

Development of a five-protein signature for predicting the prognosis of head and neck squamous cell carcinoma

Xinyuan Zhao¹, Xianwen Liu¹, Li Cui^{1,2}

¹Stomatological Hospital, Southern Medical University, Guangzhou 510280, China

²UCLA School of Dentistry, Los Angeles, CA 90095, USA

Correspondence to: Xinyuan Zhao, Li Cui; **email:** zhaoxinyuan1989@smu.edu.cn, zsucllj@ucla.edu

Keywords: head and neck squamous cell carcinoma, prognostic signature, survival analysis, HER3_pY1289, the cancer proteome atlas

Received: February 15, 2020

Accepted: August 19, 2020

Published: October 13, 2020

Copyright: © 2020 Zhao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Currently no reliable indicators are available for predicting the clinical outcome of head and neck squamous cell carcinoma (HNSCC). This study aimed to develop a protein-based model to improve the prognosis prediction of HNSCC. The proteome data of HNSCC cohort was downloaded from The Cancer Proteome Atlas (TCPA) portal. The TCPA HNSCC cohort was randomly divided into the discovery and validation cohort. A protein-based risk signature was developed with the discovery cohort, and then verified with the validation cohort. The prognostic value of HER3_pY1289 was further determined. We have constructed a five-protein risk signature which was strongly associated with the overall survival (OS) in the discovery cohort. Similar findings were observed in the validation cohort. The protein-based risk signature was identified as an independent prognostic factor for HNSCC. A nomogram model built on the protein-based risk signature exhibited good performance for predicting OS. Our immunohistochemistry (IHC) analysis showed that higher HER3_pY1289 staining intensity was closely associated with unfavorable prognosis of HNSCC. HER3 suppression inhibited the proliferation and invasion capacity of HNSCC cells. Collectively, we have developed a protein-based risk signature for accurately predicting the prognosis of HNSCC, which might provide valuable information for optimal individualized treatment regimens.

INTRODUCTION

Head and neck cancer (HNC) represents the sixth most common cancer worldwide, and squamous cell carcinoma (SCC) consist of more than 90% of HNC [1]. Accumulation of genetic alterations, exposure to environmental risk factors, viral infection and unhealthy lifestyles contribute to the initiation and progression of head and neck squamous cell carcinoma (HNSCC) [2]. Although great progress has been achieved for the therapeutic methodologies, the clinical outcome of HNSCC remains unsatisfactory in the past few decades [3]. Stratifying the HNSCC patients with different prognosis is of great importance for selecting the optimal therapeutic strategies for individual patient. Unfortunately, currently no reliable indicator is

available for accurately predicting the prognosis of HNSCC. Therefore, development of robust models for distinguishing the HNSCC patients with different risks is urgently needed.

The recent development of next-generation sequencing technology has significantly enriched and expanded our understanding on the association between DNA (genomic level) or RNA (transcript level) and cancer progression. Many DNA or RNA based prognostic signatures have also been developed for predicting the prognosis of human cancer including HNSCC [4–6]. However, proteins are the basic functional units for executing the biological processes. In addition, most cancer therapy targeted the proteins instead of DNA and RNA. Moreover, poor correlation might be observed between

DNA and RNA levels with protein levels [7]. Therefore, it is important for measuring the protein levels directly and explore their associations with the clinical outcome of HNSCC. Reverse-phase protein arrays (RPPAs) is a powerful proteomic approach for assessing the levels of interested proteins across different samples in a high-throughput manner [8, 9]. The Cancer Proteome Atlas (TCPA) is a public accessible platform which spans more than 8,000 patient samples through The Cancer Genome Atlas (TCGA). In addition, approximately 300 protein markers have been examined in the tumor samples [10, 11]. Therefore, TCPA is highly valuable resource for developing protein-based risk signature for predicting the clinical outcome of HNSCC.

In this study, the TCPA HNSCC cohort was randomly divided into discovery cohort and validation cohort. A five-protein risk signature was developed with the discovery cohort, and then successfully verified in the validation cohort. In addition, a nomogram model based on the protein-based risk signature was constructed and it demonstrated impressive predictive accuracy for overall survival (OS) of HNSCC. Our immunohistochemistry (IHC) analysis showed that the staining intensity of HER3_pY1289, a protein in the risk signature, was positively associated with unfavorable clinical outcome of HNSCC. The expression level of HER3_pY1289 was markedly reduced following downregulation of HER3, and knockdown of HER3 inhibited the oncogenic activities of HNSCC cells.

RESULTS

Candidate OS-related proteins of HNSCC patients in the TCPA discovery cohort

A total of 33 proteins that significantly associated with OS are identified by the univariate cox proportional hazards regression analysis. The proteins with the hazard ratio (HR) larger than 1 are risky proteins, while those with the HR less than 1 are protective proteins. As shown in Figure 1A and Supplementary Table 1, 13 risky (dark red) and 20 protective (green) proteins are identified. Then the above 33 OS-associated proteins are subjected to least absolute shrinkage and selection operator (LASSO) analysis, which is a regression analysis methodology that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model. It is widely used for optimal selection of features in high-dimensional data with a strong predictive value and low correlation between one another to prevent overfitting. Therefore, it is highly efficient for identifying the most useful predictive markers and generating a prognostic signature for predicting clinical outcome. The dashed vertical line indicates the optimal

value of $\log \lambda$ with the minimum partial likelihood deviance, and therefore 6 proteins (CyclinD1, E-cadherin, HER3_pY1289, PAI-1, XRCC1 and Raptor) are selected for the subsequent multivariate analysis (Figure 1B). The LASSO coefficient of the 6 proteins is shown in Figure 1C. Then multivariate cox proportional hazards regression analysis is performed to identify the independent prognostic proteins. The results show that CyclinD1, HER3_pY1289, PAI-1, XRCC1 and Raptor are the prognostic proteins independently associated with OS in the discovery cohort (Figure 1D). Subsequently, the independent prognostic proteins are used to build up a risk score model. A risk score for each patient is calculated as the sum of each protein's score, which is obtained by multiplying the expression level of the protein and its corresponding coefficient. The coefficients of the proteins are obtained from the multivariate analysis. The following formula is developed to calculate risk score for each patient: risk score = $(0.960 \times \text{CyclinD1}) + (1.673 \times \text{HER3_pY1289}) + (0.330 \times \text{PAI-1}) + (-1.169 \times \text{XRCC1}) + (1.681 \times \text{Raptor})$.

Construction of a protein-based prognostic model using the TCPA discovery cohort

The discovery cohort is divided into high risk group (n=86) and low risk group (n=87) using the median value of the risk scores. The distribution of risk scores in the high and low-risk groups is depicted in Figure 2A. Figure 2B shows the patterns of survival time and survival status between high and low-risk groups. The relative expression levels of the five prognostic proteins for each patient are shown in Figure 2C. The survival analysis demonstrates that the patients in the high-risk group have a significant shorter OS than the patients in the low-risk group ($p=2.254e-07$) (Figure 2D). The differences in OS stratified by the common clinicopathological parameters are further analyzed between high and low-risk groups. One hundred and forty-nine patients remain in the discovery cohort after removing the cases with missing information in age, gender, tumor grade or TNM stage. The clinical information of the discovery cohort is summarized in Supplementary Table 2. As shown in Figure 3A–3D, high-risk group has consistently worse OS than low-risk group for the subgroups stratified by age, gender, tumor grade or TNM stage.

Validation of the protein-based prognostic model using the TCPA validation cohort

Similarly, the validation cohort is divided into high-risk group (n=93) and low-risk group (n=79) using the median value of the risk scores generated in the discovery cohort. The distribution of risk scores as well as the distribution of survival time and survival status

between high and low-risk groups are shown in Figure 4A and Figure 4B, respectively. The expression patterns of the five prognostic proteins in the validation cohort are revealed in Figure 4C. The high-risk group patients have poorer OS than the low-risk group patients ($p=2.747e-02$) (Figure 4D). One hundred and fifty-three patients remain in the validation cohort after removing the cases with missing information in age, gender, tumor grade or TNM stage. The clinical information of the validation cohort is described in Supplementary Table 3. For the cases with age<60, or male cases, or patients with G1-G2 or those at stage III-IV, the high-risk group patients suffer worse OS compared to the low-risk group. No significant difference in OS is found between high and low-risk groups for the patients with age>=60, or female cases, or patients with G3-G4, or those at stage I-II (Figure 5A–5D).

The protein-based risk signature is an independent prognostic factor for HNSCC

The clinicopathological parameters including age, gender, tumor grade, and TNM stage as well as the protein-based risk signature are subjected to the univariate and multivariate analyses. As shown in Figure 6A–6D, the univariate analysis reveals that the risk signature is significantly associated with the OS of HNSCC both in the discovery cohort ($p<0.001$, HR=1.320, 95% CI=1.187-1.467) and validation cohort ($p<0.001$, HR=1.152, 95% CI=1.062-1.249). The multivariate analysis shows that the risk signature is an independent prognostic factor both in the discovery cohort ($p<0.001$, HR=1.366, 95% CI=1.218-1.532) and validation cohort ($p=0.003$, HR=1.136, 95% CI=1.044-1.235).

