Observational Study

Medicine

OPEN

Construction of a prognostic model for radical esophagectomy based on immunohistochemical prognostic markers combined with clinicopathological factors

Bo Wang, MD[a,b], Anna Su, MS[c], Mengyan Li, PhD[a], Xuecheng Wang, MS[d], Yongmei Gao, MS[a], Liping Su, PhD[a], Wan Li, MS[a], Yuqing Ma, PhD[a,*]

Abstract

Esophageal squamous cell carcinoma (ESCC) has a poor prognosis and lacks effective biomarkers to evaluate prognosis and treatment. Glycoprotein nonmetastatic melanoma protein B (GPNMB) is a protein highly expressed in ESCC tissues screened by isobaric tags for relative and absolute quantitation proteomics, which has significant prognostic value in a variety of malignant tumors, but its relationship with ESCC remains unclear. By immunohistochemical staining of 266 ESCC samples, we analyzed the relationship between GPNMB and ESCC. To explore how to improve the ability of ESCC prognostic assessment, we established a prognostic model of GPNMB and clinicopathological features. The results suggest that GPNMB expression is generally positive in ESCC tissues and is significantly associated with poorer differentiation, more advanced American Joint Council on Cancer (AJCC) stage, and higher tumor aggressiveness ($P < .05$). Multivariate Cox analysis indicated that GPNMB expression was an independent risk factor for ESCC patients. A total of 188 (70%) patients were randomly selected from the training cohort and the four variables were automatically screened by stepwise regression based on the AIC principle: GPNMB expression, nation, AJCC stage and nerve invasion. Through the weighted term, we calculate the risk score of each patient, and by drawing the receiver operating characteristic curve, we show that the model has good prognostic evaluation performance. The stability of the model was verified by test cohort. Conclusion: GPNMB is a prognostic marker consistent with the characteristics of tumor therapeutic targets. For the first time, we constructed a prognostic model combining immunohistochemical prognostic markers and clinicopathological features in ESCC, which showed higher prognostic efficacy than AJCC staging system in predicting the prognosis of ESCC patients in this region.

Abbreviations: AJCC = American Joint Council on Cancer, AUC = area under the curve, ESCC = esophageal squamous cell carcinoma, GPNMB = glycoprotein nonmetastatic melanoma protein B, IHC = immunohistochemical, ITRAQ = isobaric tags for relative and absolute quantitation, OS = overall survival, ROC curve = receiver operating characteristic curve.

Keywords: esophageal squamous cell carcinoma (ESCC), GPNMB, immunohistochemical (IHC), prognostic

1. Introduction

Esophageal squamous cell carcinoma (ESCC) is one of the most common cancer types in China, with poor prognosis.[1,2] Xinjiang is an area with a high incidence of ESCC, and studies have shown that the morbidity and mortality of ESCC are higher in Hazak than in Han.[3,4] ESCC lacks effective therapeutic targets and prognostic biomarkers. Therefore, exploring the differences in biomarkers and prognosis between Hazak and Han patients with ESCC is of great significance for studying the pathogenesis of esophageal squamous cell carcinoma.

BW and AS contributed equally to this work.

This study was supported by Xinjiang Uygur Autonomous Region Youth Science and Technology Talent Project (no. WJWY-XCZX202224); Natural Science Foundation of Xinjiang Uygur Autonomous Region (no. 2021D01A12); National Natural Science Foundation of China (no. 81860422).

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.; All data generated or analyzed during this study are included in this published article [and its supplementary information files].

All patients involved in this study were informed and signed the scientific research consent form; At later follow-up, the current study was re-explained to patients or their families. Patients' hospital numbers and hospitalization information were omitted from the study.

Supplemental Digital Content is available for this article.

[a] Department of Pathology, The First Affiliated Hospital, Xinjiang Medical University, Urumqi, China, [b] Xinjiang Medical University, Urumqi, China, [c] Internal Medicine, Urumqi First People's Hospital, Urumqi, China, [d] Thoracic Surgery, The Fifth Affiliated Hospital, Xinjiang Medical University, Urumqi, China.

**Correspondence: Yuqing Ma, Department of Pathology, The First Affiliated Hospital, Xinjiang Medical University, Urumqi, Xinjiang, China (e-mail: yuqingm0928@126.com).*

How to cite this article: Wang B, Su A, Li M, Wang X, Gao Y, Su L, Li W, Ma Y. Construction of a prognostic model for radical esophagectomy based on immunohistochemical prognostic markers combined with clinicopathological factors. Medicine 2023;102:9(e32889).

Received: 13 July 2022 / Received in final form: 12 January 2023 / Accepted: 18 January 2023

http://dx.doi.org/10.1097/MD.0000000000032889

Glycoprotein nonmetastatic melanoma protein B (GPNMB) is a type I transmembrane protein composed of 576 amino acids and contains three domains: extracellular domain, transmembrane region and cytoplasmic domain.[5] GPNMB gene is located on chromosome 7q15 and is involved in cell proliferation, apoptosis and differentiation.[6,7] Studies have shown that GPNMB expression is up-regulated in head and neck squamous cell carcinoma, and its high expression is associated with poor prognosis.[8] Huang et al[9] found that GPNMB was an independent prognostic factor for triple negative breast cancer recurrence and metastasis. It promotes the invasiveness of triple negative breast cancer by participating in epithelial-mesenchymal transition. In non-small-cell lung cancer, GPNMB can activate epidermal growth factor receptor mutations and promote tumor progression.[10] The relationship between GPNMB and ESCC remains unclear.

The prognostic assessment of ESCC relies on American Joint Council on Cancer (AJCC) staging system, which is backward compared with the biomarker classification of breast cancer, lung cancer and other malignant tumors. Clinicopathological factors alone may not be sufficient to identify patients at high risk of disease progression. The identification of biomarkers with prognostic value and the construction of disease prediction models can enable clinicians to better judge the prognosis of patients.

Immunohistochemistry (IHC) is a practical clinical tool, and many biomarkers for IHC to predict ESCC prognosis have been reported.[11,12] Researchers used 6-IHC prognostic markers to construct a prediction model for ESCC recurrence and proved the reliability of the model.[13] However, the ESCC prognostic model constructed by IHC prognostic markers combined with clinicopathological factors has not been found yet.

In this study, Isobaric tags for relative and absolute quantitation (ITRAQ)[14] was used to screen out the highly expressed GPNMB in ESCC tissues of Han and Hazak patients in this region, and the relationship between GPNMB and ESCC was analyzed according to IHC results. The proportional Hazards Model was used to establish and verify a prognostic model based on GPNMB IHC and clinicopathological parameters. Finally, bioinformatics was used to explore the biological functions of GPNMB in ESCC (Fig. 1).

2. Materials and methods

2.1. Patients and samples

Immunohistochemical samples: Complete clinical data of ESCC patients undergoing esophagotomy in the First Affiliated Hospital of Xinjiang Medical University from January 2014 to December 2020 were collected. A total of 340 pairs of tumor and adjacent tissue samples (formalin fixation and paraffin embedding) were collected to create tissue chips (Supplementary Figure S1 and S2, Supplemental Digital Content, http://links.lww.com/MD/I440). Inclusion criteria: patients with ESCC who had received radical resection of esophageal carcinoma in the first affiliated hospital of Xinjiang Medical University from January 2014 to December 2020 with complete clinical data; no preoperative radiotherapy or chemotherapy was received; esophagus is the primary lesion site; and Han and Hazak patients. Exclusion criteria: patients without ESCC, such as esophageal adenocarcinoma; patients who have received radiotherapy or chemotherapy before surgery; non-esophageal primary tumor, such as tumor metastasis to other sites of the esophagus; patients without radical resection

