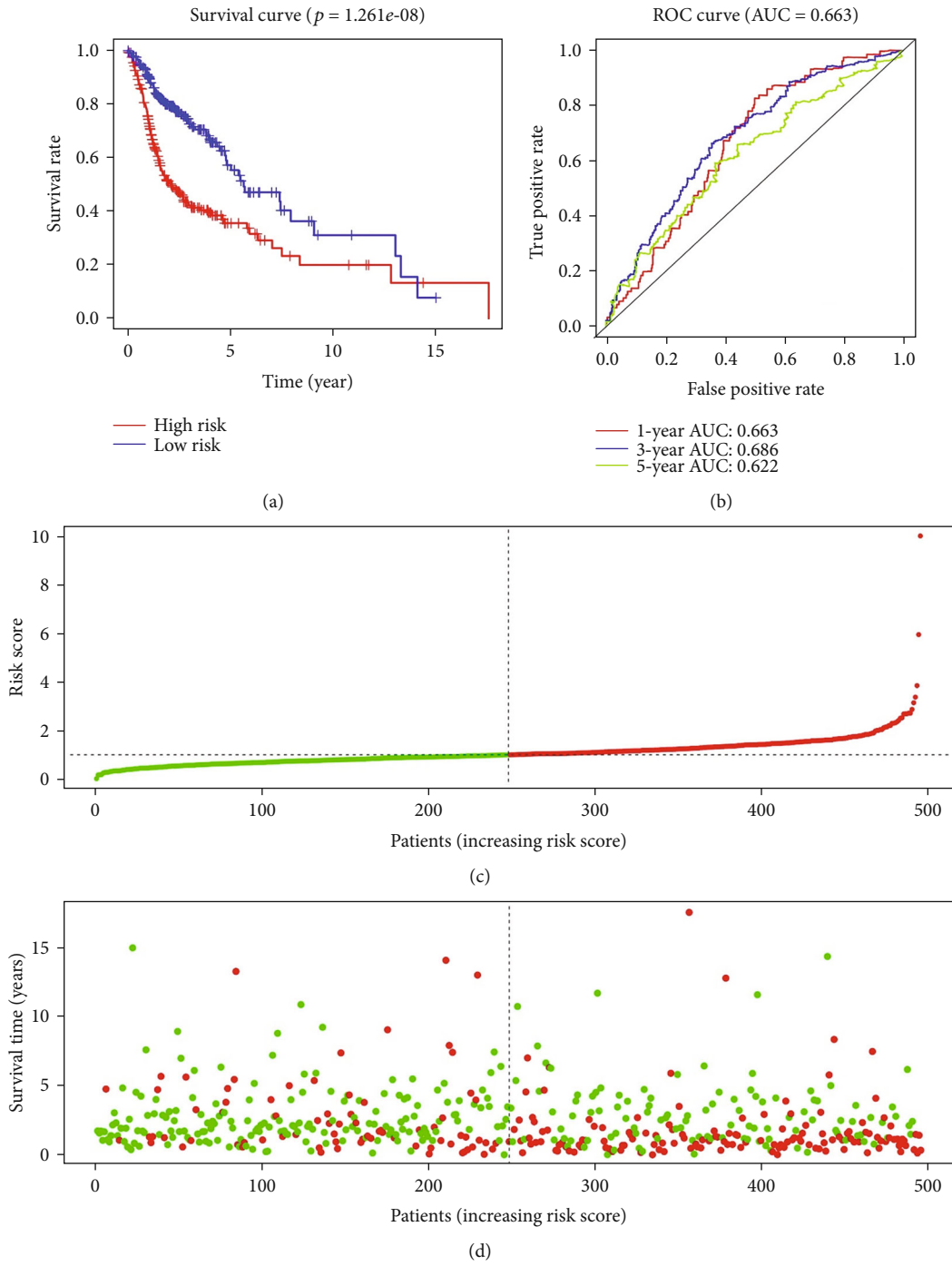
(a)



(b)
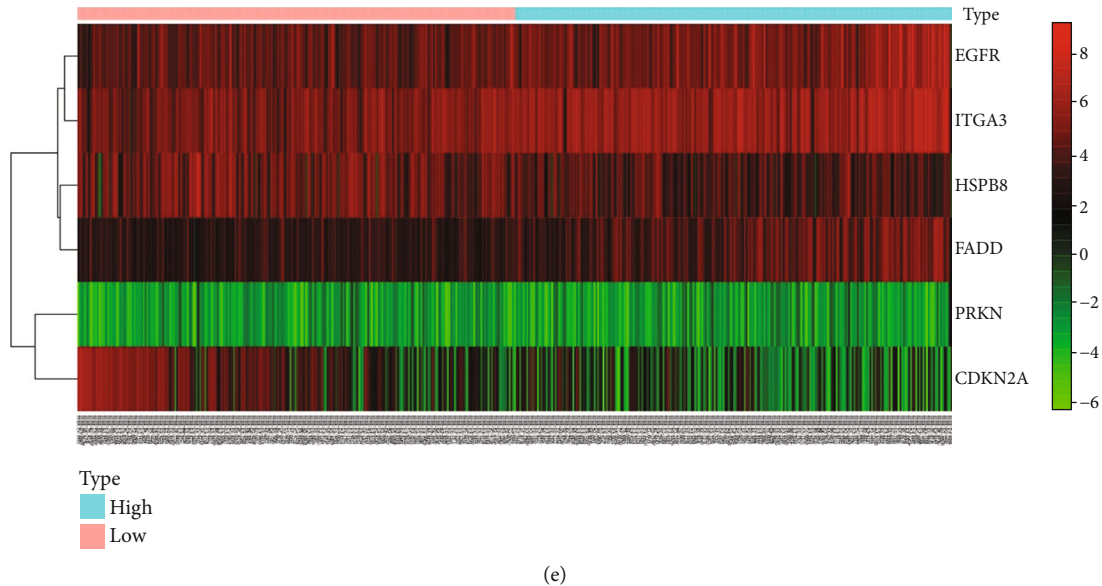


(c)



(d)

Figure 5: Continued.

(e)

FIGURE 5: (a) The Kaplan–Meier plot of the overall survival (OS) for high-risk and low-risk patient cohorts divided by the 6-ARGs signature in the training set ($N = 495$). The OS differences are determined by the two-sided log-rank test. (b) Receiver operating characteristic analysis for the 6-ARG signature in predicting the patients of 1, 3, and 5 years OS in the training set. (c) The distribution of the risk score. Each dot represents a specific patient. The color red indicates high-risk level, and green indicates low-risk level. (d) The distribution of patients' survival status. Each dot represents a specific patient. The color red indicates dead, and green indicates alive. (e) The heat map displays the relationship between risk score and respective gene expression level of this signature in the training set.

$* \mathrm{HSPB8}) + (0.6109 * \mathrm{PRKN}) + (-0.0144 * \mathrm{CDKN2A}) + (0.0075 * \mathrm{FADD}) + (0.0046 * \mathrm{ITGA3})$.

Based on the median expression value of PI, patients in the training cohort were stratified into the high-risk and low-risk groups with specific cut-off point at 1.023. Survival analysis indicated that patients in the high-risk group had a significantly worse prognosis (Figure 5(a)). The results of ROC analysis revealed that this PI model had good accuracy and efficiency in predicting the 1-, 3-, and 5-year survival with respective areas under the ROC curve (AUCs) which were 0.663, 0.686, and 0.622 (Figure 5(b)). With an increasing risk score came shorter overall survival and more death events (Figures 5(c) and 5(d)). The heat map was also performed to evaluate the relationship between risk score and respective gene expression level in this signature (Figure 5(e)). EGFR, FADD, and ITGA3 were upregulated in the high-risk group (Figures 6(a), 6(e), and 6(f)), which indicated they acted as risk factors, while HSPB8 and CDKN2A were discovered to be downregulated in the high-risk group (Figures 6(b) and 6(d)), indicating that they were protective factors for HNSCC. However, we did not observe a significant difference of PRKN expression between the high-risk and low-risk groups (Figure 6(c)). Kaplan–Meier survival analysis of the 6 ARGs demonstrated similar results. The high expression of EGFR, FADD, and ITGA3 was significantly correlated with inferior overall survival of HNSCC patients (Figures 7(a), 7(e), and 7(f)). In contrast, the downregulation of HSPB8 and CDKN2A indicated inferior OS (Figures 7(b) and 7(d)). Also, there was no significant difference in OS concerning PRKN expression level (Figure 7(c)). Since each gene in this risk model played a

different role in predicting prognosis, there may not be of statistical difference for every single gene in above analysis.