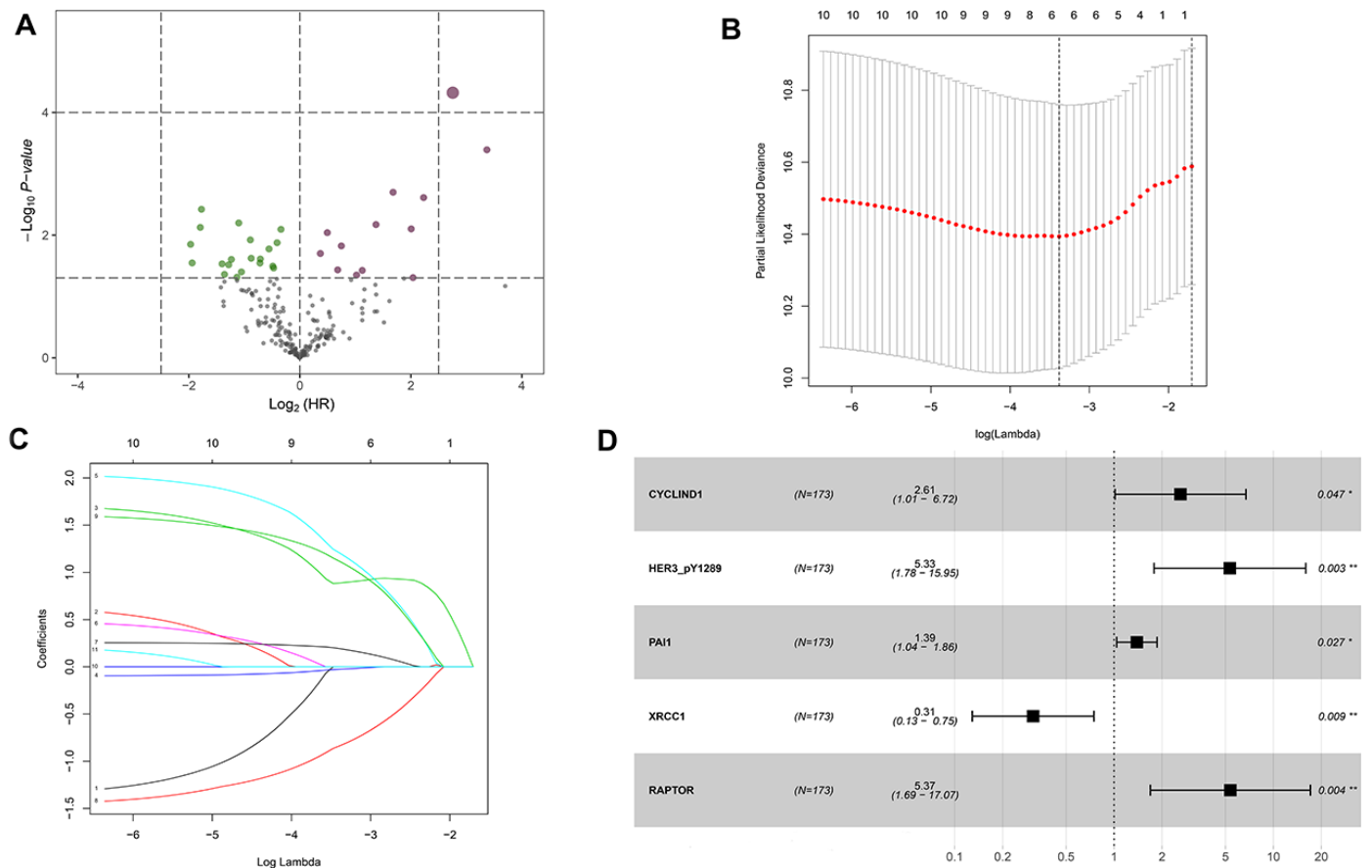


Figure 1. Identification of the overall survival (OS)-associated proteins in the TCPA discovery cohort. (A) Volcano plot of proteins that are significantly associated with OS of HNSCC. Y-axis indicates the p values ($-\log_{10}$ scale), whereas the X-axis shows the hazard ratio (\log_2 scale). Each symbol represents a different protein, and the dark red and green symbols categorize the risky ($n=13$) and protective ($n=20$) proteins, respectively. **(B)** Tuning parameter ($\log\lambda$) selection cross-validation error curve for OS-associated proteins. The vertical dotted line is drawn at the optimal value by the minimum criteria and the 1-SE criteria. **(C)** The LASSO coefficient profile of 6 OS-related proteins and the vertical dotted line is drawn at the value chosen by 10-fold cross-validation. **(D)** Multivariate Cox proportional hazards regression analysis reveals five independent prognostic proteins of HNSCC patients in the discovery cohort.

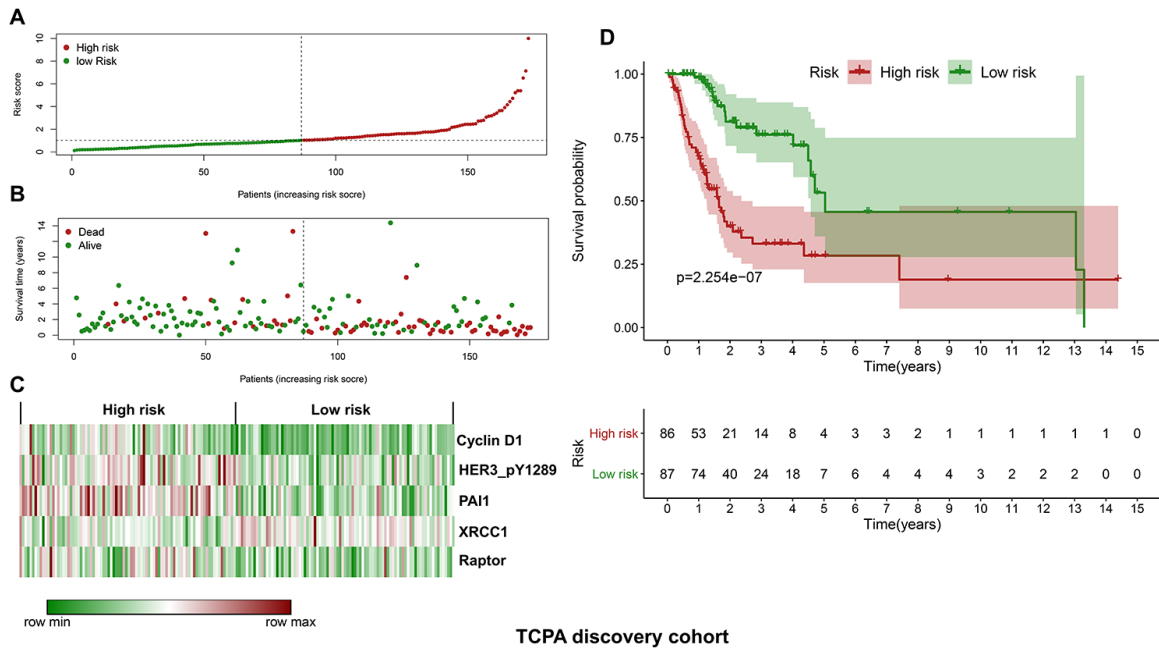


Figure 2. Construction of a protein-based prognostic signature based on the TCPA discovery cohort. (A) The distribution of risk scores in the low and high-risk groups. The risk scores for all patients in discovery cohort are plotted in ascending order and are divided by the threshold (vertical dotted line). The dots in the left (green) and right (dark red) side of the vertical dotted line belong to the low and high-risk groups, respectively. The risk scores are gradually increased from the low-risk group to high-risk group. (B) The pattern of survival time and survival status in low and high-risk groups. The dots in the left and right side of the vertical dotted line belong to the low and high-risk group, respectively. The dark red and green dots indicate death and survival, respectively. The high-risk group has a significantly higher mortality rate than the low-risk group. (C) The expression levels of the five prognostic proteins for each patient in the discovery cohort, with dark red indicating higher expression and green representing lower expression. (D) Survival analysis demonstrates that the patients in the high-risk group have statistically significant worse OS than those in low-risk group.

