ORIGINAL RESEARCH

Development and Validation of an Anoikis-Related Gene Signature for Prognostic Prediction in Cervical Cancer

Silu Meng [1,2,*], Xiangqin Li[3,*], Jianwei Zhang[1], Xiaodong Cheng [1]

[1]Department of Gynecologic Oncology, Women's Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, People's Republic of China; [2]Department of Gynecologic Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, People's Republic of China; [3]Xiangya Hospital, Zhongnan University, Changsha, Hunan, People's Republic of China

*These authors contributed equally to this work

Correspondence: Xiaodong Cheng, Email chengxd@zju.edu.cn

Background: Cervical cancer still has high incidence and mortality rates worldwide. This study aimed to evaluate the prognostic value of anoikis-related genes (ARGs) and develop a risk scoring model for accurate survival prediction in cervical cancer patients.
Methods: The expression profiles of cervical cancer tissue and survival data were downloaded from TCGA-CESC and CGCI-HTMCP-CC. We identified 83 ARGs significantly associated with patients' survival. Subsequently, we developed a risk-scoring model based on 10 key genes. We assessed the predictive performance of our model by survival analysis, ROC curve analysis, and a nomogram that incorporated clinical factors. Additionally, we validated the expression of Granzyme B (GZMB) by immunohisto-chemical staining. Furthermore, we compared the biological processes and pathway enrichment in high-risk and low-risk patient groups, using differential gene expression and functional enrichment analysis. Finally, we investigated the immune microenvironment of patients in both high-risk and low-risk groups.
Results: Patients in the high-risk group had significantly poorer survival compared to those in the low-risk group. The immunohis-tochemical results suggested that GZMB was associated with the prognosis of cervical cancer patients. The risk scoring model showed high accuracy in predicting the prognosis of cervical cancer patients. Differential gene expression analysis revealed enriched pathways related to tumor invasion and metastasis in the high-risk group. Conversely, the low-risk group showed a strong association with the activation of immune response pathways.
Conclusion: This study concluded that anoikis-related genes played a crucial role in determining the prognosis of individuals with cervical cancer. This discovery not only presented potential biomarkers but also provided valuable insights for informing treatment strategies. The risk scoring model may assist clinicians in better identifying high-risk patients and personalizing treatment plans.
Keywords: cervical cancer, anoikis-related genes, risk model, immune microenvironment, GZMB

Introduction

Cervical cancer plays a significant role in global morbidity and mortality rates among women, with 604,127 new cases and 341,831 deaths in 2020 globally.[1] Current screening methods, such as cytology, human papillomavirus testing, and colposcopy, have limitations in their specificity and sensitivity, leading to a high rate of false-negative results.[2] Therefore, novel biomarkers with improved accuracy and reliability are urgently needed to enhance the early detection, personalized treatment, and prevention of misdiagnosis and underdiagnosis of cervical cancer.[3] This research area is of great importance as it aims to improve the management and outcomes of cervical cancer patients.

Anoikis is a specialized form of programmed cell death that occurs when cells lose their attachment to the extracellular matrix (ECM) or neighboring cells.[4] This mechanism plays a critical role in maintaining cellular home-ostasis and facilitating proper tissue development. In typical physiological conditions, cells rely on these connections

with the ECM and adjacent cells to regulate their proliferation and differentiation. However, when these connections are compromised or disrupted, anoikis is initiated to ensure the preservation of normal physiological states.[5] Anoikis is primarily regulated through the interplay of two apoptotic pathways, involving the activation of cell surface death receptors and interference with mitochondrial function.[6] Notably, in certain pathological conditions, such as cancer, cells can develop resistance to anoikis. Extensive research has shown that cancer cells' resistance to anoikis enables them to detach from their original anchorage-dependent growth, migrate through the lymphatic and circulatory systems to distant sites, and continue to grow and proliferate.[7–9] Consequently, Anoikis resistance is considered a critical initial step in tumor invasion and metastasis. Furthermore, this resistance can reconfigure the tumor microenvironment, facilitating immune evasion and rendering the tumor cells resistant to chemotherapy.[9–11] Nonetheless, the specific mechanisms underlying the role of anoikis-related genes (ARGs) and pathways in cervical cancer remain poorly understood. Therefore, there is a significant need to identify cervical cancer-specific anoikis-related biomarkers, investigate their underlying mechanisms, and explore their implications for prognosis. This endeavor has substantial clinical relevance and with the potential to significantly enhance the early diagnosis and personalized treatment of cervical cancer.

The study aimed to identify and validate key ARGs that are associated with prognosis in cervical cancer based on public datasets from The Cancer Genome Atlas (TCGA) and the Cancer Genome Characterization Initiative (CGCI). We also developed a risk score model to facilitate individualized prediction of survival outcomes. The main objective was to identify novel biomarkers and methodological foundations for personalized screening, targeted therapy, and prognostic assessment for cervical cancer.

Materials and Methods
Data Acquisition and Organization
We obtained the gene expression data (RNA-seq) for TCGA cervical squamous cell carcinoma and endocervical adenocarcinoma (TCGA-CESC) and CGCI HIV+ Tumor Molecular Characterization Project (CGCI-HTMCP-CC) from the GDC database maintained by the National Cancer Institute, with pertinent clinical and survival data (https://portal.gdc.cancer.gov/repository as of October 14, 2023). Table 1 provides comprehensive clinical details for bothcohorts. To harmonize the expression data from these two datasets, we applied the R package 'sva' to mitigate batch effects inherent in different cohorts. Furthermore, we retrieved 700 ARGs from the Genecards (https://www.genecards.org/, as of October 2, 2023) and the Harmonizome portal (https://maayanlab.cloud/Harmonizome/, as of October 2, 2023). The whole-exome sequencing (WES) and copy number data were also downloaded from the TCGA database. The expression array datasets (GSE9750 and GSE52903) were downloaded from Gene Expression Omnibus (GEO) database. The immunohistochemistry result was retrieved from Human Protein Atlas (HPA) database.