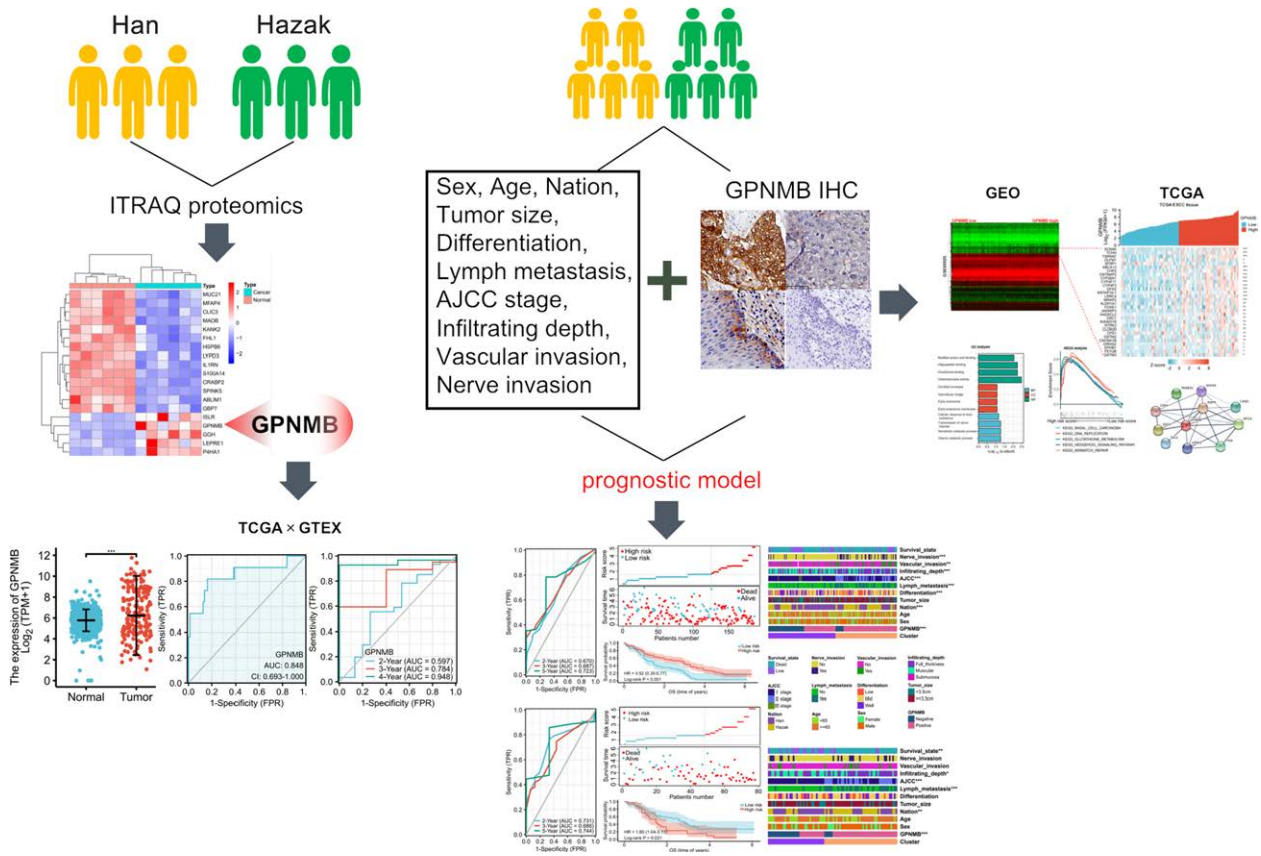

Figure 1. Abstract of pictures. AJCC = American Joint Council on Cancer, AUC = area under the curve, ESCC = esophageal squamous cell carcinoma, FPR = false positive rate, GEO = gene expression omnibus data base, GO = gene ontology, GPNMB = glycoprotein nonmetastatic melanoma protein B, IHC = immunohistochemical, KEGG = Kyoto Encyclopedia of Genes and Genomes, OS = overall survival, TPM = transcripts per kilobase of exon model per million mapped reads, TPR = true positive rate.

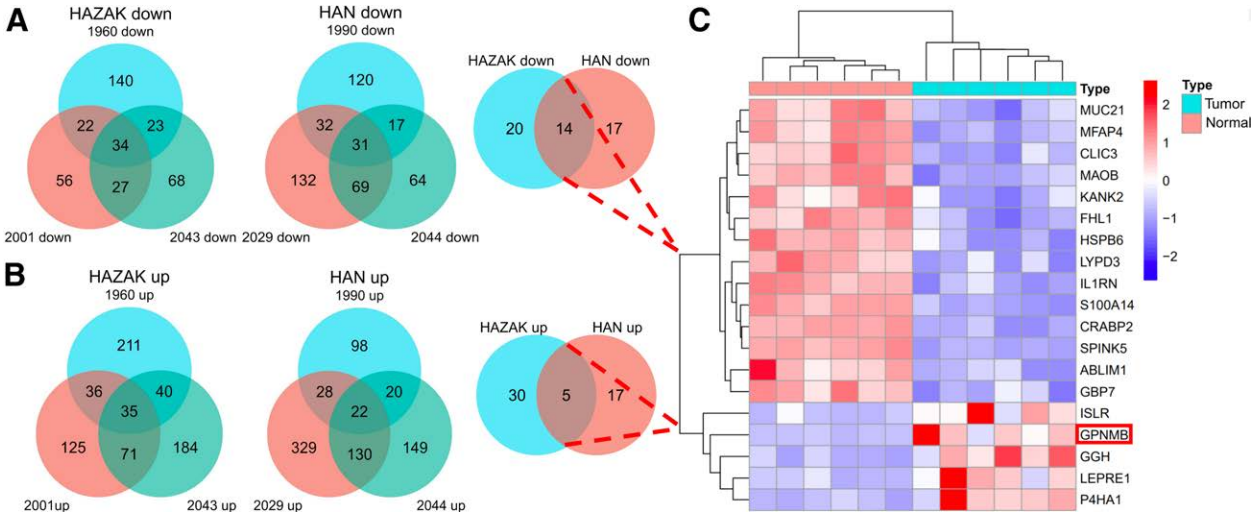

Figure 2. ITRAQ screen for differentially expressed proteins in ESCC. A, Up-regulated proteins in Han and Hazak cancer tissues. B, Down-regulated proteins in Han and Hazak cancer tissues. C, Heat map showing proteins that are co-up- and down-regulated in Han and Hazak. ESCC = esophageal squamous cell carcinoma, GPNMB = glycoprotein nonmetastatic melanoma protein B, ITRAQ = isobaric tags for relative and absolute quantitation.

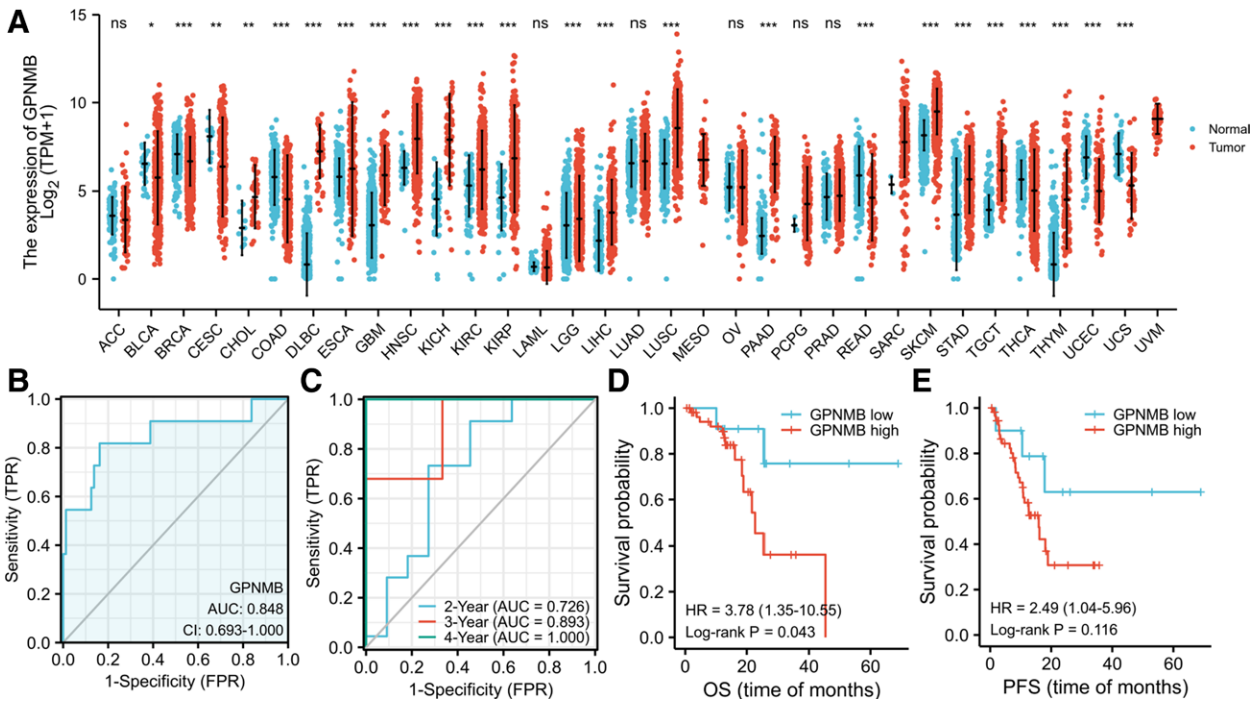

Figure 3. Evaluation of GPNMB expression, diagnosis, and prognosis in ESCA in public databases. A, GPNMB expression in pan-cancer. B, Diagnostic ROC curve of GPNMB expression. C, Time-dependent ROC curve of GPNMB expression. D–F, GPNMB expression versus OS and PFS grouped based on the smallest *P* value. GPNMB = glycoprotein nonmetastatic melanoma protein B, OS = overall survival, PFS = progression free survival, ROC curve = receiver operating characteristic curve.

of esophageal cancer; patients from other ethnic groups, such as Uygur and Mongolian; patients who died during surgery or hospitalization; Patients with distant metastasis before surgery; Tis and T1a stage; patients with incomplete tissue samples; and lost patients who cannot be contacted. According to the above criteria, 14 cases of distant metastasis, 16 cases of T1a stage, 21 cases of sparse tissue or specimen detachment, and 23 cases of lost follow-up were excluded from the 340 enrolled samples. The histological diagnosis and differentiation degree of ESCC were confirmed by 2 deputy chief pathologists in 266 samples, including 149 cases of Han nationality and 117 cases of Hazak nationality, including the following clinicopathological data: age (<65; ≥65 years old), gender (male; Female), tumor size (<3 cm; ≥3 cm), degree of differentiation (high;; Low), lymph node metastasis (yes; No), vascular invasion (yes; No), nerve invasion (yes; No), AJCC(esophageal cancer is divided into stage I/II and III according to the eighth edition of AJCC and Union for International Cancer Control in 2017), depth of invasion (mucosa; muscular layer; full thickness). The follow-up period was up to December 2021.