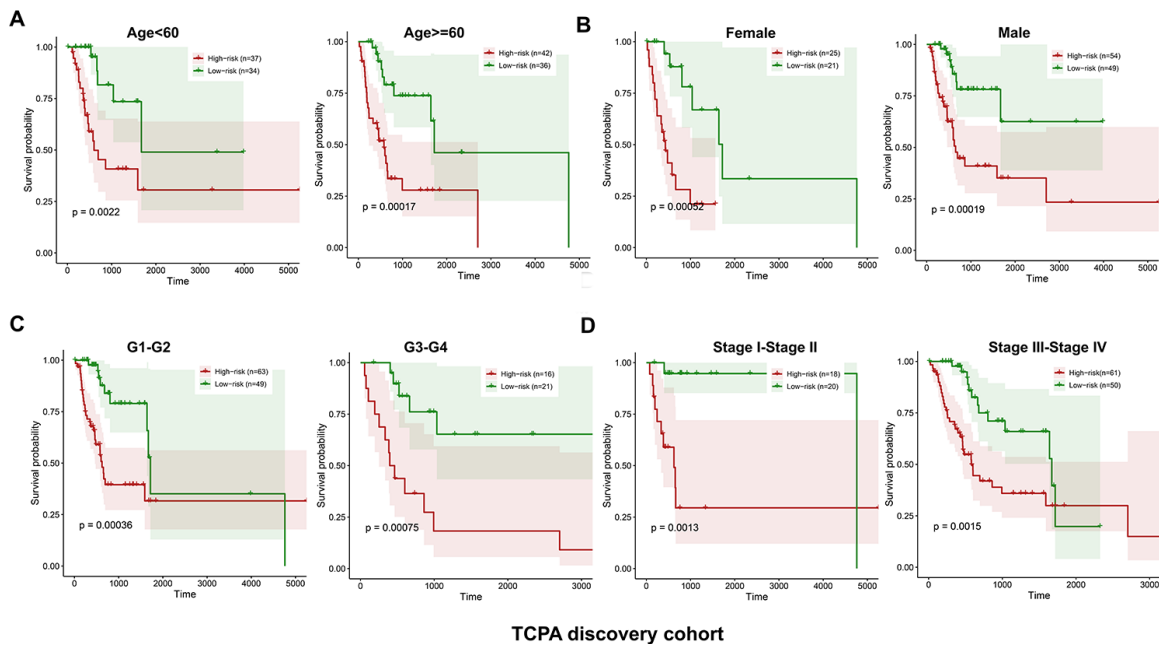


Figure 3. Kaplan-Meier curves of OS differences stratified by age, gender, tumor grade or TNM stage between low and high-risk group in the TCPA discovery cohort. The patients with low-risk scores have significantly better OS than patients with high-risk scores in different subgroups stratified by age (A), gender (B), tumor grade (C) and TNM stage (D).

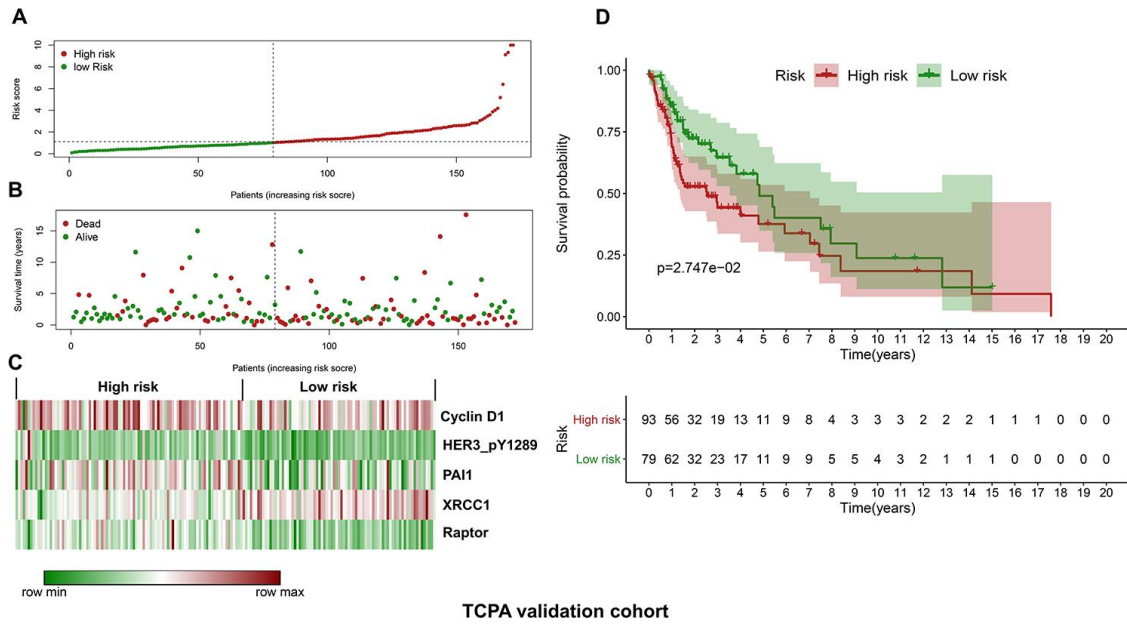


Figure 4. Validation of the protein-based prognostic signature with the TPCA validation cohort. (A) The distribution of risk scores in the low and high-risk groups. The risk scores for all patients in the validation cohort are plotted in ascending order and are divided by the threshold (vertical dotted line). The green and dark red dots belong to the low and high-risk groups, respectively. The risk scores are gradually increased from the low-risk group to high-risk group. (B) The pattern of survival time and survival status in low and high-risk groups. The dots in the left and right side of the vertical dotted line indicate the patients in the low and high-risk group, respectively. The dark red and green dots represent death and survival, respectively. The mortality rate is markedly higher in the high-risk group than in the low-risk group. (C) The expression levels of the five prognostic proteins for each patient in the validation cohort. The dark red color is indicative of higher relative expression and the green color represents lower expression (D) Survival analysis of the association between risk score and OS of HNSCC in the validation cohort. The OS is significantly shorter in patients in the high-risk group than those in the low-risk group.

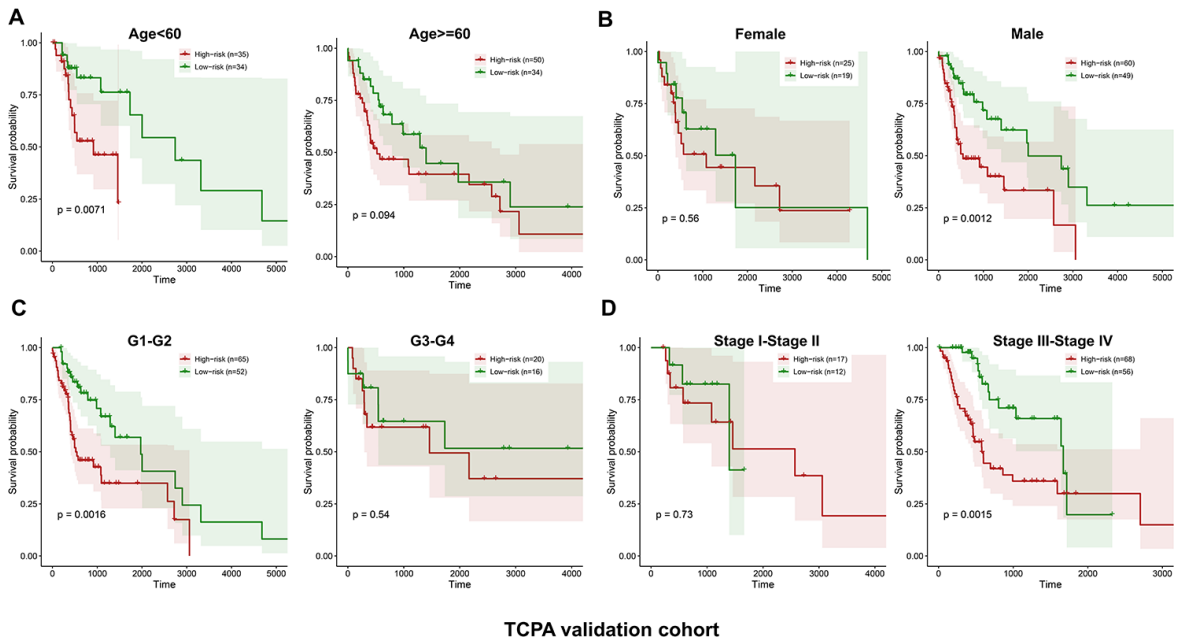


Figure 5. Stratified analysis of the five-proteins signature for HNSCC patients in the TPCA validation cohort with age, gender, tumor grade or TNM stage. The patients with high-risk scores have significantly worse OS than those with low-risk scores in the subgroups of age<60, male, grade 1-2 and stage III-IV. No significant difference in OS is found for the patients in the subgroups of age>=60, female, grade 3-4 and stage I-II (A–D).

Nomogram model construction and prediction

To facilitate the potential clinical application, a more sensitive nomogram predictive model is developed. As shown in Figure 7A, the risk signature, age, gender, tumor grade and TNM stage are included into the nomogram model to predict the prognosis of HNSCC. A nomogram-based score is calculated for each patient based on their risk scores and the clinicopathological parameters on the point scale. The calibration curves show that the nomogram model exhibits excellent performance for predicting the 1-year OS and 3-year OS of HNSCC (Figure 7B–7C).