*3.4. Correlation between the Autophagy-Related Gene Signature and Clinicopathologic Characteristics in HNSCC Patients.* Considering that clinical parameters may influence the performance of the risk signature, the patients were also grouped by sex, age, clinical stage, histological grade, anatomic subdivision, alcohol history, and smoking status and to explore their association with the 6-ARG signature. Patients with incomplete grade, alcohol history, and smoking information were removed, and 457 samples were retained for this procedure. The results suggested that the signature was significantly correlated with clinical stage ($p = 0.037$, Figure 8(a)). Besides, the 6 ARGs displayed different expression with respect to various clinicopathologic features. Different expression of CDKN2A was found across different sexes, ages, grades, and anatomic subdivisions (Figures 8(b)–8(e)). HSPB8 was expressed differentially with different anatomic sites and alcohol histories (Figures 8(f) and 8(g)). Higher PRKN expression was observed in male and stage III and IV HNSCC patients (Figures 8(h) and 8(i)), and higher FADD expression was found in patients with higher clinical stages and tobacco use (Figures 8(j) and 8(k)). ITGA3 also showed different expression across different sexes, anatomic subdivisions, and smoking statuses (Figures 8(l)–8(n)). Since HPV has been revealed to have some impacts on tumor progression in HNSCC [11], we also took HPV status into analysis. Because many patients in TCGA cohort were lack of HPV status information, only 90 samples were selected. As shown in
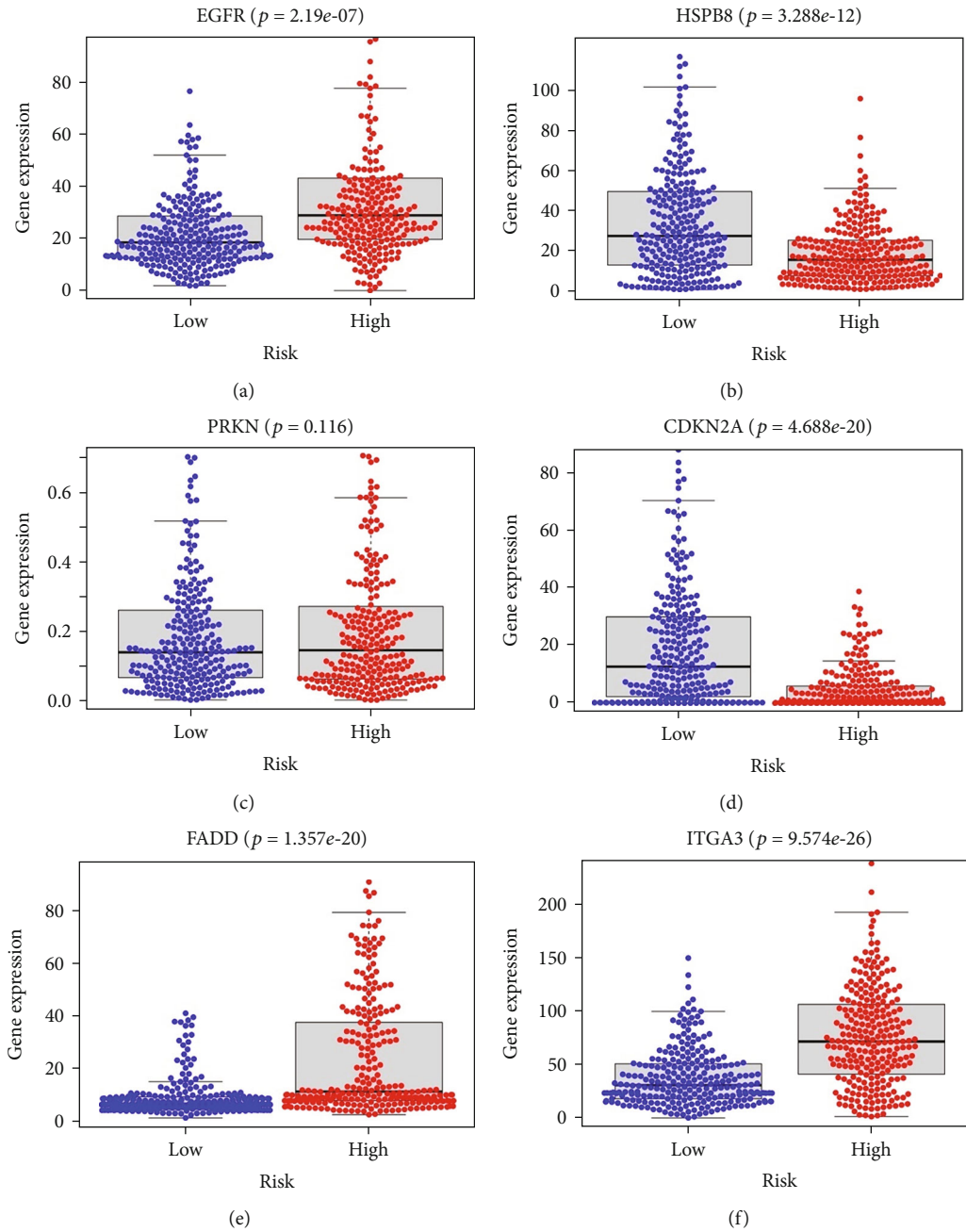
(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 6: Different expression of the six key genes between the high-risk group and low-risk group.

Figure 9, higher expression levels of EGFR and ITGA3 were discovered in patients with HPV-negative HNSCC, while higher CDKN2A expression was found in patients with HPV-positive HNSCC.

*3.5. The Autophagy-Related Gene Signature Is an Independent Prognostic Indicator in Patients with HNSCC.* To identify other possible contributors on overall survival, the patients in different clinical subgroups were used for univariate and multivariate Cox regression analyses. Tumor mutation burden (TMB) has been shown to correlate with cancer aggressiveness and immunotherapeutic response

[12, 13] and thus was also included in this procedure. 450 HNSCC patients with available TMB information were stratified in the high/low TMB group based on median TMB value and were selected for further study. The results suggested that the autophagy-related gene signature could independently predict prognosis of HNSCC (Figures 10(a) and 10(b)). It was also noticed that age and clinical stage were also significantly associated with patient survival. However, ROC analysis demonstrated that age (AUC = 0.586) and stage (AUC = 0.554) alone were not reliable indicators while risk score signature (AUC = 0.665) kept a stable and reliable performance (Figure 10(c)). Likewise, we brought HPV status
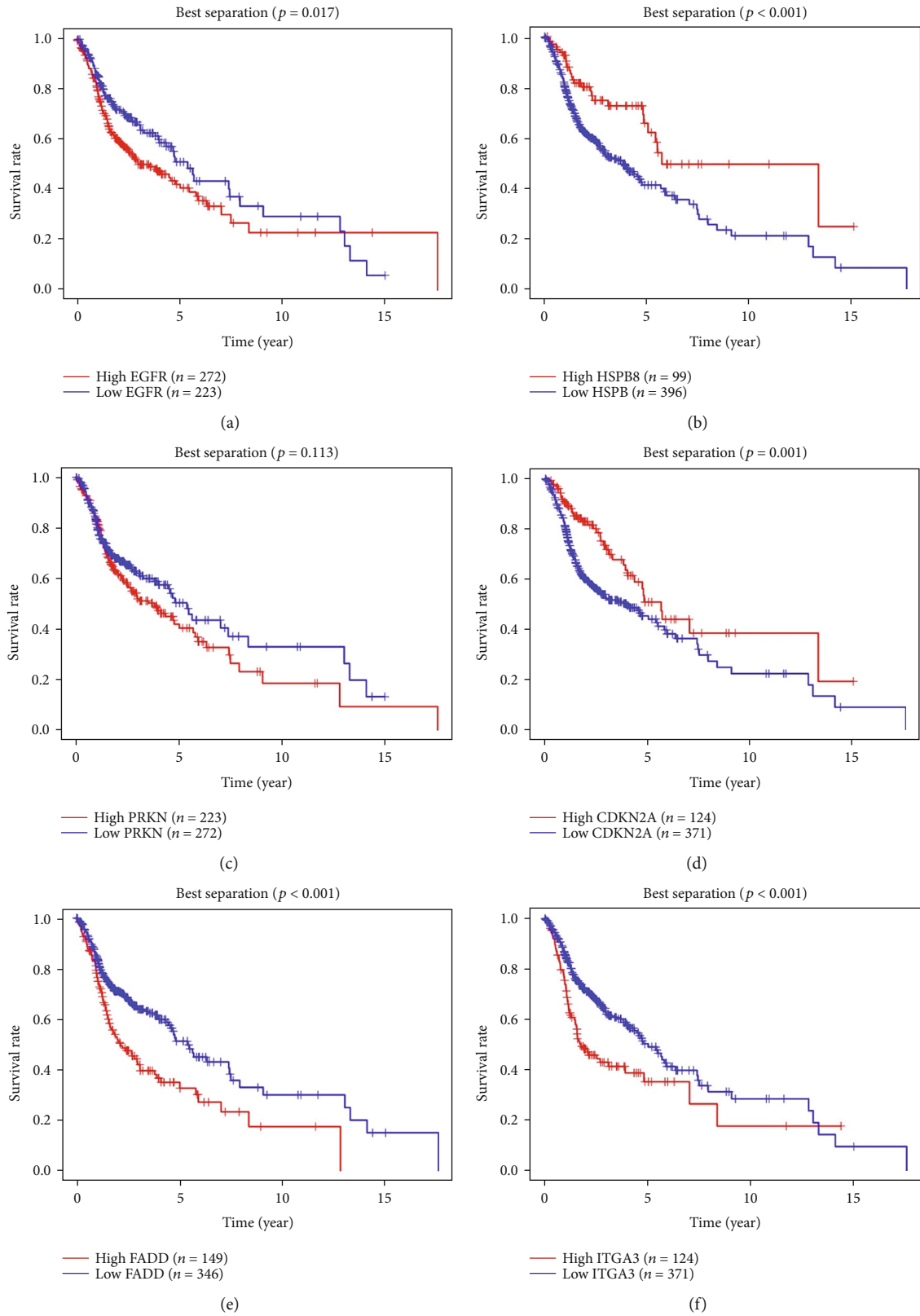
Figure 7: The correlation between the six genes included in the signature and HNSCC patients' overall survival. Kaplan–Meier plots demonstrate results from the analysis of correlation between each gene expression level and OS, all using the best separation.
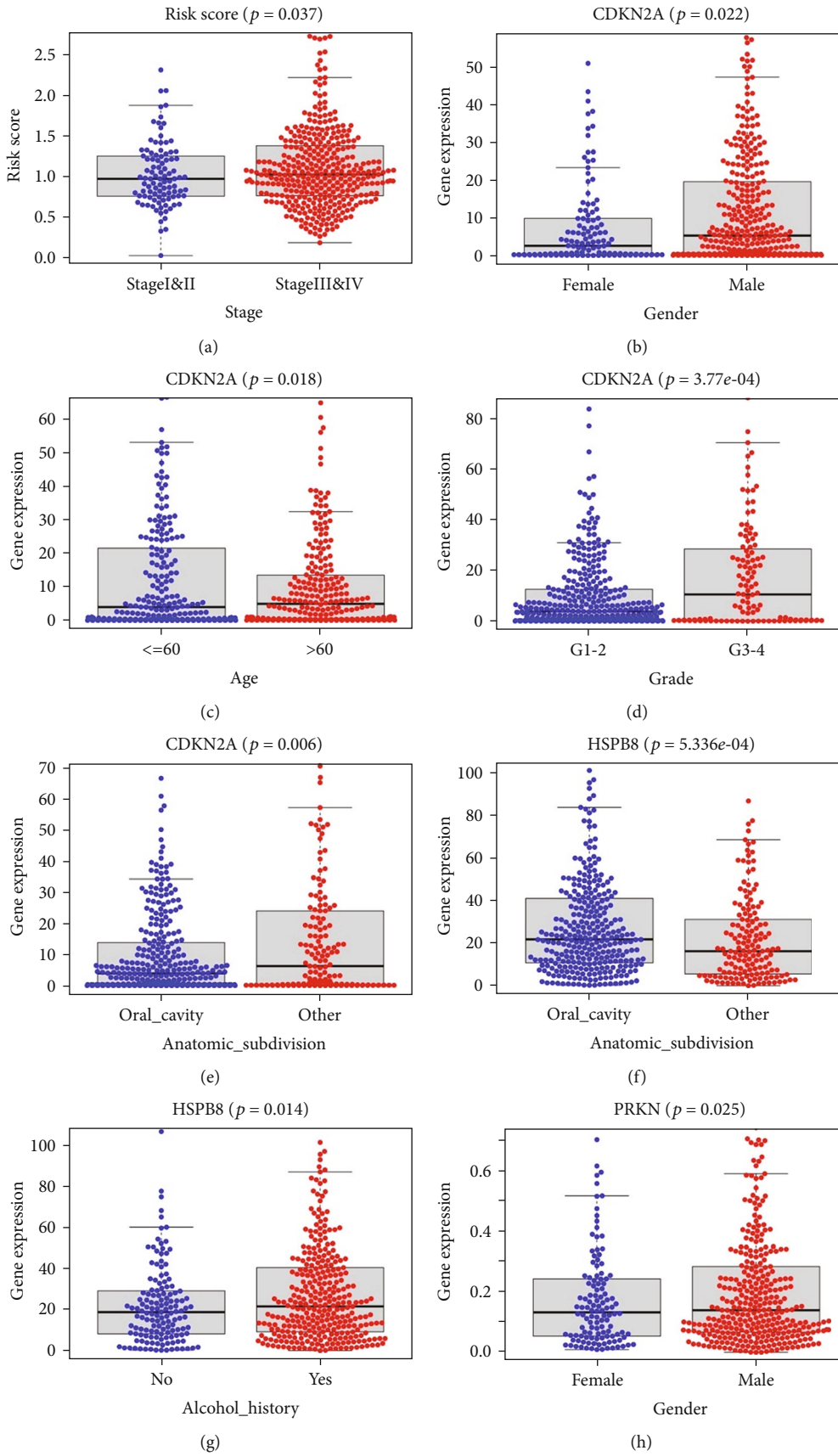
(a)