ITRAQ samples: 6 pairs of fresh tissue and normal adjacent tissue of esophageal squamous cell carcinoma underwent radical esophageal carcinoma operation in the First Affiliated Hospital of Xinjiang Medical University (3 pairs of Han nationality and 3 pairs of Hazak nationality) were collected.

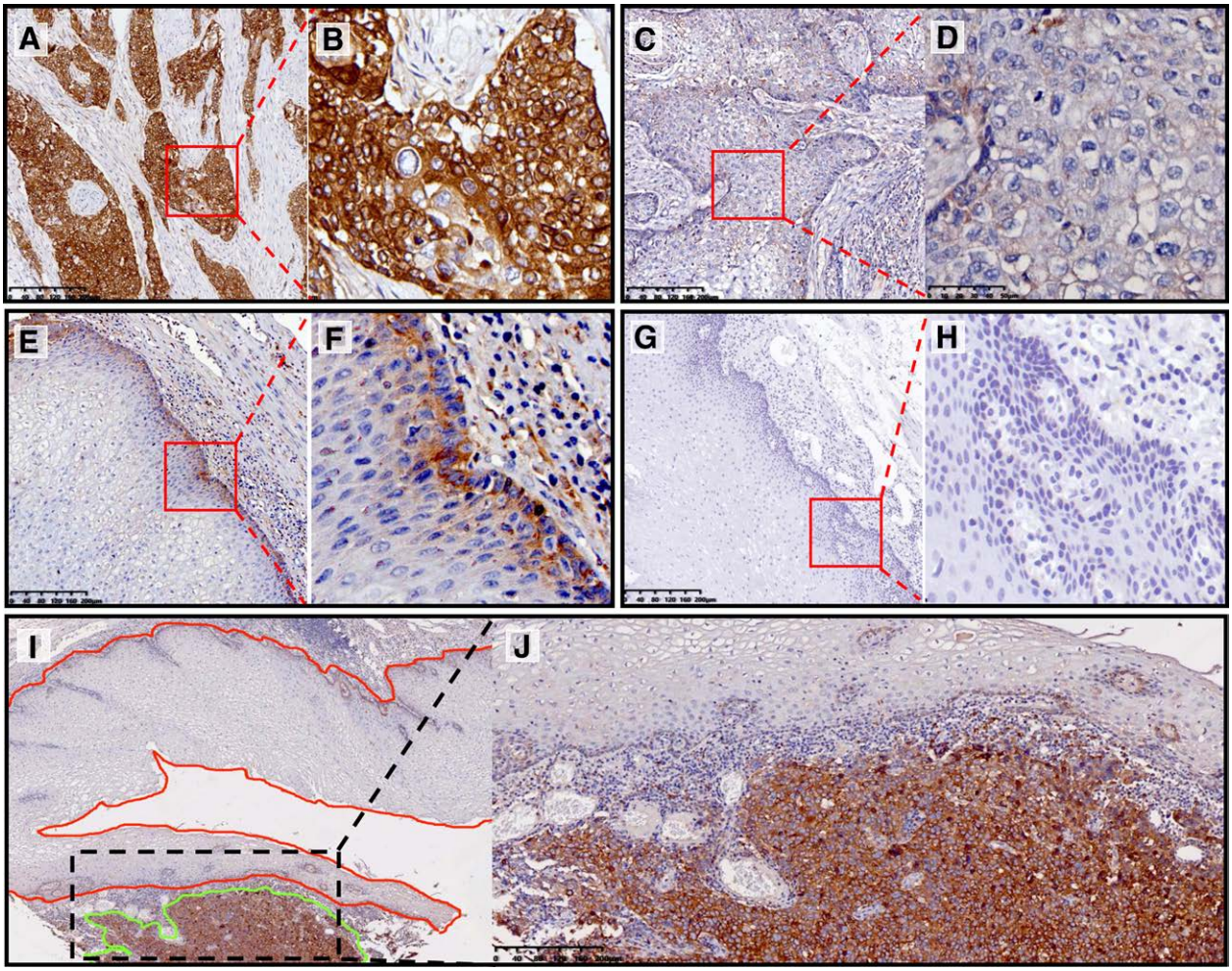

Figure 4. Illustration of GPNMB IHC staining. A–B, GPNMB was positively expressed in ESCC tissues under low and high magnification. C–D, GPNMB was negatively expressed in ESCC tissues under low and high magnification. E–F, GPNMB is positively expressed in esophageal squamous epithelium under low and high magnification. G–H, GPNMB is negatively expressed in esophageal squamous epithelium under low and high magnification. I–J, A section contained both cancerous tissue and normal esophageal tissue (The red outline marks the normal squamous epithelium and the green outline marks the ESCC tissue). ESCC = esophageal squamous cell carcinoma, GPNMB = glycoprotein nonmetastatic melanoma protein B, IHC = immunohistochemical.

This study was approved by the Institutional Review Board of the First Affiliated Hospital of Xinjiang Medical University and carried out in accordance with the Declaration of Helsinki.

2.2. Proteomics analysis of ITRAQ

Protein expression data of fresh samples were obtained by protein extraction, concentration measurement, sodium dodecyl sulfate electrophoresis, protein enzymatic hydrolysis, ITRAQ labeling, strong cation exchange separation and liquid tandem mass spectrometry. LogFC > 1.5, $P < .05$ was used to screen differential proteins in paracancer and cancer tissues.

2.3. Immunohistochemistry (IHC)

The paraffin-embedded ESCC tissues were made into tissue chips, sliced into 4 μm sections, deparaffinized in xylene and rehydrated in 100%, 95%, 80%, and 70% ethanol. Following treatment with 3% hydrogen peroxide to block endogenous peroxidase activity, the sections were heated with ethylenediaminetetraacetic acid (pH9.0) in boiling water at 100°C for antigen retrieval. The sections were then treated with goat serum (ZSGB-BIO, ZLI-9022) at room temperature to block nonspecific antigens for 30 minutes. Then use anti-GPNMB Antibody (1:1000, Abcam Inc., United Kingdom, AB222109) overnight at 4°C. Sections were incubated with a peroxise-labeled polymer (ZSGB-BIO, Beijing, China, PV-6001) as a secondary antibody for 30 minutes. The slides were subsequently stained with DAB, dehydrated, sealed and observed under a light microscope (DM300; Leica Microsystems GmbH, Hesse, Germany; magnifications, ×10 and ×40). GPNMB immunohistochemical staining of ESCC and normal squamous epithelial tissue chips is shown in Supplementary Figures S1 and S2, Supplemental Digital Content, http://links.lww.com/MD/I440.

2.4. Immunohistochemical score

GPNMB positive particles were mainly located in the cell membranes of ESCC and normal squamous epithelial tissues. Staining index = percentage of positive cells × staining intensity, the presence of light yellow or brown-yellow fine particles in cell membrane or cytoplasm was considered as positive color, and the scoring rules of staining intensity were as follows: 0 (negative), 1 (light brown), 2 (brown), 3 (dark brown); The scoring rules of percentage of staining positive cells: 0 points (0–10%), 1 point (11–25%), 2 points (26–50%), 3 points (51–75%), 4 points (76–100%) (positively stained cells as a percentage of the same type of cells, not as a percentage of all cell types). The total staining index was 12. A score greater than 4 was defined as positive expression, otherwise it was defined as negative expression.

Table 1

Association between GPNMB expression and clinicopathological parameters of patients with ESCC.