HER3_pY1289 is overexpressed in HNSCC and its potential prognostic value

Immunohistochemistry (IHC) is performed to evaluate the expression level of HER3_pY1289 in HNSCC samples. The IHC score is calculated for each sample in the SMUSH cohort. The median value of IHC score is used as the cut-off value to divide the SMUSH cohort into high IHC score group (n=54) and low IHC score group (n=58). Figure 8A shows the representative samples in the high and low IHC score group, and the

staining intensity is significantly higher in high IHC score group compared to the low IHC score group. A higher percentage of HNSCC patients at the advanced stages (32/54 vs 18/58) or with positive lymph node metastasis (40/54 vs 25/58) or with human papillomavirus (HPV) infection (18/48 vs 10/46, 94 out of 112 HNSCC patients have the results of the HPV test) is found in the high IHC score group compared to the low IHC score group (Figure 8B–8D). The survival analysis reveals that the patients in the high IHC score group have markedly shorter OS than those in the low IHC score group ($p=4.774e-03$) (Figure 8E).

Knockdown of HER3 suppresses the proliferation and invasion of HNSCC cells

The HER3 mRNA is markedly reduced in HNSCC cells following siHER3 transfections (Figure 9A). The western blotting results show that the expression levels of HER3 and HER3_pY1289 are significantly lower in siHER3 treated cells compared to the siCTRL treated cells (Figure 9B). The MTT assay reveals that the OD values are lower in HER3 knockdown cells at various time points (48h, 72h and 96h) compared to the control cells (Figure 9C). Similarly, the percentage of EdU positive cells is lower in HNSCC cells with HER3

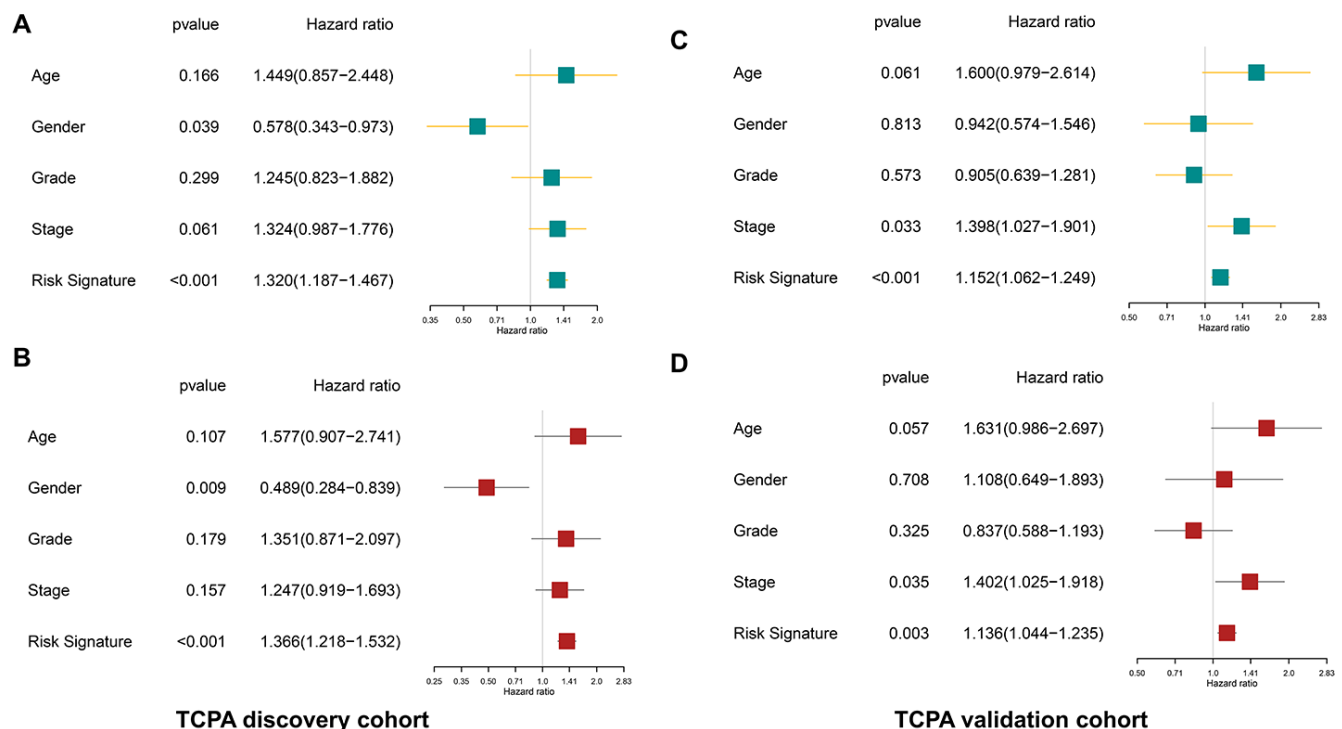


Figure 6. The protein-based risk signature is an independent prognostic factor both in the discovery and validation cohort. (A) The univariate analysis shows that the risk signature is significantly associated with OS of HNSCC in the discovery cohort. (B) The multivariate analysis demonstrates the risk signature is an independent prognostic factor in the discovery cohort. (C, D) Similar findings are observed in the validation cohort.

downregulation (Figure 9D–9E). The number of HER3 knockdown cells that invaded through the membrane is significantly less than that of the control cells (Figure 9F, 9G).

DISCUSSION

To the best of our knowledge, this is the first study to construct a protein-based prognostic signature for HNSCC with the TCGA dataset. A five protein-based risk signature was developed with the discovery cohort, and further validated in the validation cohort, indicating the risk signature was strongly associated with the OS of HNSCC. In addition, the multivariate cox model which included the common clinicopathological variables showed that the protein-based risk signature was an independent risk indicator for HNSCC both in the discovery and validation cohort. Moreover, the risk signature-based nomogram model exhibited good performance for predicting the OS of HNSCC, indicating it might have promising potential for clinical application. Our IHC analysis showed that high HER3_pY1289 staining intensity was positively correlated with aggressive clinical variables and unfavorable OS. Downregulation of HER3 suppressed

the expression of HER3_pY1289 as well as the oncogenic activities of HNSCC cells. These results suggested that HER3_pY1289 played a tumor promoting role in HNSCC carcinogenesis, which further supported the findings that HER3_pY1289 was a risky protein in the risk signature model.

Cyclin D1 plays a critical role in regulating cell cycle progression. Higher expression of Cyclin D1 was strongly correlated with advanced tumor stage and positive lymph node metastasis in HNSCC [12]. A recent systemic review also demonstrated that CCND1 amplification or Cyclin D1 overexpression was significantly associated with HNSCC progression and malignant transformation of potentially malignant disorders [13]. Positive PAI-1 membrane expression was demonstrated to be an independent risk factor for local disease relapse of oral squamous cell carcinoma [14]. In addition, high PAI-1 level was strongly associated with aggressive clinicopathological parameters and unfavorable survival in HNSCC [15]. XRCC1 polymorphisms are closely associated with the risk and survival of HNSCC [16]. In addition, reduced XRCC1 levels has been found in HNC patients compared to normal controls [17], suggesting XRCC1