Identification of Anoikis-Related Patterns
In this study, we initially screened 503 candidate ARGs using univariate Cox regression, a standard method for preliminary prognostic biomarker identification.[12,13] Although univariate Cox analysis cannot account for gene interactions, it is an efficient initial screening tool to identify candidate genes significantly associated with survival ($p < 0.05$), reducing dimensionality for subsequent modeling. The primary purpose was to establish a prognostically relevant gene set as input for LASSO regression analysis and further exploration. This initial screening identified 83 prognostically relevant ARGs from the RNA-Seq data. We used the expression profiles of these ARGs to investigate potential anoikis-related molecular subtypes using the R package 'ConsensusClusterPlus'. The optimal number of clusters was determined using a consistency matrix. Principal Component Analysis (PCA) was employed to assess subtype stability, and survival differences were evaluated using Kaplan-Meier analysis. An expression heatmap was also generated to illustrate differences between the subtypes.

Construction and Evaluation of the Anoikis-Related Prognostic Signature
To establish and appraise prognostic markers associated with anoikis, we randomly allocated 385 patients into a training set (n=231) and a validation set (n=154), employing the R package 'caret'. In the training set, the Least Absolute

Table 1 Clinical Characteristics of TCGA-CESC and CGCI-HTMCP-CC Datasets

Characteristics		TCGA-CESC (N=273)	CGCI-HTMCP-CC (N=112)	Total (N=385)
Age (years)	Mean (SD)	47.9 (13.7)	46.9 (10.9)	47.6 (13.0)
Grade	G1	16 (6.5%)	6 (5.6%)	22 (7.8%)
	G2	126 (51.0%)	54 (50.0%)	180 (63.6%)
	G3	104 (42.1%)	48 (44.4%)	152 (53.7%)
	G4	1 (0.4%)	0 (0%)	1 (0.4%)
Stage	I	149 (55.8%)	18 (16.1%)	167 (44.1%)
	II	60 (22.5%)	54 (48.2%)	114 (30.1%)
	III	38 (14.2%)	37 (33.0%)	75 (19.8%)
	IV	20 (7.5%)	3 (2.7%)	23 (6.1%)
T	T1	128 (58.7%)	18 (16.1%)	146 (44.2%)
	T2	64 (29.4%)	55 (49.1%)	119 (36.1%)
	T3	16 (7.3%)	36 (32.1%)	52 (15.8%)
	T4	10 (4.6%)	3 (2.7%)	13 (3.9%)
N	N0	118 (69.0%)	21 (67.7%)	139 (68.6%)
	N1	53 (31.0%)	10 (32.3%)	63 (31.2%)
M	M0	101 (91.0%)	27 (96.4%)	128 (92.1%)
	M1	10 (9.0%)	1 (3.6%)	11 (7.9%)
Status	Alive	203 (74.4%)	43 (38.4%)	246 (63.9%)
	Dead	70 (25.6%)	69 (61.6%)	139 (36.1%)
Survival Time (years)	Mean (SD)	3.15 (3.16)	1.20 (0.674)	2.58 (2.83)

Shrinkage and Selection Operator (LASSO) was employed on the 83 prognostically relevant ARGs to select key predictors, balancing model robustness and overfitting control. The 'glmnet' package was used, and the optimal regularization parameter (lambda) was determined via 10-fold cross-validation based on the minimum partial likelihood deviance. This process identified 10 ARGs with non-zero coefficients. The risk score calculation formula is delineated as follows:

$$\text{Risk Score} = \sum_{1}^{n} \text{coef}_i * x_i$$

The specific formula derived was: risk score = (0.098 * SPP1) + (0.116 * CXCL8) + (−0.239 * TP73) + (−0.280 * GZMB) + (0.310 * ITGA5) + (0.314 * ENDOG) + (−0.356 * CSK) + (−0.377 * MAP2K2) + (−0.397 * PRKACA) + (0.538 * KDM3A). Patients in both the training and validation sets were categorized into high-risk and low-risk groups based on the median risk score calculated in the training set. The prognostic accuracy of the risk score was evaluated using Kaplan-Meier survival curves (Log rank test) and time-dependent receiver-operating characteristic (ROC) curve analysis at 3 and 5 years in both sets.

Comparative Analysis of Prognostic Models and Selection of the Final Model

To rigorously validate the reliability of the model constructed using our selected methodology, we conducted a comparative analysis. This analysis utilized a 10-fold cross-validation strategy, repeated 10 times, on the combined TCGA-CESC and CGCI-HTMCP-CC dataset (n=385 after filtering). Crucially, all models compared were trained and evaluated using the expression data of the same 10 ARGs (SPP1, CXCL8, TP73, GZMB, ITGA5, ENDOG, CSK, MAP2K2, PRKACA, KDM3A) identified through our initial feature selection process described before. This ensured a direct comparison of the algorithms' performance based on the final 10 gene signature. The following models were evaluated:

Cox Proportional Hazards Model: Trained using the 'survival::coxph' function in R. This standard survival analysis method estimates the hazard ratio assuming proportional hazards over time.[14]

Random Survival Forest (RSF): An ensemble method based on survival trees, implemented using the 'randomForestSRC' package.[15] We used 100 trees (ntree=100) per forest. RSF can capture non-linear effects and interactions without the proportional hazards assumption.

XGBoost Survival: A gradient boosting machine algorithm adapted for survival analysis using the Cox partial likelihood objective function, implemented with the 'xgboost' package.[16] We used 100 boosting rounds (nrounds=100).

Gradient Boosting Machine (GBM) Cox: Another gradient boosting approach using the Cox partial likelihood, implemented with the 'gbm' package.[17] We used 1000 trees (n.trees=1000) and an interaction depth of 3.