Characteristics	GPNMB		*P* value
	Negative	Positive	
Sex			
Male	31	46	.117
Female	57	132	
Age			
<65	34	66	.893
≥65	54	112	
Nation			
Han	44	105	.190
Hazak	44	73	
Tumor size			
<3 cm	30	59	.891
≥3 cm	58	119	
Differentiation			
Low + Mid	58	138	.043
Well	30	40	
Lymph metastasis			
No	64	113	.167
Yes	24	65	
Infiltrating depth			
Mucous membrane	7	5	.011
Muscle layer	45	69	
whole layer	36	104	
AJCC stage			
I	5	2	.022
II	66	122	
III	17	54	
Vascular invasion			
No	70	141	.872
Yes	17	37	
Nerve invasion			
No	68	140	.875
Yes	20	38	

AJCC = American Joint Committee on Cancer, ESCC = esophageal squamous cell carcinoma, GPNMB = glycoprotein nonmetastatic melanoma protein B.

2.5. Model construction

Using the "caret" R package, 266 patients with ESCC were randomized 7:3 into a training cohort and a test cohort. Data from 188 patients in the training cohort were used to build a prognostic model. Survival status is "dead" and "survival," and survival time is overall survival (OS). Use the Surv function of the "Survival" package to create survival objects, establish a Cox regression model, and compare the differences in survival rates; use the step function to screen the features of the cox model by stepwise regression based on AIC indicators.

2.6. Statistical methods

All statistical analyses were performed using SPSS 22.0 (SPSS, Chicago, IL) and R (version 3.1.2). Categorical variables were compared using the chi-square test or Fisher's exact test. Survival curves were drawn using the Kaplan–Meier method, and the log-rank test was used for comparison. Univariate and multivariate Cox regression analysis was used to analyze the independent effect of this feature on prognosis. receiver operating characteristic (ROC) curve analysis was performed using the "pROC" package.

3. Result

3.1. ITRAQ analysis of differential proteins in ESCC

ITRAQ analysis was performed on the ESCC and adjacent tissues of three pairs of Han and three Hazak patients. There were 34 and 31 down-regulated proteins in the cancer tissues of Hazak and Han patients, respectively, and 14 proteins were down-regulated together. There were 35 and 22 highly expressed proteins in Hazak and Han cancer tissues, respectively, and 5 proteins were commonly up-regulated (Fig. 2B). Heatmap showing differential proteins screened by ITRAQ (Fig. 2C).

3.2. Public database analysis of the expression, diagnostic and prognostic value of GPNMB in ESCC

Figure 3A shows the expression of GPNMB in various malignant tumors including ESCA, and its expression in ESCA tissue is significantly higher than that in normal esophageal tissue. ROC curve analysis of GPNMB expression has a good diagnostic value in distinguishing normal esophageal tissue from ESCC tissue (Fig. 3B). Time-dependent ROC curves showed that GPNMB expression had good prognostic value in stage II and III patients (Fig. 3C). The expression of GPNMB in stage II and III ESCC patients correlated with patient prognosis (Fig. 3D–E, representing OS and PFS, respectively).

3.3. Expression of GPNMB in ESCC and adjacent tissues by IHC

GPNMB can be well expressed in ESCC tissues, and its positive staining accounts for 66.92% (178/266), mainly expressed in the cell membrane and cytoplasm, but not in stroma and lymphocytes (Fig. 4A and B). Figure 4C and D show images of negative expression of GPNMB in ESCC. GPNMB has low expression or no expression in normal esophageal tissue, and only 6.11% (11/180) of the samples with positive staining are expressed in basal layer (Fig. 4E–H). The expression of GPNMB in ESCC and in normal squamous epithelium can be represented by a single section where they coexist (Fig. 4I and J).

3.4. Association of GPNMB with clinicopathological features and prognosis of ESCC

Statistical analysis was performed according to the results of GPNMB IHC combined with the clinicopathological factors of patients. No significant correlation was found between the expression of GPNMB and sex, age, nation, tumor size, and lymph metastasis ($P < .05$). Positive GPNMB expression was significantly associated with worse differentiation, more advanced AJCC stage and higher tumor aggressiveness in ESCC ($P > .05$) (Table 1). Survival analysis showed that high GPNMB expression, poorer differentiation, lymph node metastasis, later AJCC staging, and nerve invasion were negatively correlated with patient survival ($P < .05$) (Fig. 5). In univariate analysis, tumor size, differentiation, AJCC stage, lymph metastasis, nerve invasion and GPNMB were important prognostic factors, which were consistent with the results of log-rank test. Multivariate analysis revealed that only nerve invasion and GPNMB positive remained independent predictors of OS (Table 2).

3.5. Establishment of a prognostic model

Baseline data of patients in the training cohort and the test cohort are shown in Table 3. GPNMB expression status and patient clinicopathological factors were used as variables. The established prognostic feature model calculated the risk score of each patient as follows: risk score = 0.446206*GPNMB + 0.248787*Nation + 0.771995*AJCC + 0.423623*Nerve invasion. In the formula, GPNMB negative is 0 and positive is 1. Han is 0 and Hazak is 1; AJCC I/II is 0 and III is 1; nerve invasion negative is 0 and positive is 1. Figure 6B shows the risk score and survival status of each patient. ROC curve showed that area under the curve (AUC) values of risk score evaluation

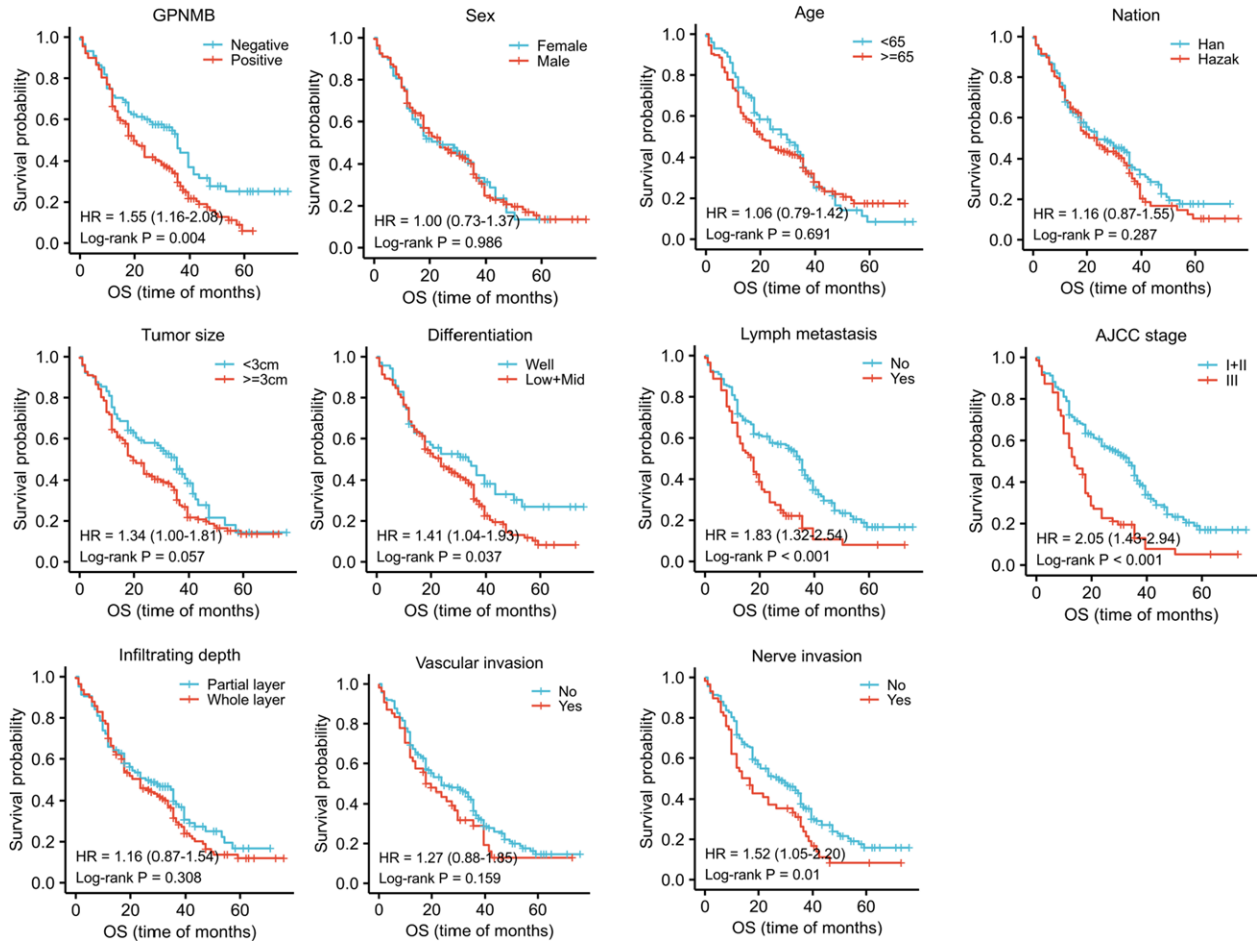

Figure 5. Association of GPNMB and clinicopathological variables with patient outcomes. AJCC = American Joint Council on Cancer, GPNMB = glycoprotein nonmetastatic melanoma protein B, HR = hazard ratio, OS = overall survival.