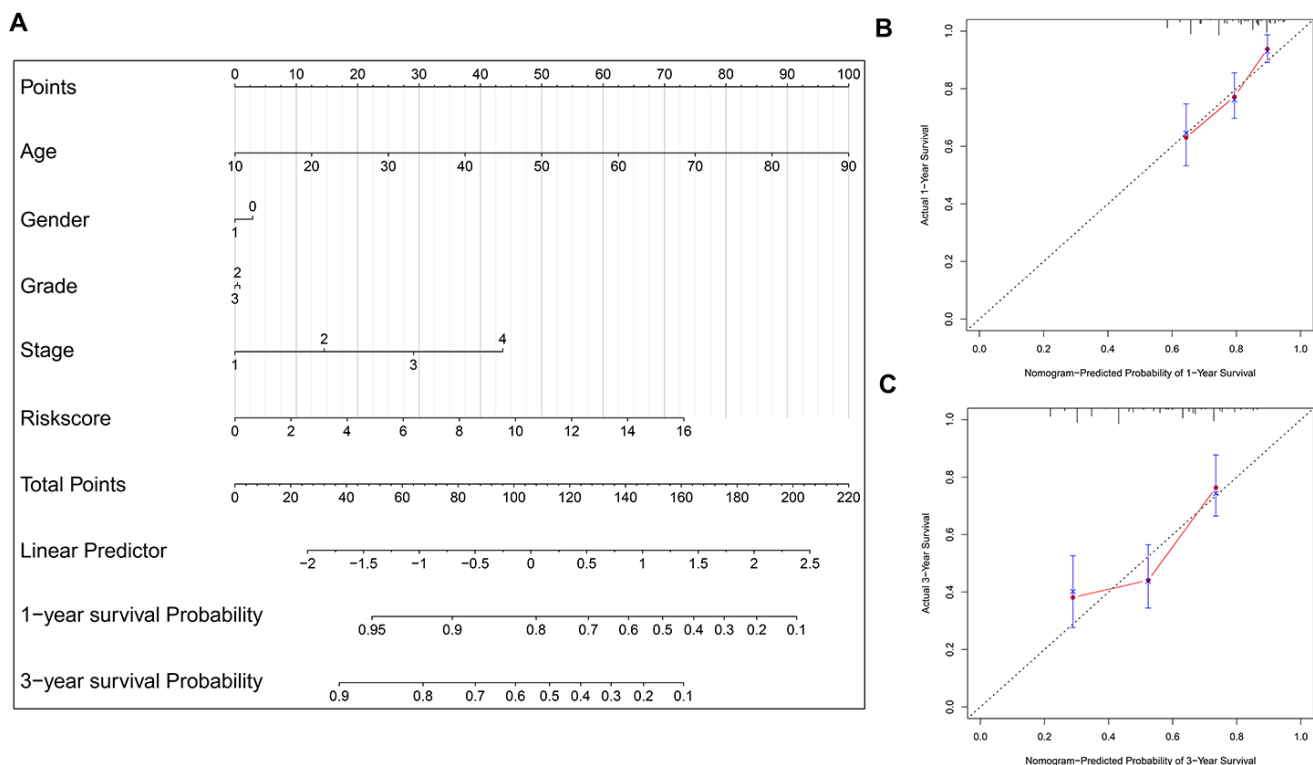


Figure 7. Nomogram model construction and prediction. (A) The nomogram model which includes the protein-based risk signature, age, gender, tumor grade and TNM stage is developed. The total nomogram score for each patient can be easily calculated based on the risk score and the clinicopathological parameters, which is then used to predict the 1-year and 3-year survival probability. (B, C) The calibration plots demonstrate that the nomogram model exhibits good predictability for 1-year OS and 3-year OS.

might play a tumor suppressive role in HNSCC. Mammalian/mechanistic target of rapamycin (mTOR) plays a critical role in tumorigenesis of HNSCC and serves as a molecular target for HNSCC [18, 19]. Raptor acts as scaffold protein and is important for maintaining the activation of mTOR [20]. HER3 is widely recognized as tumor promotor in HNSCC [21, 22]. Brand et al have demonstrated that HER3 was overexpressed in HPV positive HNSCC, and its upregulation was associated with worse overall survival in patients with pharyngeal cancer [21]. HER3 lacks the intrinsic tyrosine kinase activity and it is frequently phosphorylated by other receptor tyrosine kinases (RTKs). Phosphorylation of HER3 is important for activating oncogenic signaling, such as the PI-3K/Akt pathway and Src kinase [23]. Currently, there are at least 9 potential tyrosine phosphorylation sites in the

carboxy-terminal tail of HER3 and HER3_pY1289 (HER3 phosphorylated at the tyrosine residue of 1289) is one of them. Therefore, HER3 upregulation in HNSCC does not definitely indicate that HER3_pY1289 is overexpressed in HNSCC, which was the major reason for choosing HER3_pY1289 for validation. To the best of our knowledge, this is the first study to demonstrate high HER3_pY1289 expression is associated with unfavorable prognosis of HNSCC.

The results of the stratified survival analysis of the validation cohort are not consistent with the discovery cohort in several subgroups of patients, including age \geq 60 years, female, G3-G4 and stage I-II. One possible reason accounting for the inconsistencies of the stratified survival analysis between the discovery and validation cohort is that the sample size in the

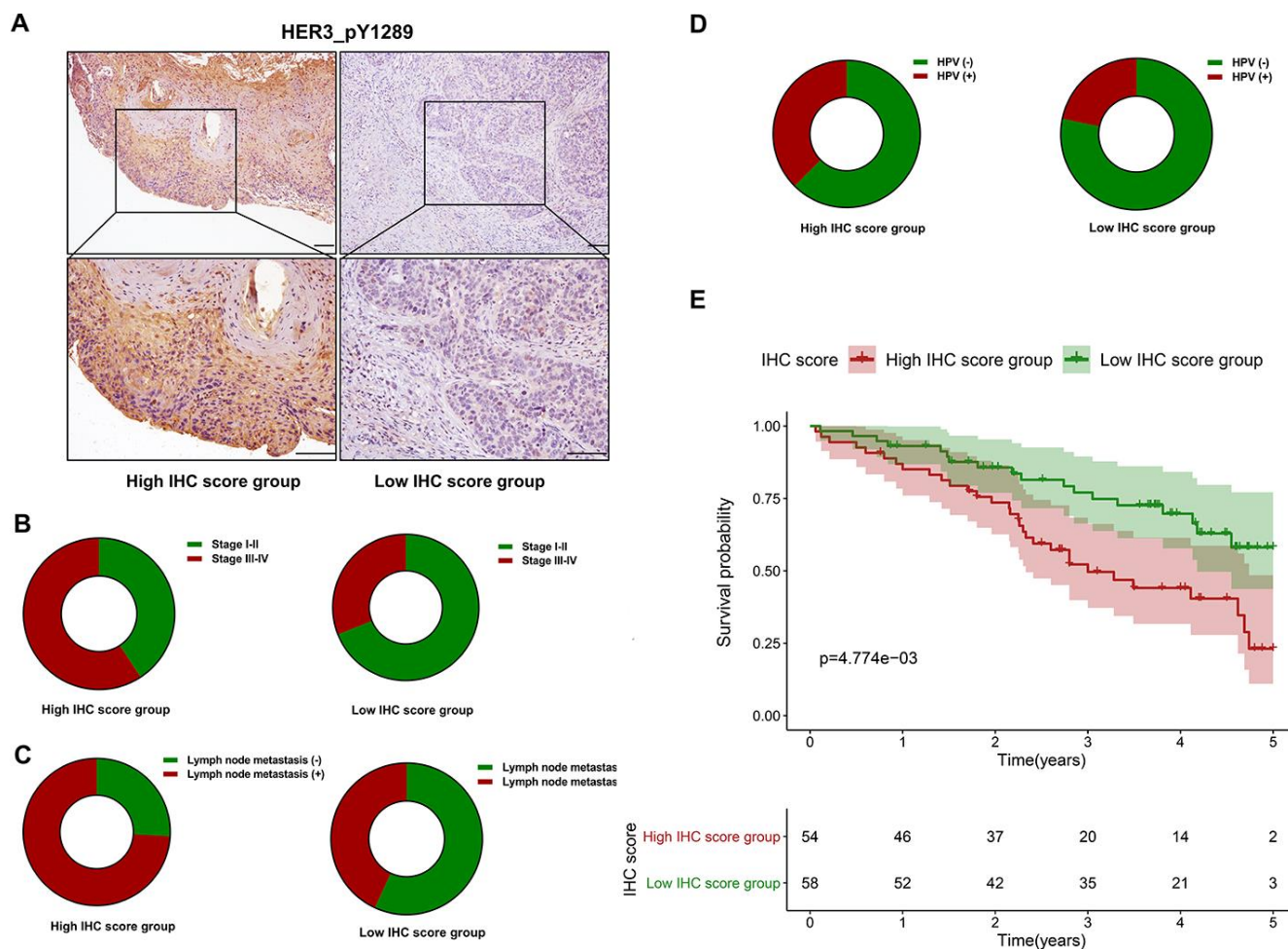


Figure 8. The prognostic value of HER3_pY1289 in HNSCC. (A) The staining intensity of HER3_pY1289 is higher in the high IHC score group compared to the low IHC score group (upper panels: scale bar=100 μ m, lower panels: scale bar=100 μ m). (B) A higher percentage of patients at the advanced stage is observed in the high IHC score group than in the low IHC score group. (C) A higher percentage of patients with positive lymph node metastasis is found in the high IHC score group. (D) A higher percentage of patients with positive HPV infection is found in the high IHC score group. (E) The survival analysis shows that the patients in the high IHC score group have worse OS than those in the low IHC score group.