Model performance was assessed in each fold of the cross-validation using time-dependent Area Under the Curve (tdAUC) at 1, 3, and 5 years (calculated using the timeROC package) on both the training and testing sets. The average performance across all repeats and folds was calculated.

Based on this comparative analysis, considering model robustness (similarity between training and test performance), predictive accuracy, and interpretability, the LASSO-Cox approach (used for initial feature selection) leading to the 10-gene Cox model was chosen as the final prognostic model for subsequent analyses and nomogram construction.

Immunohistochemistry

To evaluate the prognostic function of Granzyme B (GZMB), we used the tissue microarray (TMA) obtained from Shanghai Outdo Biotech Co., Ltd. (Shanghai, China). It consisted of 126 cervical cancer tissues and 42 pair-matched normal samples, and the corresponding clinical information is available online. The TMA slide was dried at 60°C for 1 hour, dewaxed in xylene, and dehydrated in a gradient ethanol series. Antigen retrieval was performed by heating the slide in a microwave oven in a container filled with EDTA antigen retrieval buffer (pH 9.0). Subsequently, the slide was immersed in 3% hydrogen peroxide for 30 minutes to inhibit the activity of endogenous peroxides. Next, the TMA tissue was incubated with 10% goat serum (Boster Biological Technology Co., Ltd., Wuhan, China) and blocked at room temperature for 1 hour. Then, the TMA slide was incubated with anti-GZMB (1: 100 dilution, 13588-1-AP, Proteintech Group, Wuhan, China) for 2 hours at room temperature. The tissue was incubated at room temperature for 50 minutes with horseradish peroxidase (HRP)-labeled goat anti-rabbit secondary antibody (1:200 dilution, GB23303, ServiceBio, Wuhan, China). Then, the tissue was stained with diaminobenzidine (DAB) for 90 seconds. Finally, the slide was counterstained with hematoxylin, dehydrated, and mounted. To evaluate IHC staining, semi-quantitative scoring criteria were used. The stained slide was scored by 2 observers using a semi-quantitative hybrid (H) scoring method: H-score = (% cells of 0 intensity) + (% cells of 1+ intensity × 1) + (% cells of 2 + intensity × 2) + (% cells of 3 + intensity × 3) + (% cell of 4 + intensity × 4). The staining intensity was classified as follows: 0 (negative), 1 (weak), 2 (moderate), or 3 (strong). The staining area was defined as the distribution of positively stained tumor cells.

Clinical Factor Stratification Analysis and Development and Validation of Nomogram

We conducted an analysis to assess the correlation between the risk score and various clinical characteristics, including patient age, grade, TNM and FIGO stage. For analyses involving clinical variables (age, grade, T stage, FIGO stage), patients with missing data in any of these variables were excluded from the corresponding analysis. Additionally, we performed stratified survival analysis based on tumor stage (T and FIGO) to evaluate differences in survival between high- and low-risk groups. To enhance the clinical applicability of the risk score, we integrated it with clinical features to develop a nomogram for predicting 3- and 5-year survival in patients with cervical cancer. The accuracy of the nomogram was assessed through ROC curves, C-index, calibration curves, and DCA curves.

Identification of Differentially Expressed Genes and Functional Enrichment Analysis

Differentially expressed genes (DEGs) between high- and low-risk subgroups were identified using the 'limma' R package, with thresholds of adjusted p-value<0.05 and |logFC| > 1. We performed GO, KEGG, and GSEA analyses using the R package 'clusterProfiler' to elucidate the functions and pathways associated with DEGs concerning cervical cancer prognosis.

Tumor Immune Microenvironment Assessment in Risk Subgroups

Using the ESTIMATE algorithm, we calculated immune, stromal, and tumor purity scores for each patient. Single-sample GSEA (ssGSEA) was performed on immune gene sets with the 'GSVA' R package. The CIBERSORT algorithm was employed to determine the differential abundance of 22 tumor-infiltrating immune cells (TICs) in both high and low-risk subgroups.

Statistical Analysis

All statistical analyses were performed using R software (version 4.3.1). We conducted univariate and multivariate Cox regression analyses using the 'survminer' package and assessed the survival curves via the Kaplan-Meier Log rank test. The nomogram was created using the 'rms' package. Decision curve analysis was performed using the 'ggDCA' package. For data visualization, we used tools such as 'ggplot2', 'ggpubr', and 'pheatmap'. Group comparisons were made using t-tests or Mann–Whitney U-tests, and differences within groups were assessed through one-way ANOVA. Statistically significant was defined as a p-value < 0.05 (*p < 0.05; **p < 0.01; ***p < 0.001).

Results

Identification of Anoikis-Related Genes with Prognostic Value

The study's workflow is illustrated in Supplementary Figure 1. Initially, we obtained data from the TCGA-CESC and CGCI-HTMCP-CC cohorts from the GDC database, encompassing a total of 27,733 genes. To investigate the connection between ARGs and survival, we further analyzed these genes by intersecting them with the acquired ARGs, ultimately isolating 503 ARGs with expression data. Through univariate Cox regression analysis, we pinpointed 83 statistically significant (p <0.05) ARGs that exerted a substantial impact on survival prognosis (Supplementary Table 1). To explore the biological functions represented by the prognostically significant ARGs, we performed GO and KEGG functional enrichment analyses on the 83 genes significantly associated with prognosis in univariate Cox regression analysis. Results showed that these genes were significantly enriched in biological processes and signaling pathways closely related to tumor progression, including apoptosis regulation, extracellular matrix remodeling, immune response, and angiogenesis (Supplementary Tables 2 and 3). These functional enrichment results support the biological rationale of our prognostic model based on ARGs.