(b)

(c)

(d)

(e)
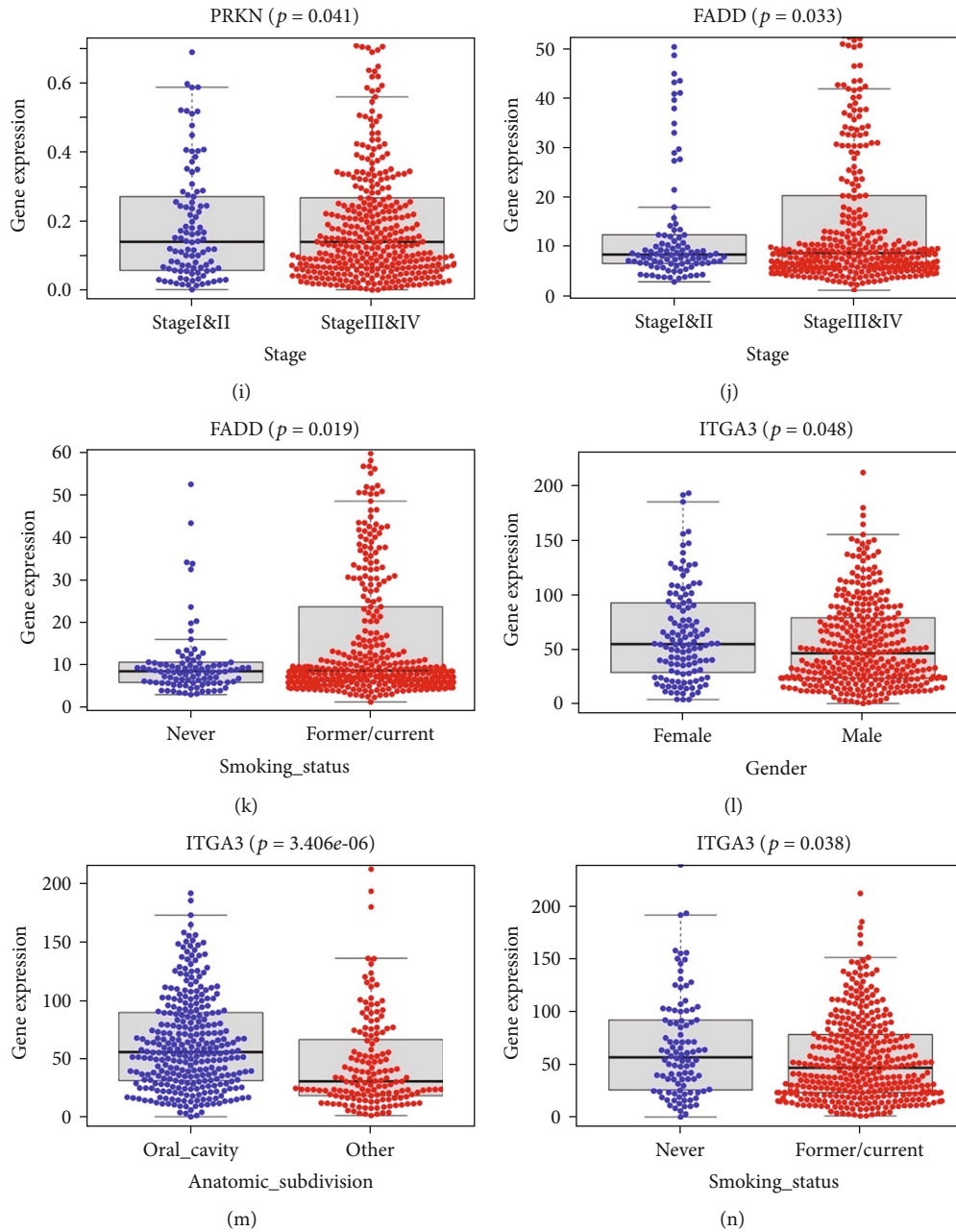
(f)

(g)

(h)

Figure 8: Continued.

FIGURE 8: (a) The 6-ARG signature in the cohort stratified by clinical stages. (b–e) CDKN2A expression in the cohorts stratified by sexes, ages, grades, and anatomic subdivisions. (f, g) HSPB8 expression in the cohorts stratified by anatomic sites and alcohol histories. (h, i) PRKN expression in the cohorts stratified by sexes and clinical stages. (j, k) FADD expression in the cohorts stratified by clinical stages and smoking statuses. (l–n) ITGA3 expression in the cohorts stratified by sexes, anatomic subdivisions, and smoking statuses.

into Cox regression analysis and 88 samples were selected. The results also indicated that the risk signature was an independent prognostic factor for HNSCC patients with even better predicting accuracy (AUC = 0.709, Figure 11). Besides, we identified the most frequently mutated genes in patients with HNSCC (Figure 12(a)). Researches have indicated TP53 to be the most commonly mutated gene in HNSCC [14, 15], which was also confirmed in our study. Since mutational status of the samples might influence the performance of the risk signature, the patients were grouped by mutational

status of TP53. The high-risk groups showed significantly shorter OS in both HNSCC patients with TP53 mutation and patients without TP53 mutation (Figures 12(b) and 12(c)), indicating the independence of the risk signature once again. In conclusion, the 6-ARG signature could be applied as an independent prognostic indicator for HNSCC in clinical practice.