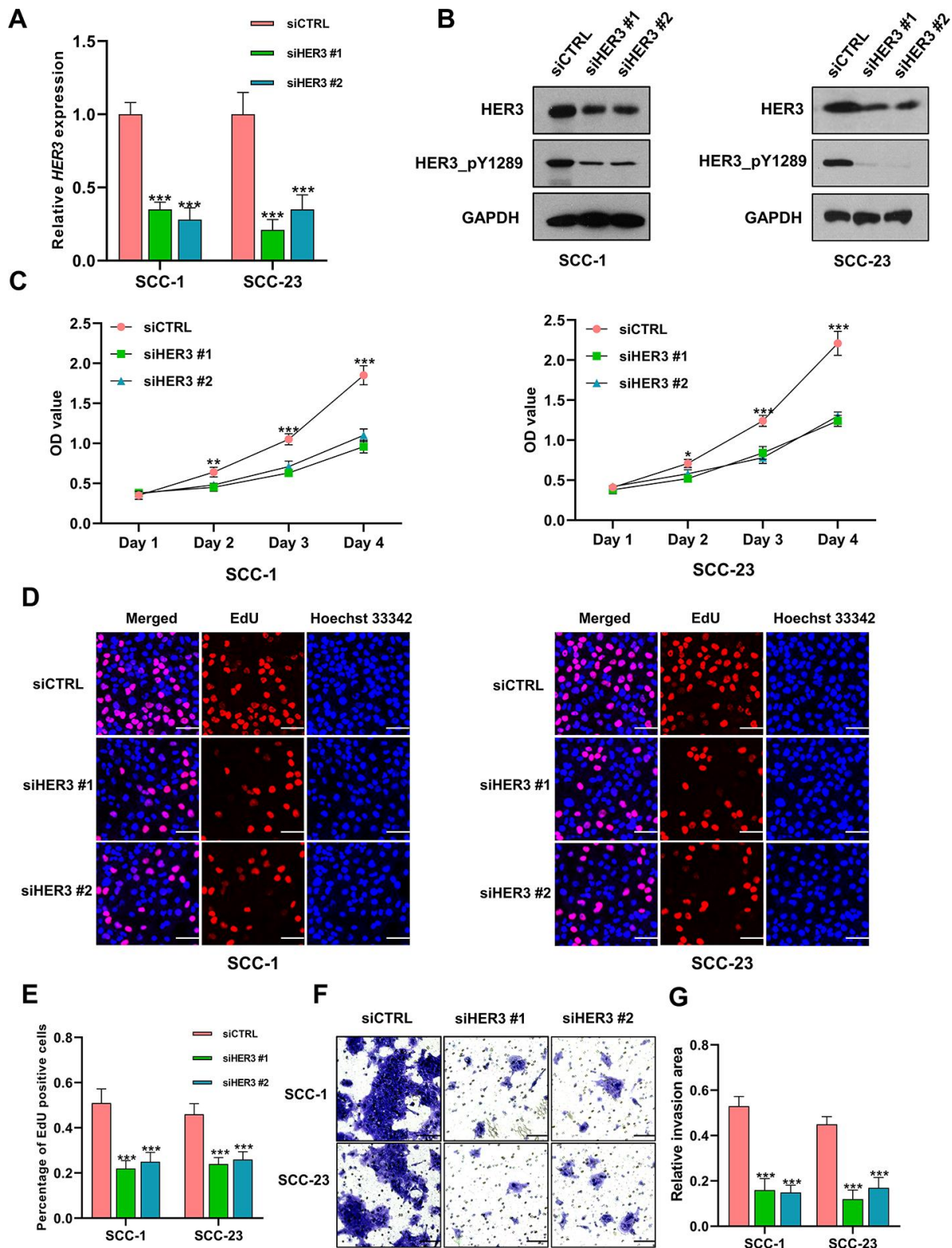


Figure 9. Knockdown of HER3 suppresses the proliferation and invasion of HNSCC cells. (A) The expression level of HER3 mRNA is significantly lower in siHER3 treated cells compared to the siCTRL treated cells. (B) Western blot shows that the levels of HER3 and HER3_pY1289 are lower in siHER3 treated cells compared to the siCTRL treated cells. (C) MTT assay reveals that the OD values are lower in the cells subjected to HER3 downregulation. (D, E) The EdU assay shows that the percentage of EdU positive cells is lower in siHER3 treated cells compared to the siCTRL treated cells (scale bar=20 μ m). (F, G) The invasion assay shows the number of cells invading the membrane is lower in siHER3 treated group than in the control group (scale bar=100 μ m).

subgroups are too small. For instance, for the validation cohort, there are only 44 patients, 36 patients and 29 patients in the female subgroup, G3-G4 subgroup and stage I-II subgroup, respectively. It is very difficult to reach a statistical significance due to the small sample size, and variations in a few samples can greatly affect the results. Therefore, increasing the sample size might reduce the inconsistency. For example, there are 84 patients in the age \geq 60 subgroup of the validation cohort. Although the p value ($p=0.094$) is larger than 0.05, we can still observe the trend that the HNSCC patients in the high-risk group suffered worse OS than those in the low risk group.

Four clinicopathologic parameters (age, gender, tumor grade and TNM stage) are chosen into the univariate and multivariate analyses. There are two reasons for choosing these features. Firstly, age, gender, tumor grade and TNM stage are important clinical variables for HNSCC. Secondly, most HNSCC patients in the TCGA dataset have data for these clinicopathological parameters. If other clinicopathological parameters are selected, most HNSCC cases will be removed. For instance, only 56 out of 346 cases in the TCPA cohort have the results of the HPV test. Therefore, it is unappropriated to include the HPV status into the univariate and multivariate models. In that case, the sample size (only around 20-30 patients in both the discovery and validation cohort) would be very small, which might significantly affect the results of univariate and multivariate analyses. In addition, the results of univariate and multivariate analyses might be not robust and reliable due to the small sample size.

We have also analyzed the associations between the protein-based prognostic signature and the common clinicopathological parameters of HNSCC (age, gender, tumor grade, and TNM stage), and no significant correlation is observed (data not shown). For other clinical features and molecular characteristics, the TCGA HNSCC dataset provides no or little information on these parameters. Further studies with more detailed information are warranted to explore the clinical significance of the protein-based prognostic signature.

There were several limitations for our current study. Firstly, the predictive power of the protein-based risk signature was only evaluated in TCPA validation cohort. Validation of the 5-protein prognostic model in other independent cohorts is very important for confirming its robustness. RPPA is used to determine the expression level of proteins in the HNSCC tissues for the TCPA cohort. RPPA is a high-throughput antibody-based technique with the procedures similar to that of western blots [24]. Thus, the IHC staining results are not suitable for validating the protein-based

prognostic signature which is built on the RPPA methodology. Instead, RPPA should be performed to validate our 5-protein prognostic model in large-scale independent patient cohorts. Secondly, it is important to optimize the risk score formula to minimize the deviations in the tumor samples across different independent studies.

Collectively, we have developed a robust protein-based risk signature for accurately predicting the clinical outcome of HNSCC, which might contribute to the improvement of individualized treatments.

MATERIALS AND METHODS

Public data source

The proteome data of HNSCC cohort (level 4 data) which includes 346 HNSCC samples were downloaded from the TCPA portal (<https://tcpaportal.org/tcpa/>).

The clinical information of HNSCC cases were obtained from The National Cancer Institute Genomic Data Commons (NCI-GDC) (<https://gdc.cancer.gov/>). As the overall survival (OS) time was missing in one case, the remaining 345 HNSCC patients were subjected to the subsequent analysis.

Construction of a protein-based prognostic signature based on the TCPA discovery cohort

The HNSCC dataset was randomly split into the discovery cohort ($n=172$) and validation cohort ($n=173$) using a computer-generated random sequence. The proteins that strongly associated with the OS of HNSCC patients were determined with the univariate cox proportional hazards regression analysis. The LASSO regression analysis was performed to identify the most optimal OS-associated proteins into the multivariate cox proportional hazards regression model. The multivariate analysis was used to determine the prognosis-related proteins and their coefficients. A risk score for each patient was calculated as the sum of each protein's score, which was obtained by multiplying the expression level of the protein and its corresponding coefficient. The TCPA discovery cohort was divided into high-risk group and low-risk group with the median value of the risk scores. The differences in OS and the OS stratified by clinicopathological parameters were compared between high and low-risk group.

Validation of the protein based-prognostic signature using the TCPA validation cohort

Similarly, the TCPA validation cohort was divided into high-risk group and low-risk group with the same cut-

off value in the T CPA discovery cohort. The OS and the OS stratified by clinicopathological parameters were compared between high and low-risk group.

Nomogram model construction

The risk signature, age, gender, TNM stage and tumor grade were used to construct a nomogram model. Calibration curves were used to determine the agreement between model prediction outcome and actual outcome for one-year OS and three year OS.

Tissue samples and IHC analysis

One hundred and twelve formalin-fixed paraffin-embedded (FFPE) tissue specimens and the corresponding clinical information were obtained from the Stomatological Hospital, Southern Medical University (SMUSH cohort). All the HNSCC cases were pathologically confirmed. The detailed information of the patient cohort was summarized in Table 1. This study was approved by the Institutional Research Ethics Committee at the Stomatological Hospital, Southern Medical University. Written informed consent was obtained from all patients for the using their tissue samples.

For IHC analysis, FFPE tissue sections were deparaffinized by sequential washing with xylene, 100% ethanol, 95% ethanol, 80% ethanol and PBS. Followed by quenching with 0.3% H₂O₂ in methanol for 5 min, the slides were blocked with 5% BSA for 30 min. Then the sections were incubated with HER3_pY1289 primary antibody (1:200, Cell Signaling Technology, Danvers, MA, USA) overnight at 4 °C. After washing with PBS, the slides were incubated with horseradish peroxidase (HRP)-conjugated secondary antibody for 2 h at room temperature. For the quantitative analysis, the IHC score of HER3_pY1289 was obtained by multiplying the staining intensity (on a scale of 0-3: negative = 0, weak = 1, moderate = 2, and strong = 3) and the percentage of cells stained (on a scale of 0-4: 0 = 0%, 1 = 1%-25%, 2 = 26%-50%, 3 = 51%-75%, and 4 = 76%-100%).