Table 2

Univariate and multivariate analyses of the OS of ESCC patients following radical resection.

	Univariate analysis		Multivariate analysis	
Variables	**HR (95% CI)**	***P* value**	**HR (95% CI)**	***P* value**
Sex (male vs female)	0.997 (0.728–1.366)	.987		
Age (yr) (>60 vs ≤60)	1.06 (0.79–1.423)	.698		
Infiltrating depth (whole layer vs partial layer)	1.157 (0.868–1.543)	.320		
Nation (Hazak vs Han)	1.164 (0.874–1.55)	.298		
Vascular invasion (yes vs no)	1.275 (0.901–1.804)	.170		
Tumor size (<3 vs ≥3 cm)	1.344 (0.983–1.838)	.064	1.196 (0.861–1.662)	.286
Differentiation (well vs low + mid)	0.703 (0.500–0.988)	.042	0.808 (0.569–1.148)	.234
AJCC stage (III vs I + II)	2.091 (1.535–2.849)	.000	1.511 (0.840–2.716)	.168
Lymph metastasis (yes vs no)	1.875 (1.391–2.528)	.000	1.305 (0.745–2.288)	.352
Nerve invasion (yes vs no)	1.524 (1.097–2.117)	.012	1.558 (1.116–2.175)	.009
GPNMB (positive vs negative)	1.567 (1.14–2.153)	.006	1.493 (1.081–2.062)	.015

The former of each scalar is the indicator light.

AJCC = American Joint Committee on Cancer, CI = confidence interval, ESCC = esophageal squamous cell carcinoma, GPNMB = glycoprotein nonmetastatic melanoma protein B, HR = hazard ratio.

patients in the training cohort were 0.669, 0.686, and 0.720 at 2, 3, and 4 years, respectively, indicating good predictive efficacy (Fig. 6A). According to the model formula, patients in the training cohort were divided into the high-risk group (96 cases, 51.06%) and the low-risk group (92 cases, 48.94%) according to the median value of risk score. Patients in the low-risk group had a better prognosis and higher 3-year OS and PFS (30.6% and 13.5% in the high-risk group; 44.7% and 32.8% in the low-risk group, respectively) (Fig. 6C and D).

3.6. Validation of prognostic models

The risk score and survival status of each patient in the test cohort (n = 78) were displayed (Fig. 6G). ROC curve verified the prognostic efficacy of risk score in the test cohort, and AUC values at 2, 3, and 4 years were 0.733, 0.683, and 0.678, respectively, indicating good stability of the model (Fig. 6F). Similarly, patients were divided into high-risk group (47 cases, 60.26%) and low-risk group (31 cases, 39.74%) based on risk score. Patients in the high-risk group had a worse prognosis, with a

Table 3

Pathoclinical characteristics of patients in discovery and validation cohort.

	Training set (n = 188)	Test set (n = 78)
Sex		
Male	135 (71.8%)	54 (69.2%)
Female	53 (28.2%)	24 (30.8%)
Age		
<65	71 (37.8%)	29 (37.2%)
≥65	117 (62.2%)	49 (62.8%)
Nation		
Han	106 (56.4%)	43 (55.1%)
Hazak	82 (43.6%)	35 (44.9%)
Tumor size		
<3 cm	67 (35.6%)	22 (28.2%)
≥3 cm	121 (64.4%)	56 (71.8%)
Differentiation		
Low	35 (18.6%)	17 (21.8%)
Mid	104 (55.3%)	40 (51.3%)
Well	49 (26.1%)	21 (26.9)
Lymph metastasis		
No	128 (68.1%)	49 (62.8%)
Yes	60 (31.9%)	29 (37.2%)
Infiltrating depth		
Mucous membrane	8 (4.3%)	4 (5.1%)
Muscle layer	82 (43.6%)	32 (41.0%)
Whole layer	98 (52.1%)	42 (53.8%)
AJCC stage		
I	6 (3.2%)	1 (1.3%)
II	136 (72.3%)	52 (66.7%)
III	46 (24.5)	25 (32.1%)
Vascular invasion		
No	145 (77.1%)	66 (84.6%)
Yes	43 (22.9%)	12 (15.4%)
Nerve invasion		
No	149 (79.3%)	59 (75.6%)
Yes	39 (20.7%)	19 (24.4%)
GPNMB		
Negative	64 (34.0%)	24 (30.8%)
Positive	124 (66.0%)	54 (69.2%)

AJCC = American Joint Committee on Cancer, ESCC = esophageal squamous cell carcinoma, GPNMB = glycoprotein nonmetastatic melanoma protein B.

3-year OS of 16.2% and a 3-year PFS of 8.4%, and patients in the low-risk group had a 3-year OS of 47.1% and a PFS of 48.9% (Fig. 6H and I). Similar differences between the two groups were noted in the combined training and test cohort (Fig. 6E and J).

3.7. Relationship between GPNMB and clinicopathological data and prognostic model

We performed statistical analyses on the distribution of GPNMB expression and clinicopathological variables in the training and test cohorts. Figure 7A shows the differences in the distribution of all variables in the training cohort between the high- and low-risk groups. The distribution of GPNMB, nation, differentiation, lymph metastasis, AJCC stage, infiltrating depth, vascular invasion and nerve invasion in the training cohort was significantly different between the low-risk group and the high-risk group. The distribution of GPNMB, nation, lymph, AJCC stage, infiltrating depth and survival state is different in test cohort (Fig. 7B). This indicated that GPNMB, Nation, Lymph metastasis, AJCC stage and infiltrating depth were the main risk factors in ESCC patients. ROC curve was used to test the independent prognostic ability of risk score, GPNMB, nation, lymph, AJCC stage and infiltrating depth. The AUC for predicting patient survival at 2, 3, and 4 years was higher than other factors, suggesting that the prognostic model had better predictive performance than any single factor (Fig. 7C–H).

3.8. To explore the molecular function of GPNMB in ESCC

According to the expression level of GPNMB, 274 ESCC samples in the gene expression omnibus database were divided into positive group and negative group, and the differential genes between the two groups were screened under the conditions of logFC > 1, fdr < 0.05 (Fig. 8A). In the positive group, we obtained 38 up-regulated and 23 down-regulated genes. These 61 genes were intersected with genes related to GPNMB expression in The Cancer Genome Atlas ESCC samples, and a total of 30 genes significantly related to GPNMB were obtained, which were displayed using a heat map (Fig. 8B). Subsequently, the molecular functions involved in GPNMB were explored with gene ontology analysis, and Figure 8C demonstrated its enrichment in Biological Process, Cellular Component, and Molecular Function. Kyoto Encyclopedia of Genes and Genomes pathway enrichment was performed using gene set enrichment analysis, and Figure 8D lists the top five pathways enriched by GPNMB upregulation. Finally, we used string networks to explore GPNMB-linked genes and display them (Fig. 8E).

4. Discussion

ESCC is a disease with an extremely poor prognosis. Studies have shown that the 5-year survival rate of ESCC is less than 20%.[15,16] Compared with malignant tumors such as breast cancer and lung cancer, ESCC lacks effective biomarkers to refine its classification and targeted therapy. Prognostication of ESCC using the AJCC classification system of the American Joint Committee on Cancer is widely accepted. Although the TNM system can provide a satisfactory prediction of survival, its accuracy is variable. We found that many patients diagnosed in the early stage had recurrence or metastasis in a short time, and the AJCC system could not evaluate such patients well. Therefore, adding effective biological prognostic markers on the basis of the clinical pathological data of patients has a greater application prospect for evaluating the prognosis of patients.