3.6. Validation of the Autophagy-Related Gene Signature in Independent HNSCC Cohorts. Three independent datasets
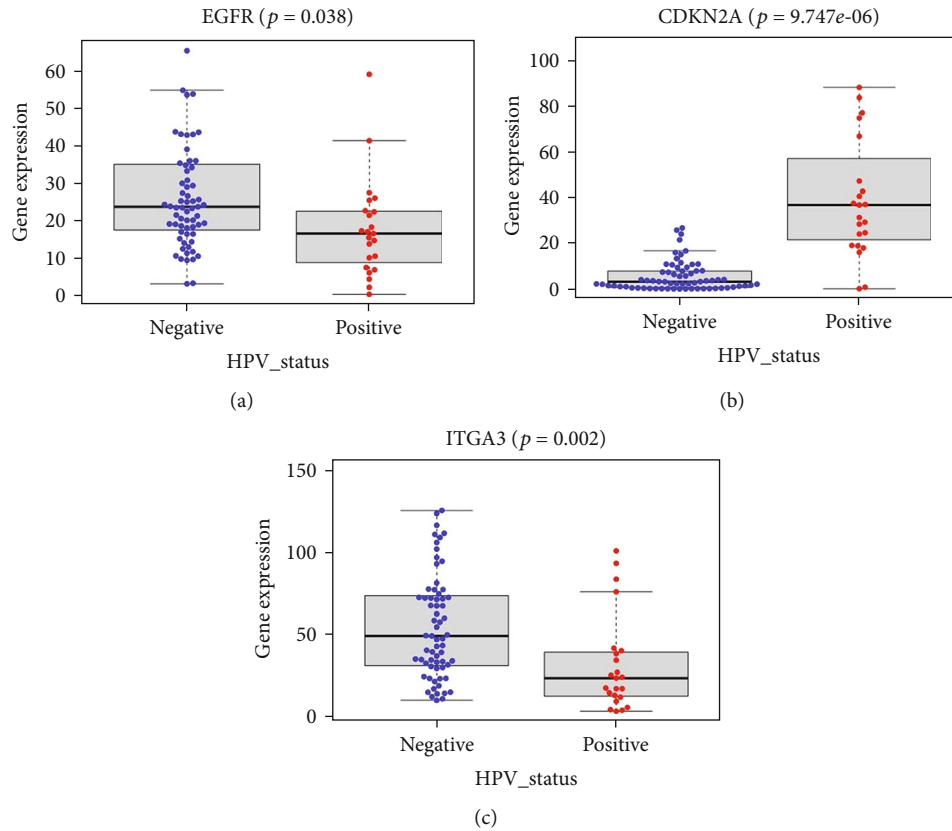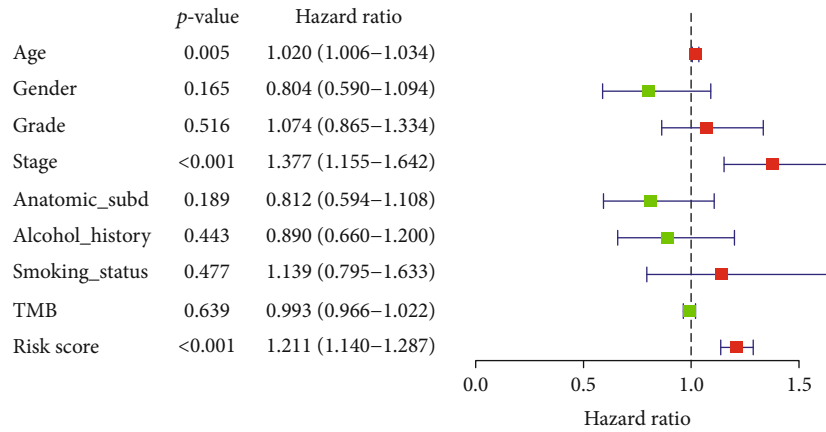
FIGURE 9: (a) EGFR expression in the cohorts stratified by HPV status. (b) CDKN2A expression in the cohorts stratified by HPV status. (c) ITGA3 expression in the cohorts stratified by HPV status.

GSE41613, GSE117973, and ORCA-IN (ICGC) were used as external validation groups. The risk score for each patient was calculated using the same PI formula. The patients were stratified into the high-risk and low-risk groups based on the median risk score. We performed Kaplan–Meier survival analysis in the three testing groups, which confirmed the prognostic value of our gene signature. As expected, patients in the high-risk groups from GSE41613 and ORCA-IN datasets showed significantly inferior overall survival (Figures 13(a) and 13(c)). The survival information of GSE117973 dataset was progression-free survival (PFS). Similarly, the high-risk HNSCC patients showed inferior PFS (Figure 13(e)). A good prognostic ability was also observed based on the time-dependent ROC analysis for the GSE41613 cohort in 1-, 3-, and 5-year follow-ups with respective AUC values which were 0.613, 0.671, and 0.631 (Figure 13(b)) and GSE117973 cohort in 1-, 3-, and 5-year follow-ups with respective AUC values which were 0.654, 0.66, and 0.604 (Figure 13(f)). Since there were only a limited number of samples acquired from ICGC and the follow-up information of most patients was restricted within 2 years, we just demonstrated the ROC curve of the 1-year follow-up (AUC = 0.838, Figure 13(d)), which suggested great predicting ability of this signature for HNSCC patients' prognosis once again. Therefore, the 6-ARG-based risk signature was proved to show great competence in predicting prognosis of HNSCC.
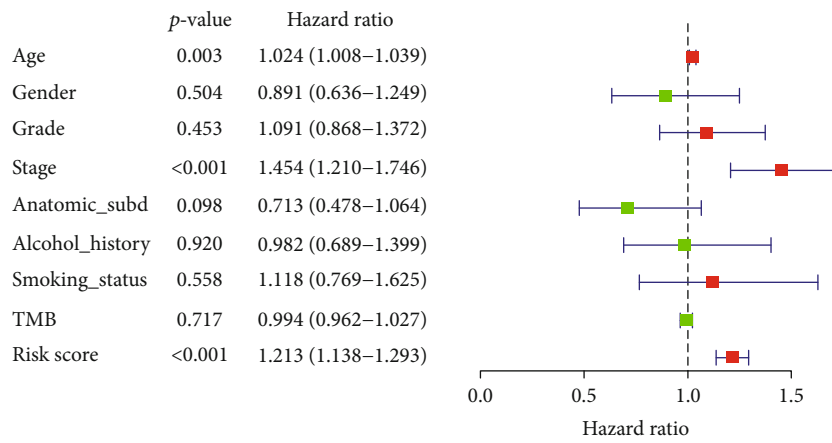
## 4. Discussion

HNSCC is one of the most prevailing and life-threatening cancers worldwide. Despite considerable improvements in diagnosis and treatment, the survival rate of HNSCC remains low. Besides, treatment is accompanied by significant long-term toxicity and morbidity [16]. Numerous studies have shown that autophagy plays a dual role in occurrence and progression of tumors. On one hand, autophagy promotes genomic stability and anticancer immunosurveillance in tumor suppression. On the other hand, it boosts high metabolic activity to sustain tumor cell survival and executives cytoprotective function under environmental stress in tumor progression [17]. Recent studies have indicated that various autophagy-related genes and corresponding protein expression profiles are related to clinicopathological features or prognosis of OSCC, including LC3, BECN1, ATG16L1, ATG9A, SQSTM1, and ATG5 [18]. However, most research mainly focuses on one particular gene related to autophagy.

The large-scale databases, such as TCGA and GEO, provide us with abundant information and effective measures to explore gene signatures. In this study, we screened autophagy-related genes and identified six key prognostic genes, which could be potential biomarkers and therapeutic targets. The 6-ARG signature was capable of distinguishing patients of different risk and prognosis, which was validated

|  | *p*-value | Hazard ratio |
| --- | --- | --- |
| Age | 0.005 | 1.020 (1.006−1.034) |
| Gender | 0.165 | 0.804 (0.590−1.094) |
| Grade | 0.516 | 1.074 (0.865−1.334) |
| Stage | <0.001 | 1.377 (1.155−1.642) |
| Anatomic_subd | 0.189 | 0.812 (0.594−1.108) |
| Alcohol_history | 0.443 | 0.890 (0.660−1.200) |
| Smoking_status | 0.477 | 1.139 (0.795−1.633) |
| TMB | 0.639 | 0.993 (0.966−1.022) |
| Risk score | <0.001 | 1.211 (1.140−1.287) |

Hazard ratio

(a)

|  | *p*-value | Hazard ratio |
| --- | --- | --- |
| Age | 0.003 | 1.024 (1.008−1.039) |
| Gender | 0.504 | 0.891 (0.636−1.249) |
| Grade | 0.453 | 1.091 (0.868−1.372) |
| Stage | <0.001 | 1.454 (1.210−1.746) |
| Anatomic_subd | 0.098 | 0.713 (0.478−1.064) |
| Alcohol_history | 0.920 | 0.982 (0.689−1.399) |
| Smoking_status | 0.558 | 1.118 (0.769−1.625) |
| TMB | 0.717 | 0.994 (0.962−1.027) |
| Risk score | <0.001 | 1.213 (1.138−1.293) |