Discovery of Anoikis Subtypes in Cervical Cancer

To investigate whether anoikis-related patterns could be employed for molecular subtyping of cervical cancer patients, we conducted unsupervised cluster analysis based on the expression profiles of the 83 prognostically relevant ARGs. Our findings suggested that k = 2 was the optimal number of clusters (Figure 1A), indicating the categorization of cervical cancer patients into two distinct anoikis-related subtypes characterized by ARGs, namely subtype A (n = 194) and subtype B (n = 233). Subsequent PCA analysis revealed clear differentiation in the distribution of patients between these two subtypes (Figure 1B). Additionally, we generated expression heatmaps to depict the ARG expression in these subtypes (Figure 1C). Kaplan-Meier survival curves demonstrated a significantly lower survival rate (p<0.001) for subtype A compared to subtype B (Figure 1D).

Prognostic Risk Model Construction and Validation

In order to pinpoint the most robust prognostic predictive features, we executed LASSO and multivariate Cox analyses using anoikis gene expression data from the training set (n=232) for screening prognostically relevant ARGs (Figure 2A and B). Consequently, we identified 10 genes (SPP1, CXCL8, TP73, GZMB, ITGA5, ENDOG, CSK, MAP2K2, PRKACA, KDM3A) for the development of prognostic signature (Figure 2C). The risk score was computed using the formula: risk score = (0.098* SPP1) + (0.116* CXCL8) + (−0.239* TP73) + (−0.280* GZMB) + (0.310* ITGA5) + (0.314* ENDOG) + (−0.356* CSK) + (−0.377* MAP2K2) + (−0.397* PRKACA) + (0.538* KDM3A). We further incorporated WES data from the TCGA database and analyzed mutations and copy number alterations of these genes. Among 304 cervical cancer samples, 25 (8.2%) carried mutations of ARGs, with the highest mutational frequency observed in KDM3A (3.3%) (Supplementary
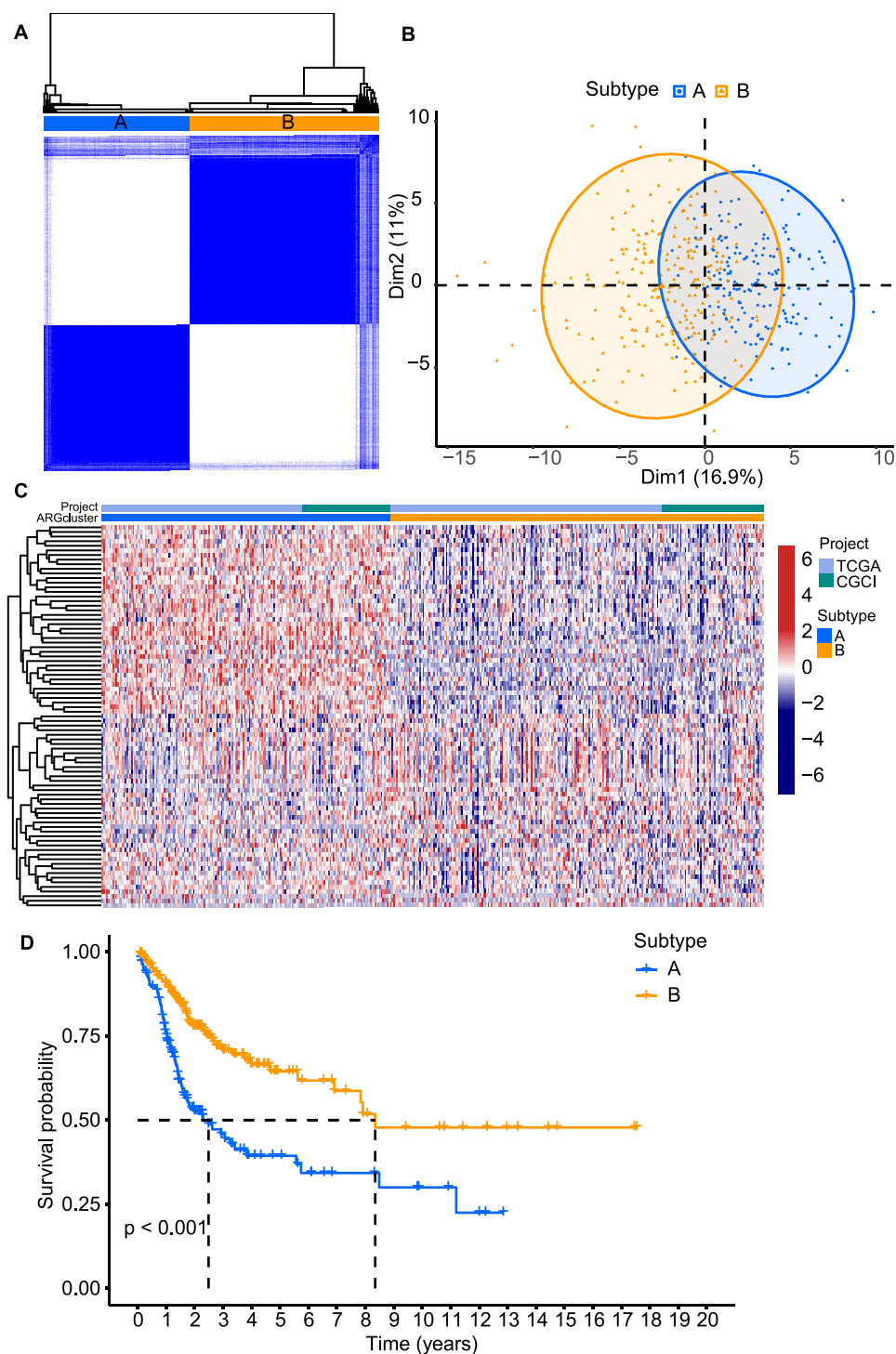
Figure 1 Identification and prognostic significance of anoikis-related subtypes. (**A**) Consensus clustering analysis identified two clusters (k = 2). (**B**) PCA analysis of distinct anoikis subtypes. (**C**) Heatmap of the ARGs expression in the two subtypes. (**D**) Survival analysis of distinct anoikis subtypes.