In our study, ITRAQ proteomics was used to screen the differential proteins in ESCC and adjacent tissues of Hazak and Han patients in Xinjiang. Among them, GPNMB is one of the proteins that are significantly up-regulated in cancer tissues of Han and Hazak patients. Studies have found that GPNMB has prognostic value in various malignant tumors such as head and neck squamous cell carcinoma, ovarian cancer, and breast cancer.[8,9,17] To explore the expression status of GPNMB in ESCC, we performed immunohistochemical staining on GPNMB in 266 ESCC tissue microarrays. To our surprise, GPNMB is not as equivocal as the expression of most molecules in tumor and paracancerous tissues, and its expression is evident in most ESCC cells but rarely in stroma, immune and esophageal squamous epithelium cells. This feature fits well with a valuable tumor therapy target. Because when it is inhibited, there is no significant loss of function to normal cells. To explore whether GPNMB has similar expression characteristics in other squamous cell carcinomas, we conducted a literature review. Biswas KB et al analyzed the expression of GPNMB in epidermal keratinocytes and believed that it was mainly expressed in basal zone cells.[18] Manevich et al[19] found that GPNMB was significantly highly expressed in laryngeal squamous cell carcinoma, while it was almost undetectable in normal epithelium. Li et al[20] studied the expression of GPNMB in head and neck squamous cell carcinoma. The results showed that GPNMB staining mainly existed in the cytoplasm of tumor cells; GPNMB staining in nasopharyngeal carcinoma cells was significantly stronger than that in normal mucosal cells; GPNMB was mainly expressed in the basal region in normal nasal mucosa. The above studies

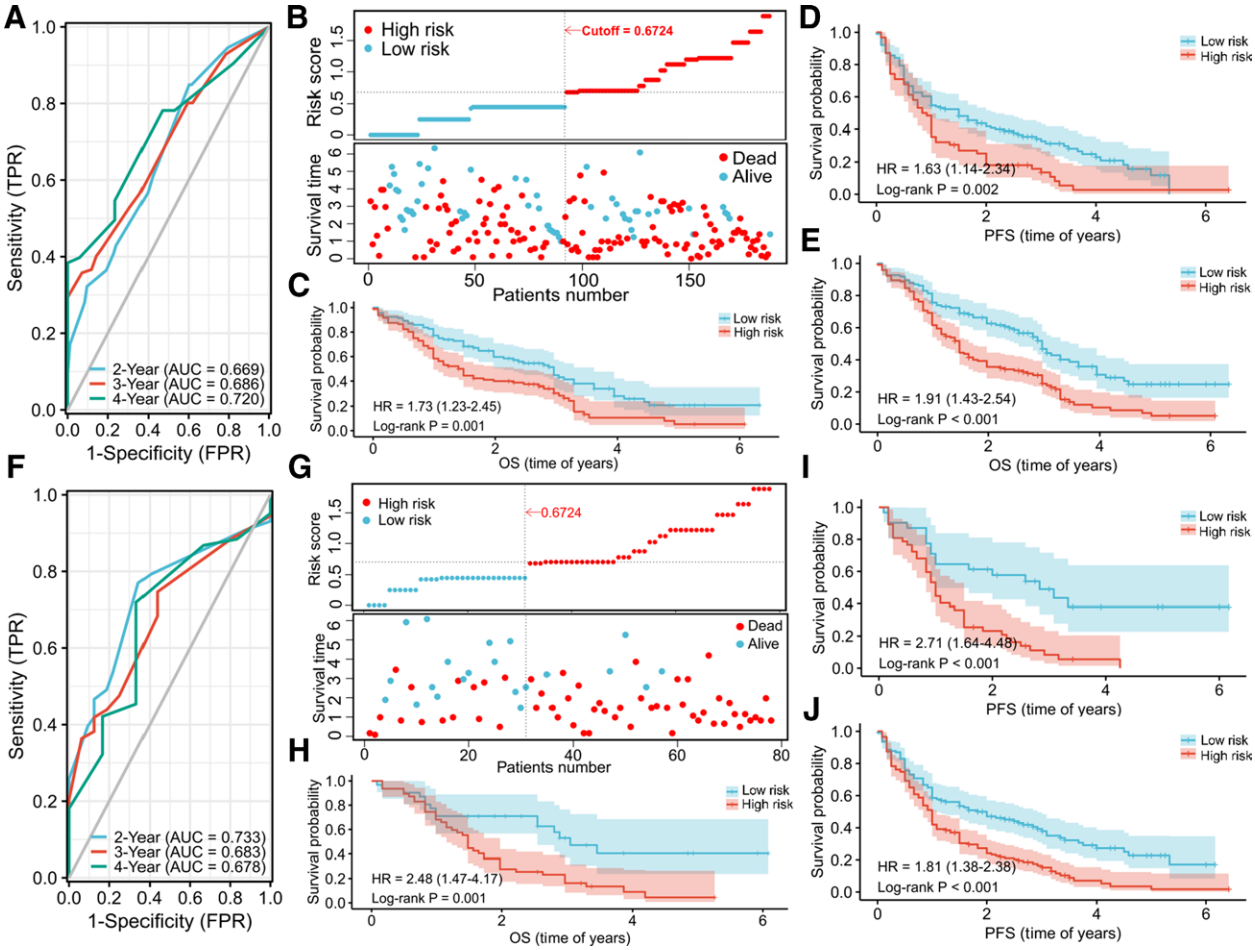

Figure 6. Model validation. A–D, training cohort (A) ROC curve based on risk score; (B) Score display of each patient; (C) OS of patients in high and low risk groups; (D) PFS of patients in high and low risk groups. F–I, validation cohort (F) ROC curve based on risk score; (G) Score display of each patient; (H) OS of patients in high and low risk groups; (I) PFS of patients in high and low risk groups. E, OS of the total study population. J, PFS of the total study population. OS = overall survival, ROC curve = receiver operating characteristic curve.

are consistent with our results. There is still a lack of research on GPNMB in cervical squamous cell carcinoma. This study analyzed the expression of GPNMB mRNA in various malignant tumors and found that its expression in cervical squamous cell carcinoma was lower than that in control tissues (Fig. 2A), which is contrary to the results of other studies. Therefore, whether the expression signature of GPNMB applies to all squamous cell carcinomas remains to be verified.

Previously, researchers have developed tools to predict prognostic risk in ESCC. For example, Wei Deng et al established a risk assessment tool for predicting OS in patients based on ESCC samples.[21] However, his assessment tool only includes traditional clinicopathological data, and lacks effective biomarkers.[11–13] However, their study only showed the prognostic characteristics of markers, and abandoned a large number of clinical case data, which is not suitable for clinical evaluation of patient prognosis. In our study, we also demonstrated the prognostic value of GPNMB in ESCC, but it is obviously not enough to evaluate the prognosis of patients only by its expression in IHC. Therefore, we further tried to add GPNMB to the traditional clinicopathological factors to improve the diagnostic performance of patient prognosis by establishing a prognostic model based on IHC + clinicopathological data. Among the multiple variables including GPNMB expression, the system selects the optimal combination of four variables to build the model. Among them, the expression of GPNMB has been considered as an important prognostic factor in other malignant tumors. In a glioma prognostic model based on multicellular gene network, GPNMB has important value as a member in evaluating the prognosis of patients.[22] The presence of nation differences in the model was unexpected because it was not a statistically significant prognostic indicator in the risk factor analysis. We argue that the penalty imposed on collinearity by the stepwise regression process when building the model causes ethnic differences to become a valid variable. Previous studies have shown that the morbidity and mortality of ESCC are higher in Hazak than in Han patients,[3,4] and the differences in lifestyles such as diet between two nation groups may have contributed to this result.[23] In the study of Huiwu Li et al, it was found that there are differentially expressed genes between Hazak and Han patients.[24] The ITRAQ results in this study also show that there are many uncrossed differential proteins between the two nation groups (Fig. 2). Therefore, ethnic differences are plausible as factors in the model. The AJCC staging system has the largest weight in the model, which also reflects that it is still an effective evaluation factor in clinical application, and also shows that the AJCC classification system cannot be missing in the ESCC prognostic model. In the past, scholars have questioned the prognostic value of nerve invasion in ESCC.[25] However, a recent study found that nerve invasion, late AJCC stage, and incomplete tumor resection were unfavorable prognostic factors for ESCC, and suggested that nerve invasion may be another way of distant metastasis besides vascular and lymphatic metastasis,[26] which is consistent with our study. By reviewing the literature, we found that most of the