Cell culture and siRNA transfection

The HNSCC cell lines SCC1 and SCC23 were cultured in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum, penicillin (100 U/mL), and streptomycin (100 µg/mL). Cultures were maintained at 37 °C in an atmosphere containing 95% air and 5% CO₂. Based on the manufacturer's manual, the cells were transfected with siRNAs targeting HER3 (siHER3 #1, siHER3 #2) and control siRNA (siCTRL) using the Lipofectamine®

RNAiMAX Transfection Reagent (Invitrogen, Carlsbad, CA, USA).

Western blotting

Equal amount of protein samples was loaded and separated on a 4-12% Bis-Tris NuPAGE gel (Invitrogen) and transferred onto a nitrocellulose membrane using a Trans-blot SD semi-dry transfer cell (Bio-Rad Laboratories Inc., Hercules, CA, USA). The membranes were blocked in TBST buffer containing 5% nonfat milk for 1 h at room temperature. Then the membranes were incubated with HER3_pY1289 primary antibody (1:500, CST) and HER3 antibody (1:500, CST) overnight in the cold room. After rinsing in TBST for three times, the membranes were incubated with HRP-link secondary antibody for 1 h at room temperature. ECL kit (Beyotime Biotech, Shanghai, China) was applied to visualize the bands.

MTT assay

The siHER3 treated cells and the control cells were seeded into 96-well plates at a density of 3000 cells/well. At the indicated time points (24h, 48h, 72h and 96h), 20 µL MTT solution (Sigma-Aldrich, St. Louis, MO, USA) was added to each well and the plates were incubated for 4 h at 37 °C. After removing the supernatant, the formazan products were dissolved by adding 200 µL dimethyl sulfoxide. Followed by shaking for 30s, the absorbance was determined using a microplate reader (Tecan, Mannedorf, Switzerland).

EdU (5-ethynyl-2'-deoxyuridine) assay

EdU assay was conducted with the Click-iT™ EdU Cell Proliferation Kit for Imaging (Invitrogen) based on the manufacturer's protocol. Briefly, EdU was added to the cells at the final concentration of 10 µM and incubated for 2 h at 37°C. After removing the supernatant, the cells were fixed with 3.7% formaldehyde in PBS for 20 min at room temperature. The fixative was removed and 0.5% Triton® X-100 (Sigma-Aldrich) in PBS was added to increase the permeability of the cellular membrane. The cells were stained with 1 × Click-iT reaction cocktail for 30 min at room temperature and protected from light. Hoechst 33342 dye was then used for nuclear staining. Images were obtained under a confocal laser scanning microscope (Olympus, Center Valley, PA, USA) and at least four random fields per well were photographed for data analysis.

Transwell matrigel invasion assay

After starvation for overnight, the cells were washed two times with PBS and then resuspended in pure

Table 1. The clinical information of the SMUSH cohort.

Clinicopathological features	Number
Age	
Mean ± SD	60.54 ± 8.13
Gender, n (%)	
Male	84 (75.00%)
Female	28 (25.00%)
Pathological diagnosis	
Squamous cell carcinoma	112 (100%)
Tumor grade	
G1	48 (42.86%)
G2	42 (37.50%)
G3	22 (19.64%)
G4	0 (0.00%)
TNM stage	
Stage I	19 (16.96%)
Stage II	43 (38.39%)
Stage III	28 (25.00%)
Stage IV	22 (19.64%)
Lymph node metastasis	
Negative	47 (41.96%)
Positive	65 (58.04%)
Distant metastasis	
No	110 (98.21%)
Yes	2 (1.79%)
HPV infection	
Negative	66 (70.21%)
Positive	28 (29.79%)

DMEM. Then 5×10^5 cells were added to the upper chamber of transwell matrigel invasion inserts (BD Biosciences, Bedford, MA, USA). The lower chamber is filled with 1 mL of complete growth medium. After 24 h, a cotton swab was used to remove the cells remaining on the upper surface of the membrane. Then the cells that had invaded through the membrane were fixed by 3.7% formaldehyde and stained with the crystal violet. At least four random fields per insert were captured for data analysis.

Statistical analysis

The statistical analysis was performed with GraphPad Prism 8.0 (GraphPad Software Inc., San Diego, CA, USA). The data of *in vitro* experiments were expressed as the mean ± standard deviation, and analyzed by the one-way ANOVA. For the SMUSH cohort, the median value of the IHC scores was used to divide the HNSCC patients into the high IHC score group and low IHC score group. The OS difference between high and low IHC score group was determined by the Kaplan-Meier method and log-rank test. All statistical analyses were two-sided. A *p* value of less than .05 was considered statistically significant.

AUTHOR CONTRIBUTIONS

XZ, XL and LC contributed to the conception and design of the study, performed the experiments, analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This study was supported by the grants from National Natural Science Foundation of China (No. 81901006), Medical Scientific Research Foundation of Guangdong Province (No. A2019299) and Pearl River S&T Nova Program of Guangzhou (201806010012).

REFERENCES

1. Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. *Lancet*. 2008; 371:1695–709.

- [https://doi.org/10.1016/S0140-6736\(08\)60728-X](https://doi.org/10.1016/S0140-6736(08)60728-X)
PMID:[18486742](https://pubmed.ncbi.nlm.nih.gov/18486742/)
2. Pai SI, Westra WH. Molecular pathology of head and neck cancer: implications for diagnosis, prognosis, and treatment. *Annu Rev Pathol.* 2009; 4:49–70.
<https://doi.org/10.1146/annurev.pathol.4.110807.092158> PMID:[18729723](https://pubmed.ncbi.nlm.nih.gov/18729723/)
 3. Liu X, Zhao X, Gou C. Identification of novel methylated DNA marker ZNF569 for head and neck squamous cell carcinoma. *J Cancer.* 2019; 10:2250–60.
<https://doi.org/10.7150/jca.31156>
PMID:[31258729](https://pubmed.ncbi.nlm.nih.gov/31258729/)
 4. Zhao X, Cui L. Development and validation of a m⁶A RNA methylation regulators-based signature for predicting the prognosis of head and neck squamous cell carcinoma. *Am J Cancer Res.* 2019; 9:2156–69.
PMID:[31720080](https://pubmed.ncbi.nlm.nih.gov/31720080/)
 5. Peng Y, Wu Q, Wang L, Wang H, Yin F. A DNA methylation signature to improve survival prediction of gastric cancer. *Clin Epigenetics.* 2020; 12:15.
<https://doi.org/10.1186/s13148-020-0807-x>
PMID:[31959204](https://pubmed.ncbi.nlm.nih.gov/31959204/)
 6. Zhao X, Sun S, Zeng X, Cui L. Expression profiles analysis identifies a novel three-mRNA signature to predict overall survival in oral squamous cell carcinoma. *Am J Cancer Res.* 2018; 8:450–61.
PMID:[29637000](https://pubmed.ncbi.nlm.nih.gov/29637000/)
 7. Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* 2009; 583:3966–73.
<https://doi.org/10.1016/j.febslet.2009.10.036>
PMID:[19850042](https://pubmed.ncbi.nlm.nih.gov/19850042/)
 8. Tibes R, Qiu Y, Lu Y, Hennessy B, Andreeff M, Mills GB, Kornblau SM. Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Mol Cancer Ther.* 2006; 5:2512–21.
<https://doi.org/10.1158/1535-7163.MCT-06-0334>
PMID:[17041095](https://pubmed.ncbi.nlm.nih.gov/17041095/)
 9. Hennessy BT, Lu Y, Gonzalez-Angulo AM, Carey MS, Myhre S, Ju Z, Davies MA, Liu W, Coombes K, Meric-Bernstam F, Bedrosian I, McGahren M, Agarwal R, et al. A technical assessment of the utility of reverse phase protein arrays for the study of the functional proteome in non-microdissected human breast cancers. *Clin Proteomics.* 2010; 6:129–51.
<https://doi.org/10.1007/s12014-010-9055-y>
PMID:[21691416](https://pubmed.ncbi.nlm.nih.gov/21691416/)
 10. Li J, Akbani R, Zhao W, Lu Y, Weinstein JN, Mills GB, Liang H. Explore, visualize, and analyze functional cancer proteomic data using the cancer proteome atlas. *Cancer Res.* 2017; 77:e51–54.
<https://doi.org/10.1158/0008-5472.CAN-17-0369>
PMID:[29092939](https://pubmed.ncbi.nlm.nih.gov/29092939/)
 11. Chen MM, Li J, Wang Y, Akbani R, Lu Y, Mills GB, Liang H. TCGA v3.0: an integrative platform to explore the pan-cancer analysis of functional proteomic data. *Mol Cell Proteomics.* 2019; 18:S15–25.
<https://doi.org/10.1074/mcp.RA118.001260>
PMID:[31201206](https://pubmed.ncbi.nlm.nih.gov/31201206/)
 12. Dhingra V, Verma J, Misra V, Srivastav S, Hasan F. Evaluation of cyclin D1 expression in head and neck squamous cell carcinoma. *J Clin Diagn Res.* 2017; 11:EC01–04.
<https://doi.org/10.7860/JCDR/2017/21760.9329>
PMID:[28384866](https://pubmed.ncbi.nlm.nih.gov/28384866/)
 13. Ramos-García P, González-Moles MÁ, Ayén Á, González-Ruiz L, Gil-Montoya JA, Ruiz-Ávila I. Predictive value of CCND1/cyclin D1 alterations in the Malignant transformation of potentially Malignant head and neck disorders: systematic review and meta-analysis. *Head Neck.* 2019; 41:3395–407.
<https://doi.org/10.1002/hed.25834> PMID:[31184805](https://pubmed.ncbi.nlm.nih.gov/31184805/)
 14. Peterle GT, Maia LL, Trivilin LO, de Oliveira MM, Dos Santos JG, Mendes SO, Stur E, Agostini LP, Rocha LA, Moysés RA, Cury PM, Nunes FD, Louro ID, et al. PAI-1, CAIX, and VEGFA expressions as prognosis markers in oral squamous cell carcinoma. *J Oral Pathol Med.* 2018; 47:566–74.
<https://doi.org/10.1111/jop.12721> PMID:[29693741](https://pubmed.ncbi.nlm.nih.gov/29693741/)
 15. Speleman L, Kerrebijn JD, Look MP, Meeuwis CA, Foekens JA, Berns EM. Prognostic value of plasminogen activator inhibitor-1 in head and neck squamous cell carcinoma. *Head Neck.* 2007; 29:341–50.
<https://doi.org/10.1002/hed.20527> PMID:[17163465](https://pubmed.ncbi.nlm.nih.gov/17163465/)
 16. Nanda SS, Gandhi AK, Rastogi M, Khurana R, Hadi R, Sahni K, Mishra SP, Srivastava AK, Bhatt ML, Parmar D. Evaluation of XRCC1 gene polymorphism as a biomarker in head and neck cancer patients undergoing chemoradiation therapy. *Int J Radiat Oncol Biol Phys.* 2018; 101:593–601.
<https://doi.org/10.1016/j.ijrobp.2018.03.039>
PMID:[29893275](https://pubmed.ncbi.nlm.nih.gov/29893275/)
 17. Kumar A, Pant MC, Singh HS, Khandelwal S. Reduced expression of DNA repair genes (XRCC1, XPD, and OGG1) in squamous cell carcinoma of head and neck in north India. *Tumour Biol.* 2012; 33:111–19.
<https://doi.org/10.1007/s13277-011-0253-7>
PMID:[22081374](https://pubmed.ncbi.nlm.nih.gov/22081374/)
 18. Tan FH, Bai Y, Saintigny P, Darido C. mTOR signalling in head and neck cancer: heads up. *Cells.* 2019; 8:333.
<https://doi.org/10.3390/cells8040333>
PMID:[30970654](https://pubmed.ncbi.nlm.nih.gov/30970654/)