Figure 2A). Additionally, exploring CNV alteration frequency determined a higher incidence of copy number amplification in the CSK, CXCL8, and PRKACA, whereas MAP2K2 and TP73 exhibited a greater frequency of copy number loss (Supplementary Figure 2B).

Subsequently, we categorized patients into a high-risk group (n=115) and a low-risk group (n=116) based on the median risk score from the training set (Figure 2D and Supplementary Table 4). The results from the survival analysis
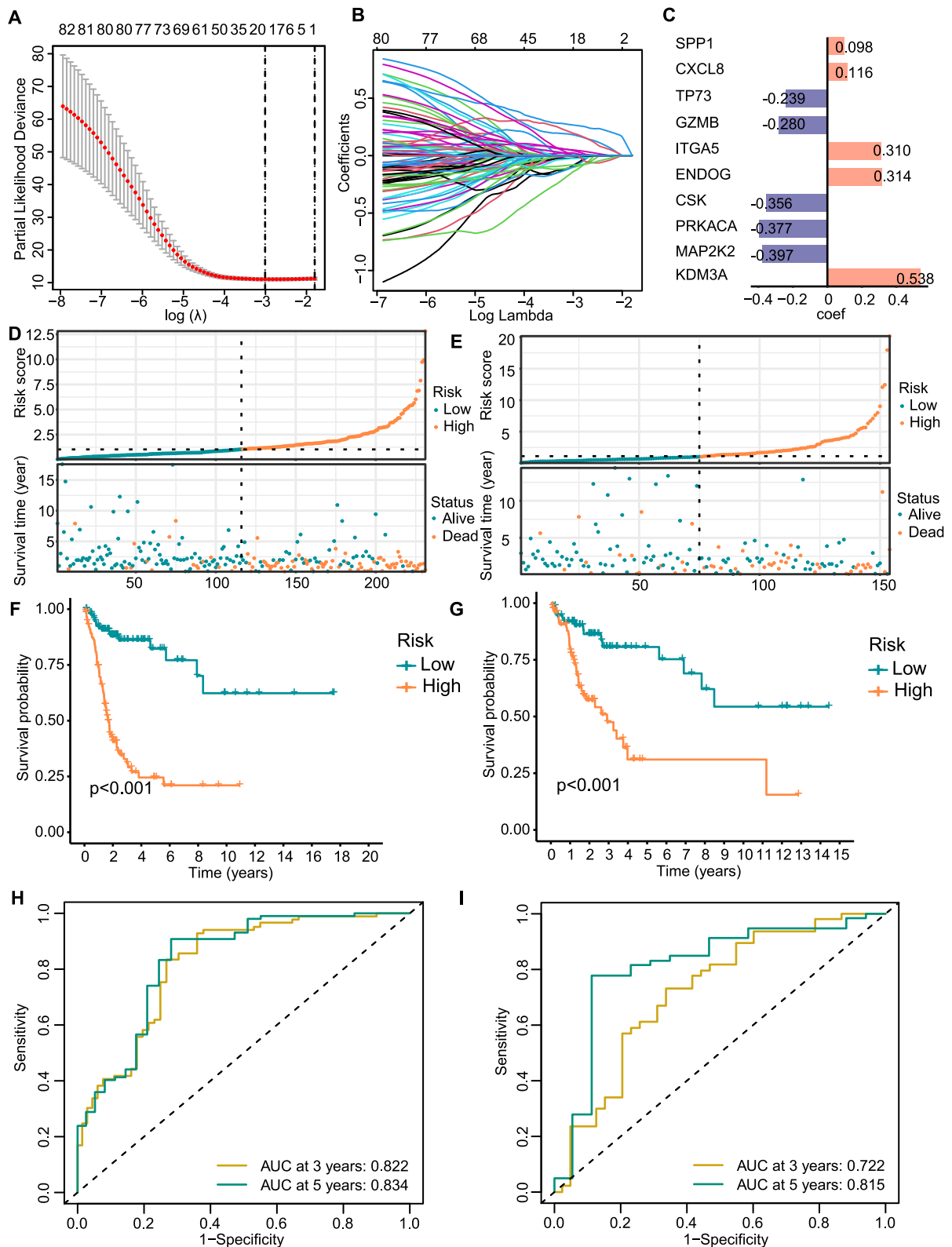
Figure 2 The establishment of ARGs risk signature. (**A** and **B**) LASSO analysis to screen candidate genes. (**C**) Coefficient of selected 10 ARGs for risk signature. Distribution of risk scores and patient status in (**D**) training cohort and (**E**) validation cohort. Survival curves for the (**F**) training cohort and (**G**) validation cohort. Time dependent ROC curve of the risk model in (**H**) training cohort and (**I**) validation cohort.

demonstrated that the high-risk group had significantly shorter survival times compared to the low-risk group (p<0.001). The Kaplan-Meier curves confirmed these results (Figure 2F). ROC curve analysis revealed that the risk signature of anoikis exhibited AUC values of 0.822 and 0.834 at 3 and 5 years, respectively (Figure 2H).

In the validation set (n=154), the same risk score formula was employed to calculate risk scores, and patients were categorized into high-risk (65 cases) and low-risk (89 cases) groups based on the median risk scores from the training set (Figure 2E and Supplementary Table 4). Consistent with the training set results, patients in the high-risk group exhibited significantly worse prognoses than those in the low-risk group (p<0.001) (Figure 2G). ROC curve analysis demonstrated that the risk profile's performance in the validation set was also excellent, with AUC values of 0.722 and 0.815 at 3 and 5 years, respectively (Figure 2I), establishing it as a valuable prognostic indicator.