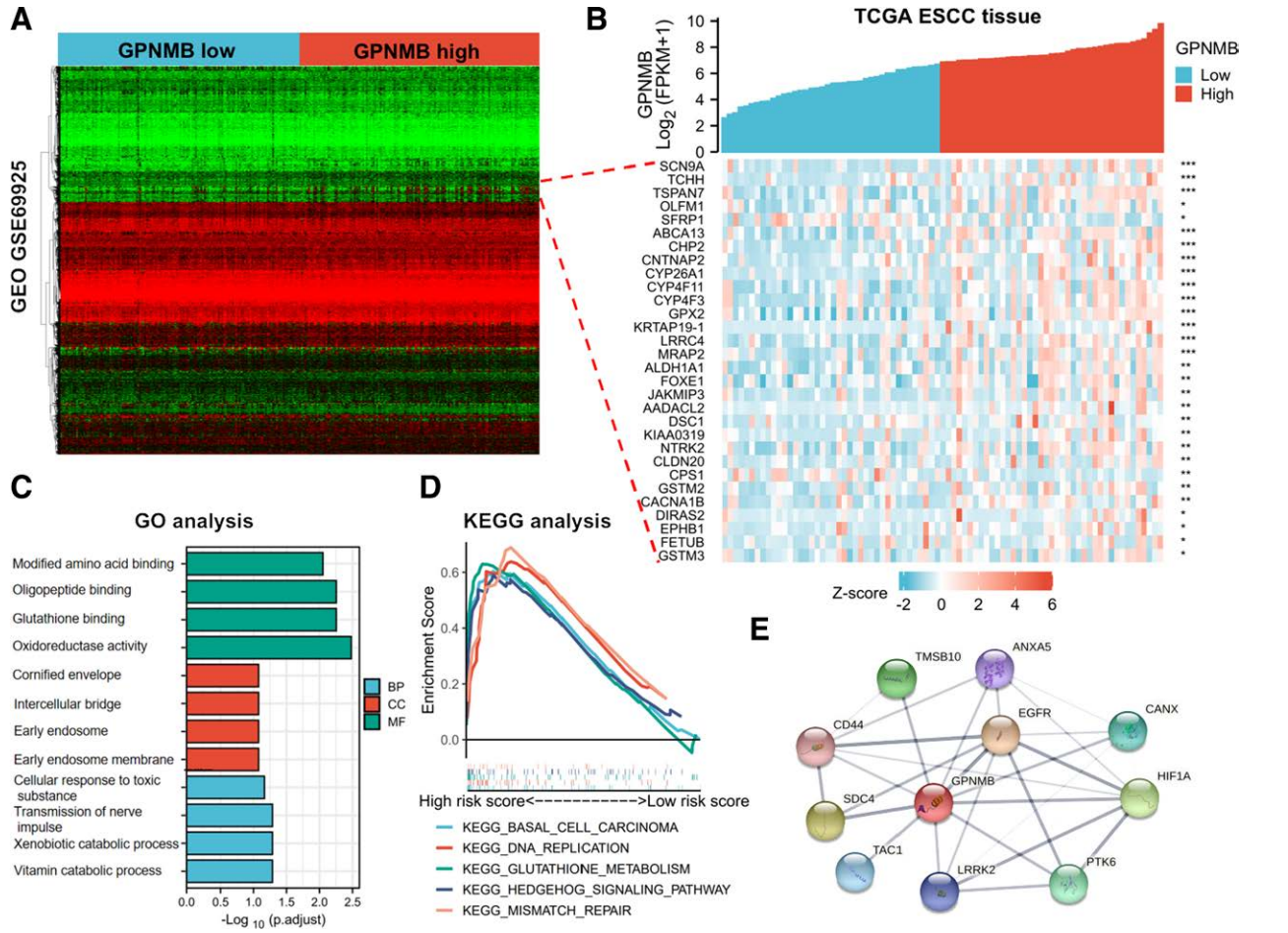

Figure 7. Relationship of prognostic model to variables. A, The distribution of each variable in the training cohort in the high and low risk groups. B, The distribution of variables in the validation cohort in the high and low risk groups. C–H, prognostic score and ROC of each variable. AJCC = American Joint Council on Cancer, AUC = area under the curve, GPNMB = glycoprotein nonmetastatic melanoma protein B, ROC curve = receiver operating characteristic curve.

prognostic models constructed based on ESCC did not obtain excellent predictive performance, which may be related to the following reasons. First, we found in the case data collection and follow-up observation that ESCC was mostly in the middle and late stages of the disease when it was diagnosed. Although after surgery and adjuvant therapy, the recurrence and metastasis time are irregular. Second, radical resection of esophageal cancer is very traumatic to patients. Although laparoscopic surgery has gradually replaced the traditional open surgery in recent years, many postoperative complications are still unavoidable for patients. Among them, the most common are esophageal anastomotic leakage, esophageal stricture, pulmonary infection and recurrent laryngeal nerve injury.[27–29] The occurrence of these complications seriously affects the quality of life of patients and may lead to non-tumor death of patients. A study showed that the median survival of patients with postoperative esophageal fistula and fistula formation was only 11 and 3.63 months, respectively.[30] Finally, treatment imbalances may also affect the accuracy of predicting patient survival. During the follow-up, we found that many patients developed the above-mentioned complications after surgery and could not receive standard radiotherapy and chemotherapy; some patients discontinued adjuvant therapy for various reasons. Therefore, under the existing treatment mode of esophageal cancer, it may be difficult to find a very satisfactory prognostic evaluation method. We demonstrated the predictive effect and stability of the model through internal validation. The patient's 2 to 4 year prognostic accuracy was predicted using the ROC curve. The 5-year survival rate for ESCC patients is poor, with more than 80% of patients dying within 5 years, and although our model predicts 5-year survival more accurately, we did not show it. As mentioned earlier, many patients with postoperative complications experience non-neoplastic deaths within 1 year, so the model estimates 1-year survival status is not stable. Although the AUC of our constructed model does not reach the excellent level above 0.8, it has surpassed the predictive power of independent factors such as AJCC.

In order to explore the molecular biological functions and pathways of GPNMB in ESCC, we used bioinformatics methods to search for genes closely related to GPNMB in The Cancer Genome Atlas and gene expression omnibus data base. The results showed that there were 30 GPNMB expressed genes in the intersection of the two databases. gene ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analysis demonstrated the biological functions and signaling pathways that GPNMB may be involved in, suggesting that GPNMB has a diverse role in ESCC, including drug resistance to chemotherapy, intercellular signaling pathway transduction, hypoxia induction and the decomposition of retinoic acid in promoting malignant differentiation of tumors.

In summary, we conducted a series of studies on the relationship between GPNMB and ESCC, and found that GPNMB not only meets the conditions as an ideal target, but also has significant prognostic value, which is sufficient to prove its effect in ESCC. By constructing a prognostic model, it is expected to provide more references for complex ESCC prognostic assessment. At present, we have not yet retrieved an ESCC prognostic

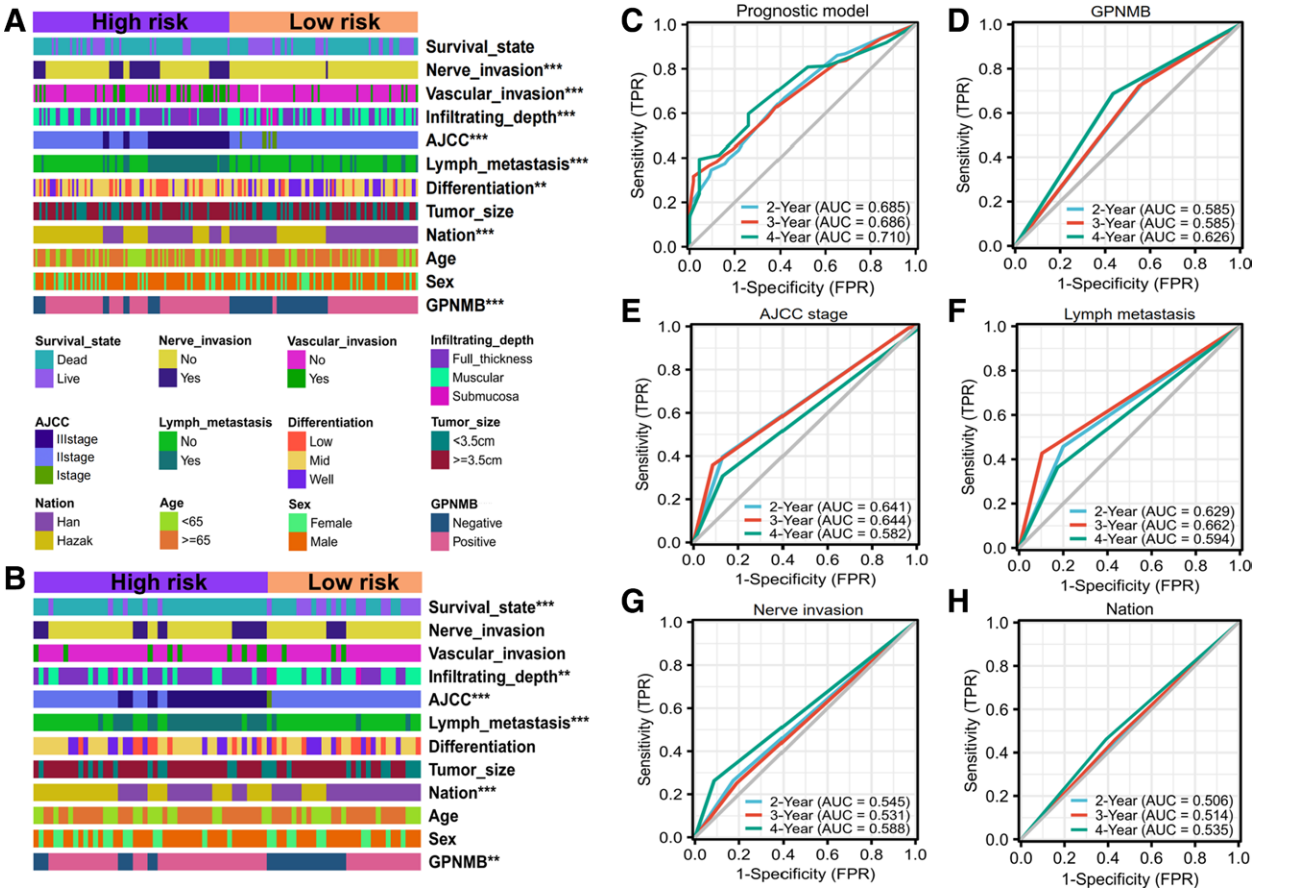

Figure 8. Molecular functional analysis of GPNMB. A, Heat map of differential genes between high and low expression groups of GPNMB in GEO database. B, Heat map of differential genes between high and low expression groups of GPNMB in TCGA database. C, GO enrichment analysis. D, KEGG pathway analysis. E, String database GPNMB-associated gene analysis. BP = biological process, CC = cellular component, ESCC = esophageal squamous cell carcinoma, GEO = gene expression omnibus data base, GO = gene ontology, GPNMB = glycoprotein nonmetastatic melanoma protein B, GSE = gene set enrichment, KEGG = Kyoto Encyclopedia of Genes and Genomes, MF = molecular function, TCGA = The Cancer Genome Atlas.