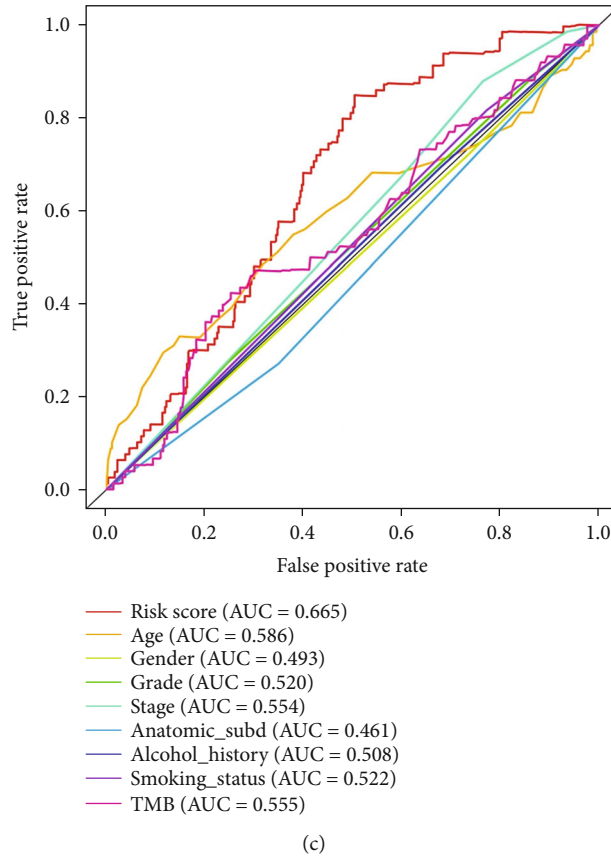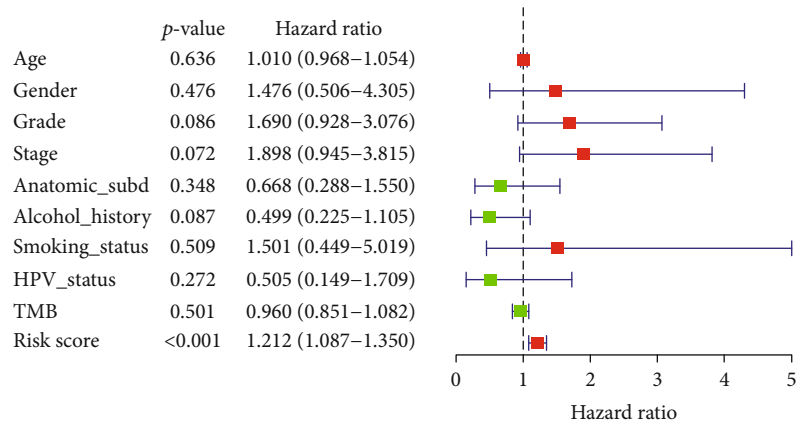Hazard ratio

(b)

Figure 10: Continued.

FIGURE 10: The forest plots of univariate (a) and multivariate (b) Cox analyses display the correlation of different indexes and overall survival of HNSCC patients. (c) Receiver operating characteristic analysis for risk signature and clinicopathological features in predicting HNSCC patients' OS in TCGA cohort.

in multiple datasets and was proved to own great prognostic ability, thus providing us with a novel insight in HNSCC. The GO and KEGG analyses were performed to exploit the molecular and biological information of in-depth ARGs. The results of both functional analyses revealed that the top enriched terms were closely related to autophagy. It was also reported that autophagy contributed to membrane trafficking and signaling pathways in a diverse way [6]. Moreover, there was an increase of enriched KEGG pathways in human cytomegalovirus infection (HCMV) and human papillomavirus infection (HPV), indicating the interplay between autophagy and the immune microenvironment. Previous observations have indicated that HPV-associated HNSCC patients tended to demonstrate improved survival compared to patients with HPV-negative HNSCC. Integration of the HPV genome into the host's genome could generate remarkable downstream consequences in immune response, especially an enhanced infiltration of immune cells and inflammatory cytokines in the HPV-positive tumor microenvironment [19, 20]. This was consistent with our results, because the risk factors in the gene signature EGFR and ITGA3 were downregulated while the protective factor CDKN2A was upregulated in HPV-positive HNSCC patients. However, our study did not observe a significant
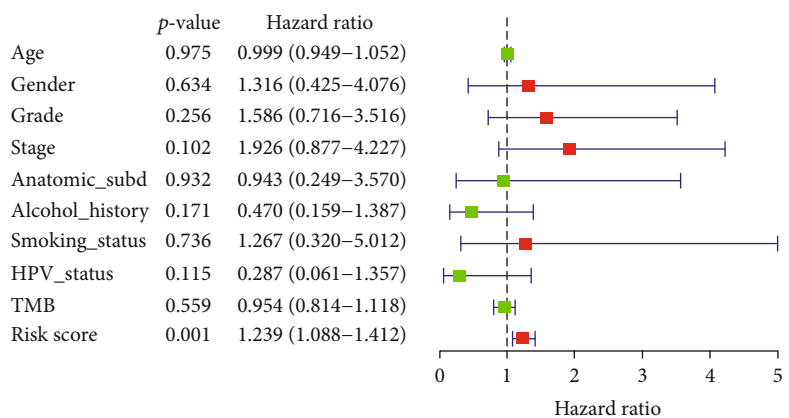
correlation between HPV and prognosis of HNSCC patients. The reasons may be that HNSCC were a heterogeneous group of malignancies and HPV was detected in 25.9% of all HNSCC, mainly the oropharyngeal squamous cell cancer subtypes [21, 22].What is more, there were only a limited number of HNSCC patients equipped with available HPV status information from TCGA database that may also affect the results. HCMV was regarded as a factor for tumorigenesis and has been revealed to have connection with several cancers including malignant glioma, cervical carcinoma, Kaposi's sarcoma, and breast cancer [23]. Some of the HCMV gene products and proteins were proved to accelerate cancer progression via certain pathways, including suppression of the local immune response against tumors and promotion of cell apoptosis [24].

The Cox regression survival analysis helped us to identify six key prognostic ARGs and construct a prognostic signature, which could be an independent indicator for HNSCC patients' OS. We observed a decreasing expression level of *HSPB8* and *PRKN* and an increasing expression level of *EGFR*, *CDKN2A*, *FADD*, and *ITGA3* in tumor samples. Conversely and interestingly, *PRKN* was upregulated while *CDKN2A* was downregulated in the high-risk group. According to the results, we speculate

| | p-value | Hazard ratio |
|---|---|---|
| Age | 0.636 | 1.010 (0.968−1.054) |
| Gender | 0.476 | 1.476 (0.506−4.305) |
| Grade | 0.086 | 1.690 (0.928−3.076) |
| Stage | 0.072 | 1.898 (0.945−3.815) |
| Anatomic_subd | 0.348 | 0.668 (0.288−1.550) |
| Alcohol_history | 0.087 | 0.499 (0.225−1.105) |
| Smoking_status | 0.509 | 1.501 (0.449−5.019) |
| HPV_status | 0.272 | 0.505 (0.149−1.709) |
| TMB | 0.501 | 0.960 (0.851−1.082) |
| Risk score | <0.001 | 1.212 (1.087−1.350) |

(a)

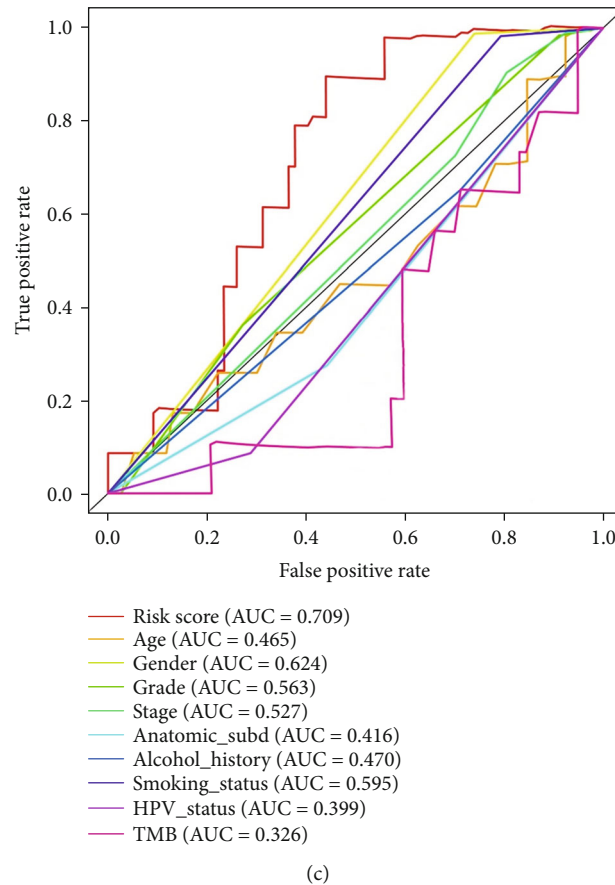| | p-value | Hazard ratio |
|---|---|---|
| Age | 0.975 | 0.999 (0.949−1.052) |
| Gender | 0.634 | 1.316 (0.425−4.076) |
| Grade | 0.256 | 1.586 (0.716−3.516) |
| Stage | 0.102 | 1.926 (0.877−4.227) |
| Anatomic_subd | 0.932 | 0.943 (0.249−3.570) |
| Alcohol_history | 0.171 | 0.470 (0.159−1.387) |
| Smoking_status | 0.736 | 1.267 (0.320−5.012) |
| HPV_status | 0.115 | 0.287 (0.061−1.357) |
| TMB | 0.559 | 0.954 (0.814−1.118) |
| Risk score | 0.001 | 1.239 (1.088−1.412) |

(b)

Figure 11: Continued.