Immunohistochemical staining was performed to experimentally verify the expression of GZMB, and representative staining fields were shown in Figure 3A–D. The results showed that GZMB was significantly overexpressed in cervical cancer compared to normal tissue (Figure 3E). High GZMB expression was defined as an H-score greater than the median. We conducted stratified survival analysis of GZMB expression and demonstrated the protective prognostic value of GZMB, as the samples with high GZMB expression had significantly better survival outcomes than the samples with low GZMB expression (Figure 3F–G). Additionally, we preliminarily validated the expression of these genes in cervical cancer using two expression profile chips (GES9750 and GSE52903) and the HPA database. The results showed that these genes were either upregulated or exhibited an upward trend in cervical cancer (Supplementary Figure 3).
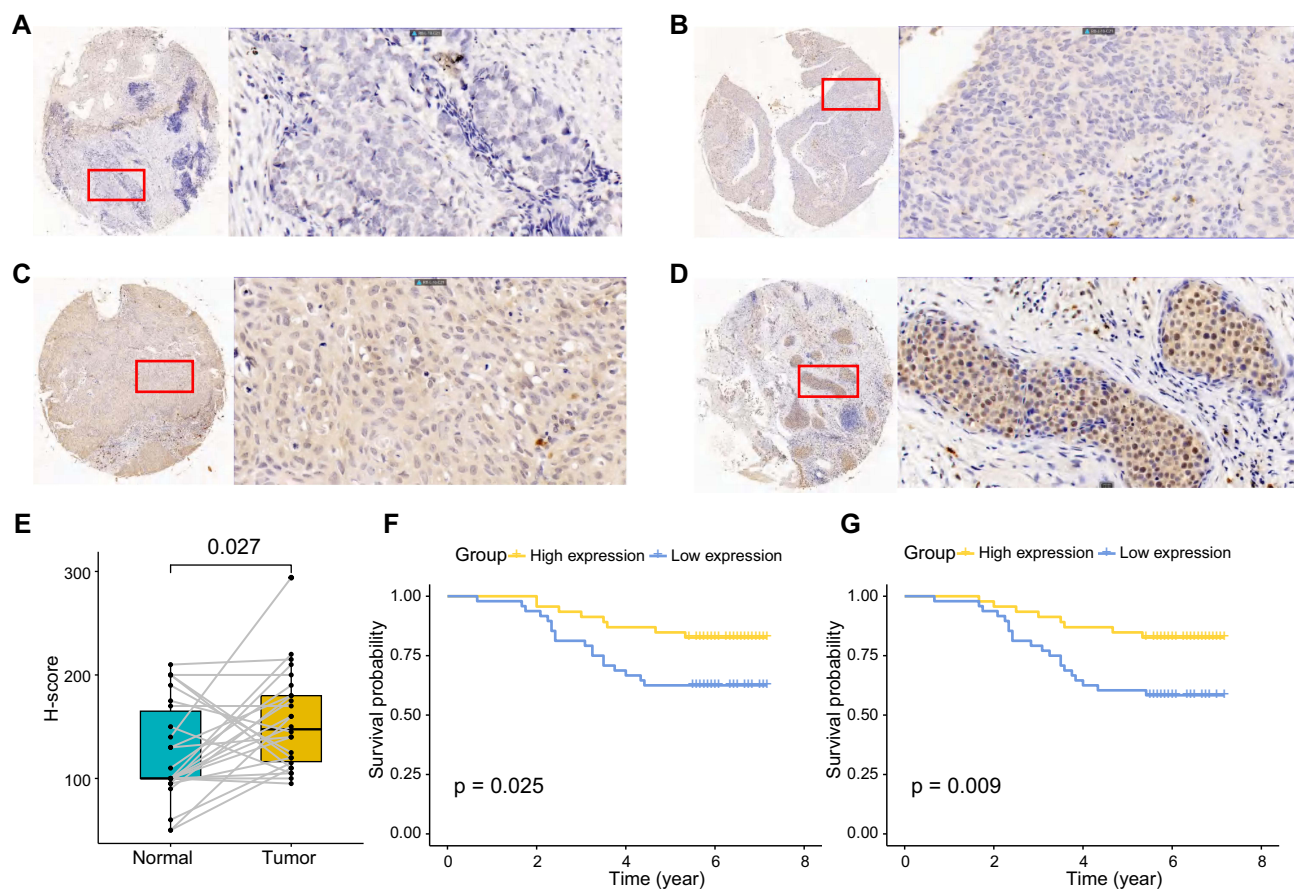

Figure 3 Experimental validation of immunohistochemical staining of GZMB in TMA cohort and its prognostic value. (**A–D**) The intensities of GZMB immunostaining were (**A**) negative, (**B**) weakly positive, (**C**) moderately positive and (**D**) strong positive in cervical cancer tissues. Left image, magnification ×50; right image, magnification ×400; the red squares indicate the area shown in higher magnification. (**E**) The H-score of GZMB in cervical cancer and adjacent normal tissues. (**F** and **G**) Kaplan–Meier survival curves to compare (**F**) OS and (**G**) DFS between the high GZMB expression and low GZMB expression patients.

Comparative Analysis of Prognostic Models

To evaluate the performance of the identified 10-gene signature using different modeling algorithms, we compared its performance against several machine learning algorithms (RSF, XGBoost, GBM) using repeated 10-fold cross-validation, with all models trained on the same 10-gene signature. The average performance metrics are summarized in Supplementary Table 5. While some machine learning models like XGBoost showed high AUC values on the training data (eg, 5-year AUC = 1.00), they exhibited a noticeable drop in performance on the validation data (eg, 5-year Test AUC ≈ 0.66), suggesting potential overfitting within this dataset. The Cox model demonstrated more stable performance across training and validation sets (eg, 5-year Train AUC ≈ 0.824, 5-year Test AUC ≈ 0.803), indicating better generalization. These results support the selection of the 10-gene Cox model for its balance of predictive accuracy, robustness, and interpretability in this context.