model constructed by IHC prognostic markers combined with clinicopathological parameters, which may be the first ESCC prediction model based on ICH and clinicopathological data. But our research also has some shortcomings. First, the patients we selected were Hazak and Han patients with a high incidence of ESCC in this region, and no patients from other ethnic groups were enrolled. In the established prognostic model, nation is a weighting term, and the model may not be applicable to patients with their nations. Second, we conducted an internal validation of random grouping, hoping to collect multi-center ESCC samples for validation in the future.

Author contributions

Conceptualization: Bo Wang, Yuqing Ma.
Data curation: Bo Wang, Yongmei Gao.
Formal analysis: Mengyan Li, Xuecheng Wang.
Funding acquisition: Anna Su, Mengyan Li.
Investigation: Bo Wang, Anna Su, Xuecheng Wang, Yongmei Gao.
Methodology: Anna Su.
Resources: Anna Su, Mengyan Li.
Software: Bo Wang, Liping Su.
Supervision: Mengyan Li, Xuecheng Wang, Yuqing Ma.
Validation: Mengyan Li.
Visualization: Liping Su.
Writing – original draft: Bo Wang, Anna Su.
Writing – review & editing: Bo Wang, Wan Li.

References

[1] Luo QY, Zhou SN, Pan WT, et al. A multi-kinase inhibitor APG-2449 enhances the antitumor effect of ibrutinib in esophageal squamous cell carcinoma via EGFR/FAK pathway inhibition. Biochem Pharmacol. 2021;183:114318.
[2] Yang J, Liu X, Cao S, et al. Understanding esophageal cancer: the challenges and opportunities for the next decade. Front Oncol. 2020;10:1727.
[3] Yang L, Salai A, Sun X, et al. Proteomic profiling of plasma exosomes reveals CD82 involvement in the development of esophageal squamous cell carcinoma. J Proteomics. 2022;265:104662.
[4] Wang Q, Liu D, Wang K, et al. Hypermethylation of miR-34b/c is associated with early clinical stages and tumor differentiation in Hazakh patients with esophageal squamous cell carcinoma. Int J Clin Exp Path. 2019;12:3119–27.
[5] Jin R, Jin YY, Tang YL, et al. GPNMB silencing suppresses the proliferation and metastasis of osteosarcoma cells by blocking the PI3K/Akt/mTOR signaling pathway. Oncol Rep. 2018;39:3034–40.
[6] Wang C, Okita Y, Zheng L, et al. Glycoprotein non-metastatic melanoma protein B functions with growth factor signaling to induce tumorigenesis through its serine phosphorylation. Cancer Sci. 2021;112:4187–97.
[7] Taya M, Hammes SR. Glycoprotein Non-Metastatic Melanoma Protein B (GPNMB) and cancer: a novel potential therapeutic target. Steroids. 2018;133:102–7.
[8] Li H, Xiao Y, Wu CC, et al. High expression of GPNMB predicts poor prognosis in head and neck squamous cell carcinoma. Histol Histopathol. 2019;34:803–10.
[9] Huang YH, Chu PY, Chen JL, et al. Expression pattern and prognostic impact of glycoprotein non-metastatic B (GPNMB) in triple-negative breast cancer. Sci Rep. 2021;11:12171.

[10] Han CL, Chen XR, Lan A, et al. N-glycosylated GPNMB ligand independently activates mutated EGFR signaling and promotes metastasis in NSCLC. Cancer Sci. 2021;112:1911–23.
[11] Yu J, Zheng Y, Han XP, et al. Three-gene immunohistochemical panel predicts progression and unfavorable prognosis in esophageal squamous cell carcinoma. Hum Pathol. 2019;88:7–17.
[12] Zhang Y, Li C, Chen M. Prognostic value of immunohistochemical factors in esophageal small cell carcinoma (ESCC): analysis of clinicopathologic features of 73 patients. J Thorac Dis. 2018;10:4023–31.
[13] Meng J, Zhang J, Xiu Y, et al. Prognostic value of an immunohistochemical signature in patients with esophageal squamous cell carcinoma undergoing radical esophagectomy. Mol Oncol. 2018;12:196–207.
[14] Vaudel M, Burkhart JM, Zahedi RP, et al. iTRAQ data interpretation. Methods Mol Biol. 2012;893:501–9.
[15] Codipilly DC, Qin Y, Dawsey SM, et al. Screening for esophageal squamous cell carcinoma: recent advances. Gastrointest Endosc. 2018;88:413–26.
[16] Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022. CA Cancer J Clin. 2022;72:7–33.
[17] Ma RQ, Tang ZJ, Ye X, et al. Overexpression of GPNMB predicts an unfavorable outcome of epithelial ovarian cancer. Arch Gynecol Obstet. 2018;297:1235–44.
[18] Biswas KB, Takahashi A, Mizutani Y, et al. GPNMB is expressed in human epidermal keratinocytes but disappears in the vitiligo lesional skin. Sci Rep. 2020;10:4930.
[19] Manevich L, Okita Y, Okano Y, et al. Glycoprotein NMB promotes tumor formation and malignant progression of laryngeal squamous cell carcinoma. Cancer Sci. 2022;113:3244–54.
[20] Li H, Xiao Y, Wu CC, et al. High expression of GPNMB predicts poor prognosis in head and neck squamous cell carcinoma. Histol Histopathol. 2019;34:803–10.
[21] Deng W, Zhang W, Yang J, et al. Nomogram to predict overall survival for thoracic esophageal squamous cell carcinoma patients after radical esophagectomy. Ann Surg Oncol. 2019;26:2890–8.
[22] Sun X, Liu X, Xia M, et al. Multicellular gene network analysis identifies a macrophage-related gene signature predictive of therapeutic response and prognosis of gliomas. J Transl Med. 2019;17:159.
[23] Chen W, Yang C, Yang L, et al. Association of roasting meat intake with the risk of esophageal squamous cell carcinoma of Hazakh Chinese via affecting promoter methylation of p16 gene. Asia Pac J Clin Nutr. 2014;23:488–97.
[24] Li H, Li H, Waresijiang Y, et al. Clinical significance of HDAC1, -2 and -3 expression levels in esophageal squamous cell carcinoma. Exp Ther Med. 2020;20:315–24.
[25] Chen JW, Xie JD, Ling YH, et al. The prognostic effect of perineural invasion in esophageal squamous cell carcinoma. BMC Cancer. 2014;14:313.
[26] Kim HE, Park SY, Kim H, et al. Prognostic effect of perineural invasion in surgically treated esophageal squamous cell carcinoma. Thorac Cancer. 2021;12:1605–12.
[27] Kubo N, Ohira M, Yamashita Y, et al. The impact of combined thoracoscopic and laparoscopic surgery on pulmonary complications after radical esophagectomy in patients with resectable esophageal cancer. Anticancer Res. 2014;34:2399–404.
[28] Wang X, Li X, Cheng H, et al. Single-port inflatable mediastinoscopy combined with laparoscopic-assisted small incision surgery for radical esophagectomy is an effective and safe treatment for esophageal cancer. J Gastrointest Surg. 2019;23:1533–40.
[29] Higaki E, Abe T, Fujieda H, et al. Significance of antimicrobial prophylaxis for the prevention of early-onset pneumonia after radical esophageal cancer resection: a retrospective analysis of 356 patients undergoing thoracoscopic esophagectomy. Ann Surg Oncol. 2022;29:1374–87.
[30] Guan X, Liu C, Zhou T, et al. Survival and prognostic factors of patients with esophageal fistula in advanced esophageal squamous cell carcinoma. Biosci Rep. 2020;40:BSR20193379.