(c)

FIGURE 11: The forest plots of univariate (a) and multivariate (b) Cox analyses display the correlation of different indexes and overall survival of HNSCC patients with available HPV status information. (c) Receiver operating characteristic analysis for risk signature and clinicopathological features in predicting HNSCC patients' OS.

that the performance of autophagy is not immutable and autophagy plays a complicated role in various kinds of tumors [25].

Indeed, four genes involved in this prognostic signature have been shown to be significantly correlated with carcinogenesis of HNSCC. The epidermal growth factor receptor (EGFR) is a cell surface receptor member of the ErbB family, which has been extensively studied in HNSCC. Activation of EGFR leads to the promotion of proliferation, differentiation, antiapoptotic signaling, angiogenesis, and metastatic spread and thus is highly correlated with tumor progression [26, 27]. EGFR is reported to be overexpressed in more than 90% of HNSCC and is an independent prognostic indicator which is associated with increased tumor size, shorter progression-free survival, and decreased overall survival [28–30]. This has made EGFR a promising therapeutic target and led to the development of monoclonal antibodies and tyrosine kinase inhibitors in anti-EGFR therapy [31]. Currently, the EGFR monoclonal antibody cetuximab is the only target drug that has been approved by the US FDA in the treatment for HNSCC [32].*CDKN2A*, a tumor-suppressor gene which encodes the cell cycle regulator p16INK4a, has been extensively investigated to be associated with HNSCC. Global genomic analyses suggest *CDKN2A* being one of the most commonly altered genes in HNSCC and noncoding

mutations of *CDKN2A* correlates with worse overall survival in HNSCC patients [33, 34]. The results of immunohistochemistry suggest that oropharyngeal tumors with p16INK4a protein expression demonstrate better prognosis than p16-negative oropharyngeal tumors [35]. Meanwhile, HPV-positive patients with p16INK4a expression have better response to chemotherapy and radiotherapy compared with HPV-negative patients, during the standard multimodality treatments [36]. FADD is originally described as an adapter molecule mediating apoptosis, which has been newly revealed to engage in cell survival and tumor development [37, 38]. High FADD expression has been identified frequently in HNSCC (>30%) and is related to a higher incidence of lymph node metastasis and shorter distant metastasis-free internal [39]. Moreover, pharmacologic modulation of *FADD* is found to be effective and promising in HNSCC patients [40]. ITGA3 is a member of the integrin family and serves as a cell surface adhesion protein [41]. Research has demonstrated that knockdown of gene *ITGA3* inhibits HNSCC cancer cell migration and invasion. Furthermore, high expression of *ITGA3* indicates poorer survival of patients with HNSCC [42, 43].

So far, existing research also has revealed some mechanism of the other two identified ARGs, though their interaction with HNSCC is rudimentary and inconclusive. PRKN,
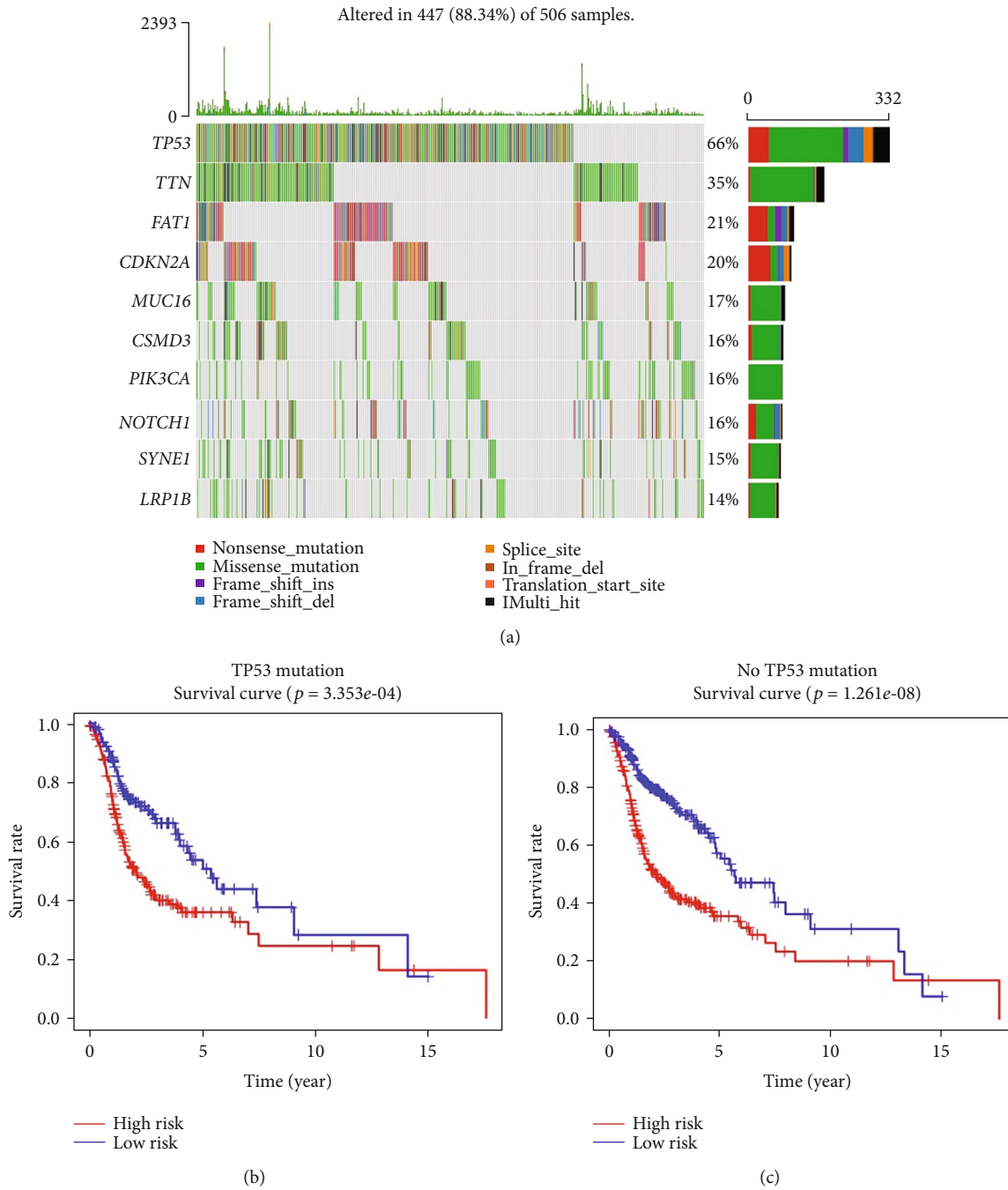
(a)



(b)



(c)

Figure 12: (a) Top ten genes with the highest mutation rates in HNSCC patients from TCGA database. (b) The results of Kaplan–Meier analysis in HNSCC patients with TP53 mutation. (c) The results of Kaplan–Meier analysis in HNSCC patients without TP53 mutation. The OS differences are determined by the two-sided log-rank test.

also named as PARK2 or PARKIN, is an E3 ubiquitin ligase that functions in mitophagy and xenophagy. Biological studies demonstrate that *PRKN* possesses both prosurvival and growth suppressive capacities and its function as a tumor promoter or suppressor is highly dependent on cancer subtype and context [44, 45]. *PRKN* has been found to be affected in tumor microenvironmental signaling networks with likely loss of expression in HNSCC tumor samples

[46]. Decreased expression of *PRKN* has also been identified in oropharyngeal squamous cell carcinoma (OPSCC) samples compared with paired normal tissues [47], which is consistent with our findings. But how *PRKN* behaves in tumor progression especially in HNSCC is still indefinite.