Stratified Analysis of Clinical Factors

We conducted a correlation analysis between risk scores and clinical characteristics (Figures 4A–F). Notably, we found that none of the differences in risk scores were statistically significant for patients of different ages, grade, N and M stage. However, patients with advanced T stage (T3-T4) and FIGO stage exhibited higher risk scores, indicating that our risk score model might encompass key information related to the biological behavior of cervical cancer. Subsequently, we performed stratified risk analysis, and the Kaplan-Meier results demonstrated that the risk signature exhibited excellent predictive capability in advanced T stage (T3-T4) and FIGO stages (p < 0.001), underscoring its reliability and clinical applicability as an independent risk factor (Figure 4G–J).

Establishment of Risk Nomogram

Both univariate Cox regression analysis and multivariate Cox regression analysis demonstrated that risk score, T stage and FIGO stage were all significantly linked to overall survival (Figure 5A and B). The findings strongly support the independent prognostic value of the risk signature in cervical cancer patients. Subsequently, we generated a nomogram illustrating the relationship between grade, age, T stage, FIGO stage and risk score in relation to the calculation of survival probability (Figure 5C). The ROC curve analysis indicated that the area under the ROC was 0.819 and 0.834 at 3 and 5 years, respectively, underscoring its strong prognostic validity (Figure 5D). Additionally, the C-index of the nomogram was 0.781. Calibration curves demonstrated significant agreement between the actual and predicted survival probabilities at 3 and 5 years (Figure 5E). Moreover, the results from the decision curve analysis at 3 and 5 years highlighted the superior predictive performance of the nomogram compared to other independent clinical factors (Figure 5F and G).

We also compared the performance of our ARG model with several previously published gene prognostic models for cervical cancer. Du et al developed an ferroptosis-based model with an AUC of 0.78 (5-year).[18] Shi et al's autophagy-based model demonstrated lower predictive ability with test AUCs of 0.603 (5-year).[19] Wang et al's immunology-based signature reported an AUC of 0.815.[20] Our ARG model (AUC = 0.83) demonstrates competitive predictive performance compared to these benchmarks. However, we acknowledge that this comparison is indirect due to differences in patient cohorts, data types, and validation methodologies across studies. Further validation in independent external cohorts is needed to definitively establish the superiority of our model.