19. Wang Z, Valera JC, Zhao X, Chen Q, Gutkind JS. mTOR co-targeting strategies for head and neck cancer therapy. *Cancer Metastasis Rev.* 2017; 36:491–502. <https://doi.org/10.1007/s10555-017-9688-7> PMID:[28822012](https://pubmed.ncbi.nlm.nih.gov/28822012/)
20. Yang H, Rudge DG, Koos JD, Vaidialingam B, Yang HJ, Pavletich NP. mTOR kinase structure, mechanism and regulation. *Nature.* 2013; 497:217–23. <https://doi.org/10.1038/nature12122> PMID:[23636326](https://pubmed.ncbi.nlm.nih.gov/23636326/)
21. Brand TM, Hartmann S, Bholra NE, Peyser ND, Li H, Zeng Y, Isaacson Wechsler E, Ranall MV, Bandyopadhyay S, Duvvuri U, LaVallee TM, Jordan RC, Johnson DE, Grandis JR. Human papillomavirus regulates HER3 expression in head and neck cancer: implications for targeted HER3 therapy in HPV⁺ patients. *Clin Cancer Res.* 2017; 23:3072–83. <https://doi.org/10.1158/1078-0432.CCR-16-2203> PMID:[27986750](https://pubmed.ncbi.nlm.nih.gov/27986750/)
22. Takikita M, Xie R, Chung JY, Cho H, Ylaya K, Hong SM, Moskaluk CA, Hewitt SM. Membranous expression of Her3 is associated with a decreased survival in head and neck squamous cell carcinoma. *J Transl Med.* 2011; 9:126. <https://doi.org/10.1186/1479-5876-9-126> PMID:[21801427](https://pubmed.ncbi.nlm.nih.gov/21801427/)
23. Lyu H, Han A, Polsdofer E, Liu S, Liu B. Understanding the biology of HER3 receptor as a therapeutic target in human cancer. *Acta Pharm Sin B.* 2018; 8:503–10. <https://doi.org/10.1016/j.apsb.2018.05.010> PMID:[30109175](https://pubmed.ncbi.nlm.nih.gov/30109175/)
24. Akbani R, Ling S, Lu Y. Generation of raw RPPA data and their conversion to analysis-ready data. *Adv Exp Med Biol.* 2019; 1188:165–80. https://doi.org/10.1007/978-981-32-9755-5_9 PMID:[31820388](https://pubmed.ncbi.nlm.nih.gov/31820388/)

SUPPLEMENTARY MATERIALS

Supplementary Tables

Supplementary Table 1. The overall survival associated proteins identified by univariate analysis.

Protein ID	HR	HR.95L	HR.95H	p value
CYCLIND1	6.75	2.69	16.93	0.00005
CMET_pY1235	10.33	2.83	37.70	0.00041
IRS1	3.20	1.53	6.71	0.00200
HER3_pY1289	4.69	1.73	12.74	0.00244
XRCC1	0.29	0.13	0.67	0.00379
MSH6	0.47	0.27	0.81	0.00633
BID	2.59	1.30	5.15	0.00673
BRAF_pS445	0.29	0.12	0.72	0.00748
RAPTOR	4.02	1.44	11.21	0.00791
ECADHERIN	0.79	0.66	0.94	0.00805
PAI1	1.41	1.09	1.82	0.00911
BAP1C4	0.54	0.33	0.87	0.01198
BETACATENIN	0.75	0.60	0.94	0.01329
ANNEXINVII	0.26	0.09	0.76	0.01413
PAXILLIN	1.68	1.11	2.55	0.01500
PDCD4	0.68	0.50	0.93	0.01680
CAVEOLIN1	1.29	1.04	1.60	0.01995
BIM	0.54	0.32	0.92	0.02387
SYK	0.61	0.40	0.94	0.02453
CYCLINE2	0.43	0.20	0.90	0.02490
TUBERIN_pT1462	0.26	0.08	0.87	0.02840
PTEN	0.61	0.39	0.95	0.02854
PCNA	0.38	0.16	0.91	0.02941
P70S6K1	0.41	0.18	0.92	0.03045
EEF2K	0.71	0.52	0.97	0.03199
CLAUDIN7	0.72	0.53	0.98	0.03439
FIBRONECTIN	1.60	1.03	2.50	0.03684
PEA15	2.18	1.05	4.54	0.03761
PI3KP110ALPHA	0.48	0.24	0.97	0.03980
ADAR1	0.39	0.16	0.97	0.04365
NF2	2.03	1.02	4.05	0.04473
MSH2	0.46	0.21	0.99	0.04851
RAD51	4.11	1.00	16.81	0.04933

Supplementary Table 2. The clinical information of the discovery cohort.

Clinicopathological features	Number
Age	
>=60	78 (52.35%)
<60	71 (47.65%)
Gender, n (%)	
Male	103 (69.13%)
Female	46 (30.87%)
Pathological diagnosis	
Squamous cell carcinoma	149 (100%)
Tumor grade	
G1	17 (11.41%)
G2	95 (63.76%)
G3	37 (24.83%)
G4	0 (0.00%)
TNM stage	
Stage I	10 (6.71%)
Stage II	28 (18.79%)
Stage III	29 (19.46%)
Stage IV	82 (55.03%)

Supplementary Table 3. The clinical information of the validation cohort.

Clinicopathological features	Number
Age	
>=60	84 (54.90%)
<60	69 (45.10%)
Gender, n (%)	
Male	109 (71.24%)
Female	44 (28.76%)
Pathological diagnosis	
Squamous cell carcinoma	153 (100%)
Tumor grade	
G1	25 (16.34%)
G2	92 (60.13%)
G3	36 (23.53%)
G4	0 (0.00%)
TNM stage	
Stage I	4 (2.61%)
Stage II	25 (16.34%)
Stage III	28 (18.30%)
Stage IV	96 (62.75%)