HSPB8 belongs to the ubiquitous small heat shock protein (sHSP) family, which cooperates with BAG3 and stimulate autophagic flux to seek out damaged cellular components
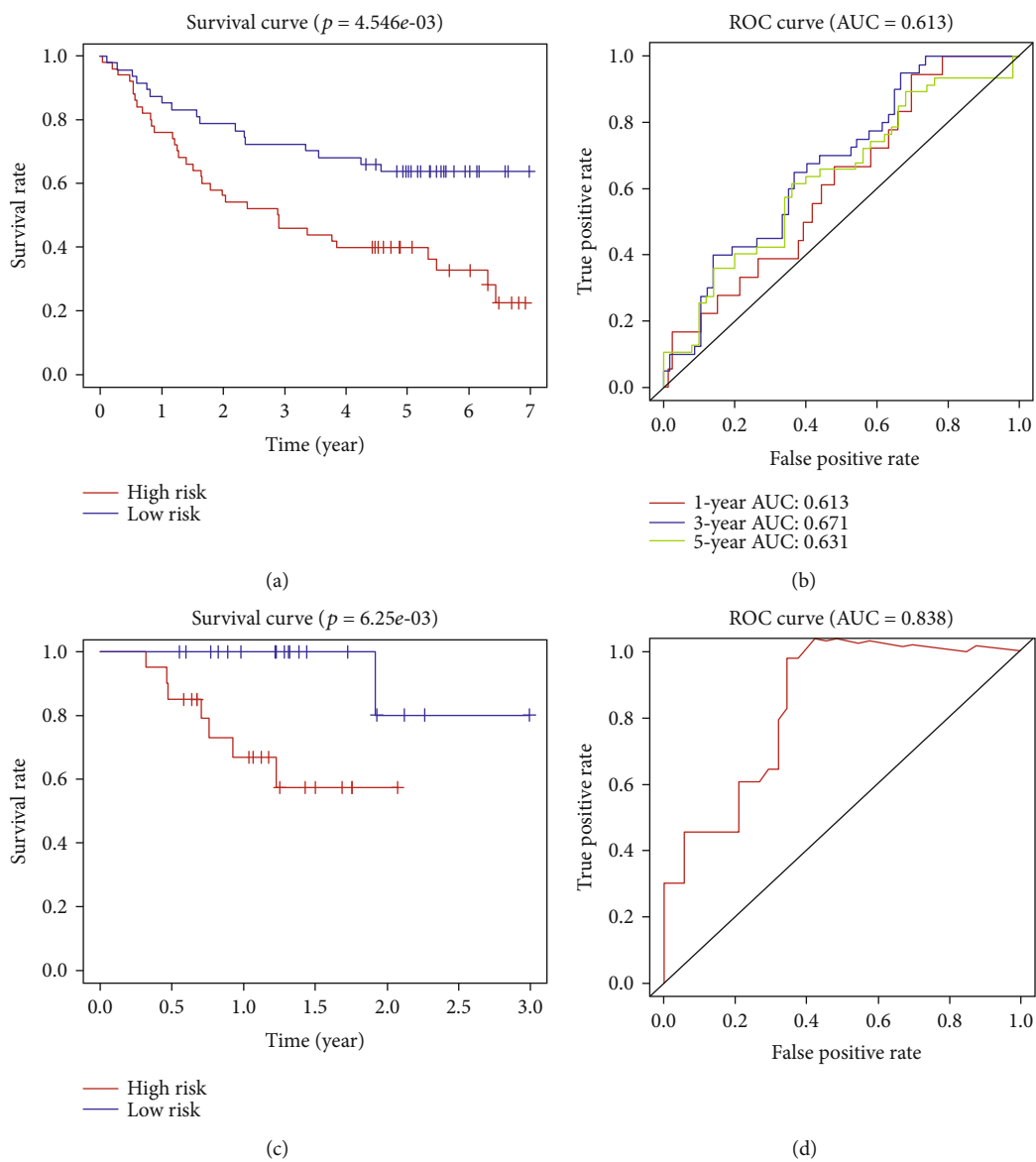
(a)

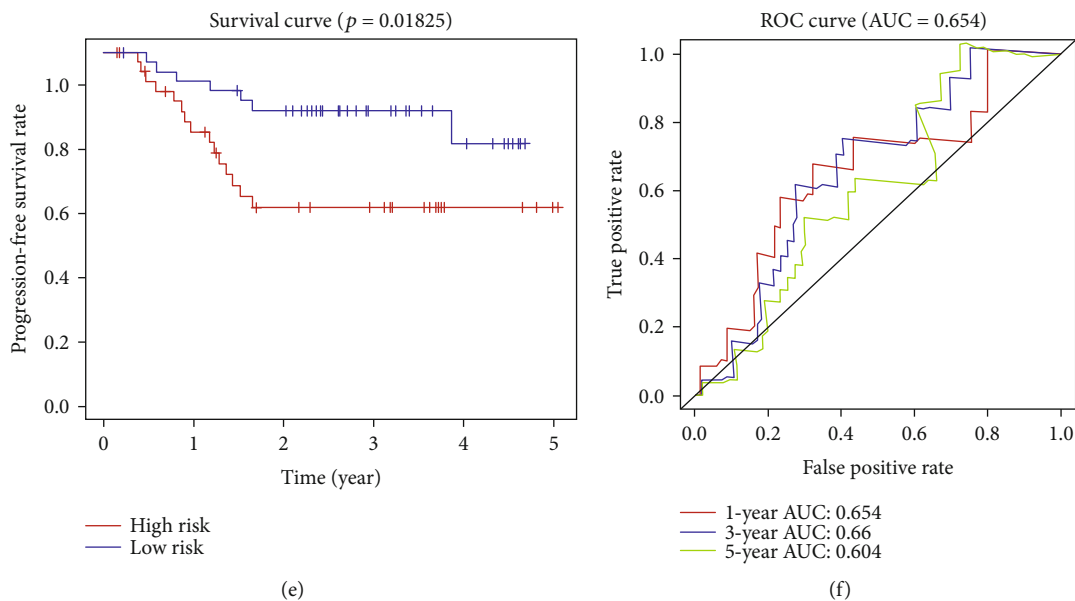(b)



(c)

(d)

Figure 13: Continued.

FIGURE 13: The Kaplan–Meier curves of the overall survival (OS) for high-risk and low-risk patient cohorts divided by the 6-ARG signature in the GEO dataset GSE41613 (a) and ICGC dataset ORCA-IN (c). (e) The Kaplan–Meier curves of the progression-free survival (PFS) for high-risk and low-risk patient cohorts divided by the 6-ARG signature in GSE117973. Receiver operating characteristic analysis in the GEO datasets GSE41613 (b) and GSE117973 (f) and the ICGC dataset ORCA-IN (d).

for autophagy. Although there have been no reports concerning HNSCC and *HSPB8*, this gene plays an important role in some diseases. It is reported that mutation of *HSPB8* is linked to neurodegenerative disorders [48]. In addition, the decreased expression of HSPB8 has been shown to lead to an increased number of cells resting in the G0/G1 phase and reduce the migratory ability of MCF-7 cells and thus is engaged in regulating cell cycle and cell migration in MCF-7 cells of breast cancer [49]. Hence, the relationship between these prognostic ARGs and HNSCC needs further research in the future.

However, there are some limitations in this study. First, this is a retrospective study, which emphasizes on data mining and data analysis. Further experiments are needed to validate these findings. Second, there is currently no other research that investigates autophagy-related gene signature for HNSCC, suggesting that we are unable to validate our findings in another independent study. Therefore, we encourage multicenter data to confirm our results.

## 5. Conclusions

In summary, we identify a six-autophagy-related gene signature that can independently predict prognosis of HNSCC and help to distinguish high-risk patients. The six identified genes provide new insights into underlying molecular mechanisms of HNSCC and can be utilized as promising therapeutic targets. Further studies are expected to verify the clinical application and explore optimal treatment strategies for HNSCC.

## Data Availability

The data used to support the findings of this study are available from TCGA database (https://cancergenome.nih.gov/),

the Gene Expression Omnibus database (https://www.ncbi .nlm.nih.gov/gds/), International Cancer Genome Consortium (https://dcc.icgc.org/releases/current/Projects/ORCA-IN), and the Human Autophagy Database (HADb; http:// autophagy.lu/clustering/index.html).

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Acknowledgments

## Supplementary Materials

Figure S1: the results of the least absolute shrinkage and selection operator regression. (a) The tuning parameter (lambda) selection in the least absolute shrinkage and selection operator (LASSO) model using 10-fold cross-validation via minimum criteria. Dotted vertical lines are drawn at the optimal values using the minimum criteria and the 1 standard error of the minimum criteria (the 1-SE criteria). (b) LASSO coefficient profiles of the 9-survival-related autophagy-related genes. *(Supplementary Materials)*

## References

[1] Global Burden of Disease Cancer Collaboration, C. Fitzmaurice, D. Abate et al., "Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and

disability-adjusted life-years for 29 cancer groups, 1990 to 2017: a systematic analysis for the global burden of disease study," *JAMA Oncology*, vol. 5, no. 12, pp. 1749–1768, 2019.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA: a Cancer Journal for Clinicians*, vol. 66, no. 1, pp. 7–30, 2016.

[3] H. Polanska, M. Raudenska, J. Gumulec et al., "Clinical significance of head and neck squamous cell cancer biomarkers," *Oral Oncology*, vol. 50, no. 3, pp. 168–177, 2014.

[4] D. Pulte and H. Brenner, "Changes in survival in head and neck cancers in the late 20th and early 21st century: a period analysis," *The Oncologist*, vol. 15, no. 9, pp. 994–1001, 2010.

[5] A. Argiris, M. V. Karamouzis, D. Raben, and R. L. Ferris, "Head and neck cancer," *The Lancet*, vol. 371, no. 9625, pp. 1695–1709, 2008.

[6] B. Levine and G. Kroemer, "Biological functions of autophagy genes: a disease perspective," *Cell*, vol. 176, no. 1-2, pp. 11–42, 2019.

[7] R. Amaravadi, A. C. Kimmelman, and E. White, "Recent insights into the function of autophagy in cancer," *Genes & Development*, vol. 30, no. 17, pp. 1913–1930, 2016.

[8] C. M. Dower, C. A. Wills, S. M. Frisch, and H. G. Wang, "Mechanisms and context underlying the role of autophagy in cancer metastasis," *Autophagy*, vol. 14, no. 7, pp. 1110–1128, 2018.

[9] P. Peixoto, C. Grandvallet, J.-P. Feugeas, M. Guittaut, and E. Hervouet, "Epigenetic control of autophagy in cancer cells: a key process for cancer-related phenotypes," *Cells*, vol. 8, no. 12, p. 1656, 2019.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort et al., "Scikit-learn: machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[11] G. D'Souza and A. Dempsey, "The role of HPV in head and neck cancer and review of the HPV vaccine," *Preventive Medicine*, vol. 53, Supplement 1, pp. S5–S11, 2011.

[12] A. M. Goodman, S. Kato, L. Bazhenova et al., "Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers," *Molecular Cancer Therapeutics*, vol. 16, no. 11, pp. 2598–2608, 2017.

[13] N. Klebanov, M. Artomov, W. B. Goggins, E. Daly, M. J. Daly, and H. Tsao, "Burden of unique and low prevalence somatic mutations correlates with cancer survival," *Scientific Reports*, vol. 9, no. 1, p. 4848, 2019.

[14] N. Stransky, A. M. Egloff, A. D. Tward et al., "The mutational landscape of head and neck squamous cell carcinoma," *Science*, vol. 333, no. 6046, pp. 1157–1160, 2011.

[15] G. Zhou, Z. Liu, and J. N. Myers, "TP53 mutations in head and neck squamous cell carcinoma and their impact on disease progression and treatment response," *Journal of Cellular Biochemistry*, vol. 117, no. 12, pp. 2682–2692, 2016.

[16] S. Miyauchi, S. S. Kim, J. Pang et al., "Immune modulation of head and neck squamous cell carcinoma and the tumor microenvironment by conventional therapeutics," *Clinical Cancer Research*, vol. 25, no. 14, pp. 4211–4223, 2019.

[17] L. Galluzzi, F. Pietrocola, J. M. Bravo-San Pedro et al., "Autophagy in malignant transformation and cancer progression," *The EMBO Journal*, vol. 34, no. 7, pp. 856–880, 2015.

[18] Y.-Q. Tan, J. Zhang, and G. Zhou, "Autophagy and its implication in human oral diseases," *Autophagy*, vol. 13, no. 2, pp. 225–236, 2017.

[19] K. K. Ang, J. Harris, R. Wheeler et al., "Human papillomavirus and survival of patients with oropharyngeal cancer," *The New England Journal of Medicine*, vol. 363, no. 1, pp. 24–35, 2010.

[20] L. A. Koneva, Y. Zhang, S. Virani et al., "HPV integration in HNSCC correlates with survival outcomes, immune response signatures, and candidate drivers," *Molecular Cancer Research*, vol. 16, no. 1, pp. 90–102, 2018.

[21] A. R. Kreimer, G. M. Clifford, P. Boyle, and S. Franceschi, "Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review," *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, vol. 14, no. 2, pp. 467–475, 2005.

[22] A. von Witzleben, C. Wang, S. Laban, N. Savelyeva, and C. H. Ottensmeier, "HNSCC: tumour antigens and their targeting by immunotherapy," *Cells*, vol. 9, no. 9, p. 2103, 2020.

[23] B. Bai, X. Wang, E. Chen, and H. Zhu, "Human cytomegalovirus infection and colorectal cancer risk: a meta-analysis," *Oncotarget*, vol. 7, no. 47, pp. 76735–76742, 2016.

[24] W. H. Teo, H.-P. Chen, J. C. Huang, and Y. J. Chan, "Human cytomegalovirus infection enhances cell proliferation, migration and upregulation of EMT markers in colorectal cancer-derived stem cell-like cells," *International Journal of Oncology*, vol. 51, no. 5, pp. 1415–1426, 2017.

[25] E. Koustas, P. Sarantis, G. Kyriakopoulou, A. G. Papavassiliou, and M. V. Karamouzis, "The interplay of autophagy and tumor microenvironment in colorectal cancer—ways of enhancing immunotherapy action," *Cancers*, vol. 11, no. 4, p. 533, 2019.

[26] P. Bossi, C. Resteghini, N. Paielli, L. Licitra, S. Pilotti, and F. Perrone, "Prognostic and predictive value of EGFR in head and neck squamous cell carcinoma," *Oncotarget*, vol. 7, no. 45, pp. 74362–74379, 2016.

[27] N. Normanno, A. de Luca, C. Bianco et al., "Epidermal growth factor receptor (EGFR) signaling in cancer," *Gene*, vol. 366, no. 1, pp. 2–16, 2006.

[28] S. Wheeler, D. R. Siwak, R. Chai et al., "Tumor epidermal growth factor receptor and EGFR PY1068 are independent prognostic indicators for head and neck squamous cell carcinoma," *Clinical Cancer Research*, vol. 18, no. 8, pp. 2278–2289, 2012.

[29] A. A. Molinolo, P. Amornphimoltham, C. H. Squarize, R. M. Castilho, V. Patel, and J. S. Gutkind, "Dysregulated molecular networks in head and neck carcinogenesis," *Oral Oncology*, vol. 45, no. 4-5, pp. 324–334, 2009.

[30] C. Boeckx, M. Baay, A. Wouters et al., "Anti-epidermal growth factor receptor therapy in head and neck squamous cell carcinoma: focus on potential molecular mechanisms of drug resistance," *The Oncologist*, vol. 18, no. 7, pp. 850–864, 2013.

[31] A. G. Sacco and F. P. Worden, "Molecularly targeted therapy for the treatment of head and neck cancer: a review of the ErbB family inhibitors," *Oncotargets and Therapy*, vol. 9, pp. 1927–1943, 2016.

[32] E. E. W. Cohen, D. Soulières, C. le Tourneau et al., "Pembrolizumab versus methotrexate, docetaxel, or cetuximab for recurrent or metastatic head-and-neck squamous cell carcinoma (KEYNOTE-040): a randomised, open-label, phase 3 study," *The Lancet*, vol. 393, no. 10167, pp. 156–167, 2019.

[33] The Cancer Genome Atlas Network, "Comprehensive genomic characterization of head and neck squamous cell carcinomas," *Nature*, vol. 517, no. 7536, pp. 576–582, 2015.

[34] S. M. Rothenberg and L. W. Ellisen, "The molecular pathogenesis of head and neck squamous cell carcinoma," *The Journal of Clinical Investigation*, vol. 122, no. 6, pp. 1951–1957, 2012.

[35] C. R. Salazar, N. Anayannis, R. V. Smith et al., "Combined P16 and human papillomavirus testing predicts head and neck cancer survival," *International Journal of Cancer*, vol. 135, no. 10, pp. 2404–2412, 2014.

[36] X. Jiang, J. Ye, Z. Dong, S. Hu, and M. Xiao, "Novel genetic alterations and their impact on target therapy response in head and neck squamous cell carcinoma," *Cancer Management and Research*, vol. 11, pp. 1321–1336, 2019.

[37] L. Tourneur and G. Chiocchia, "FADD: a regulator of life and death," *Trends in Immunology*, vol. 31, no. 7, pp. 260–269, 2010.

[38] R. B. Reddy, A. R. Bhat, B. L. James et al., "Meta-analyses of microarray datasets identifies ANO1 and FADD as prognostic markers of head and neck cancer," *PLoS One*, vol. 11, no. 1, article e0147409, 2016.

[39] W. J. Pattje, L. J. Melchers, L. Slagter-Menkema et al., "FADD expression is associated with regional and distant metastasis in squamous cell carcinoma of the head and neck," *Histopathology*, vol. 63, no. 2, pp. 263–270, 2013.

[40] D. F. Eytan, G. E. Snow, S. Carlson et al., "SMAC mimetic Birinapant plus radiation eradicates human head and neck cancers with genomic amplifications of cell death genes FADD and BIRC2," *Cancer Research*, vol. 76, no. 18, pp. 5442–5454, 2016.

[41] P. Pinon and B. Wehrle-Haller, "Integrins: versatile receptors controlling melanocyte adhesion, migration and proliferation," *Pigment Cell & Melanoma Research*, vol. 24, no. 2, pp. 282–294, 2011.

[42] K. Koshizuka, T. Hanazawa, N. Kikkawa et al., "Regulation ofITGA3by the anti-tumormiR-199family inhibits cancer cell migration and invasion in head and neck cancer," *Cancer Science*, vol. 108, no. 8, pp. 1681–1692, 2017.

[43] K. Koshizuka, N. Nohata, T. Hanazawa et al., "Deep sequencing-based microRNA expression signatures in head and neck squamous cell carcinoma: dual strands of pre-miR-150 as antitumor miRNAs," *Oncotarget*, vol. 8, no. 18, pp. 30288–30304, 2017.

[44] J. P. Bernardini, M. Lazarou, and G. Dewson, "Parkin and mitophagy in cancer," *Oncogene*, vol. 36, no. 10, pp. 1315–1327, 2017.

[45] L. Xu, D. C. Lin, D. Yin, and H. P. Koeffler, "An emerging role of PARK2 in cancer," *Journal of Molecular Medicine (Berlin, Germany)*, vol. 92, no. 1, pp. 31–42, 2014.

[46] G. Bebek, M. Orloff, and C. Eng, "Microenvironmental genomic alterations reveal signaling networks for head and neck squamous cell carcinoma," *Journal of Clinical Bioinformatics*, vol. 1, no. 1, p. 21, 2011.

[47] G. Gao, J. L. Kasperbauer, N. M. Tombers, V. Wang, K. Mayer, and D. I. Smith, "A selected group of large common fragile site genes have decreased expression in oropharyngeal squamous cell carcinomas," *Genes, Chromosomes & Cancer*, vol. 53, no. 5, pp. 392–401, 2014.

[48] F. Li, H. Xiao, Z. Hu, F. Zhou, and B. Yang, "Exploring the multifaceted roles of heat shock protein B8 (HSPB8) in diseases," *European Journal of Cell Biology*, vol. 97, no. 3, pp. 216–229, 2018.

[49] M. Piccolella, V. Crippa, R. Cristofani et al., "The small heat shock protein B8 (HSPB8) modulates proliferation and migration of breast cancer cells," *Oncotarget*, vol. 8, no. 6, pp. 10400–10415, 2017.