Differentially Expressed Genes and Functional Enrichment Analysis

To unveil differences in biological functions, we compared gene expression between patients in the high-risk and low-risk groups, ultimately identifying 193 differentially expressed genes (DEGs), with detailed information available in Supplementary Table 6. The GO enrichment results demonstrated that these DEGs are highly involved in immune system activation, production of immunoglobulins and immune mediators, antigen recognition and presentation, and other immune response processes (Figure 6A and Supplementary Table 7). The enriched KEGG pathways demonstrated that the differentially expressed genes are strongly correlated with activations of immune cell development and differentiation, autoimmunity, inflammatory cytokine production, pathogen recognition, cell adhesion and migration, antigen
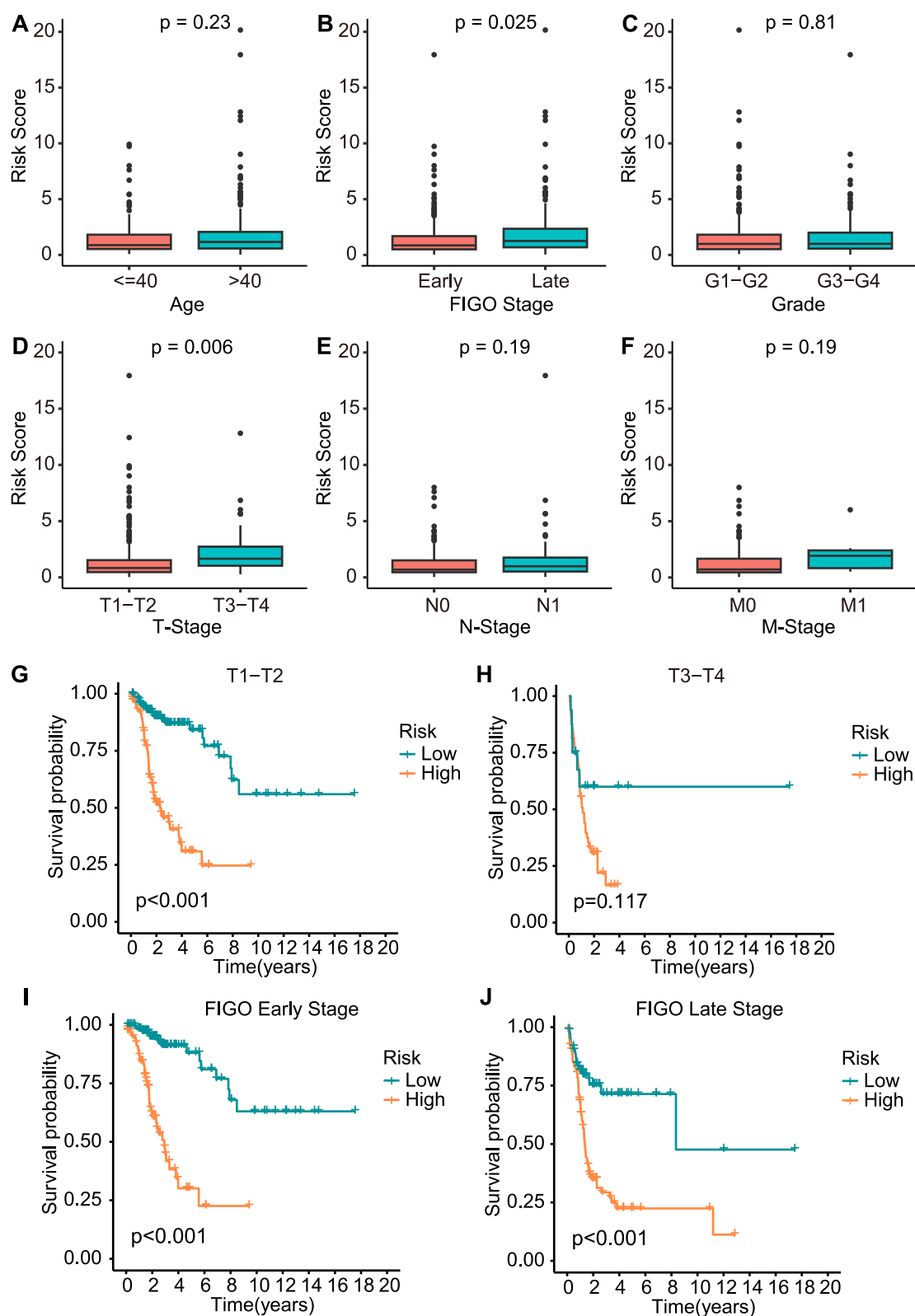
Figure 4 Relationship between ARGs risk scores and clinical characteristics. (**A–F**) Differences in risk scores of patients with regard to age, FIGO stage, grade, T, N and M stage. (**G** and **H**) Survival analysis of patients with OS regrouped according to T stage. (**I** and **J**) Survival analysis of patients with OS regrouped according to FIGO stage.
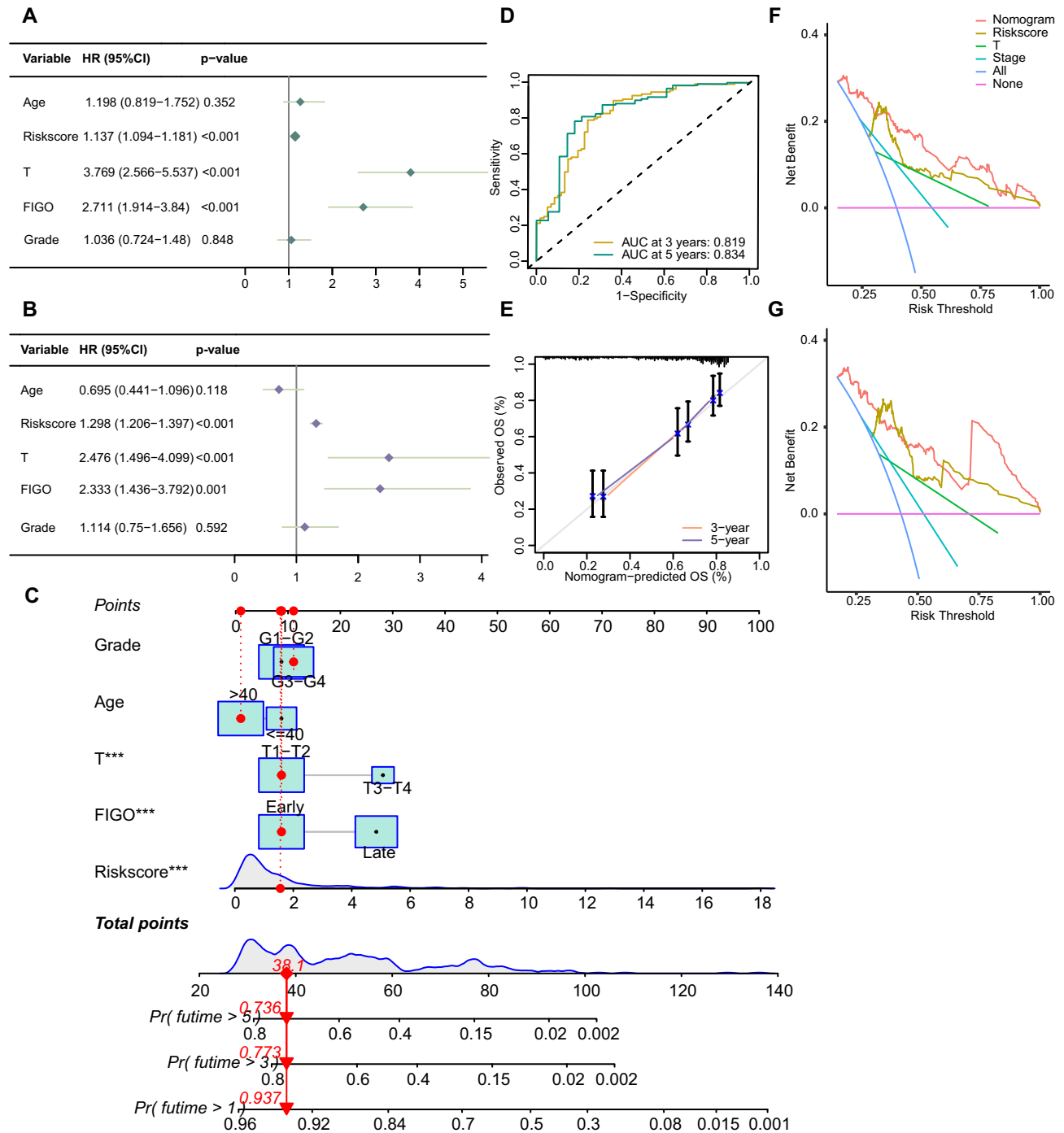
Figure 5 Establishment and evaluation of a predictive nomogram. (**A** and **B**) Results of the univariate Cox regression analysis and multivariate Cox regression analysis regarding OS of the risk signature. (**C**) The nomogram based on ARGs score and clinical characteristics to predict the prognosis at 3 and 5 years. (**D**) The ROC curves of the nomogram. (**E**) The calibration curve for evaluating the accuracy of the nomogram model. The grey dashed diagonal line represents the ideal nomogram. (**F** and **G**) The DCA curves of the nomogram at 3 and 5 years.

presentation, and infections (Figure 6B and Supplementary Table 8). The GSEA analysis based on Hallmark gene sets demonstrated increased immune responses similar to transplant rejection, activations of interferon signaling, enhanced epithelial-mesenchymal transition, and upregulated UV DNA damage responses in the samples (Figure 6C and
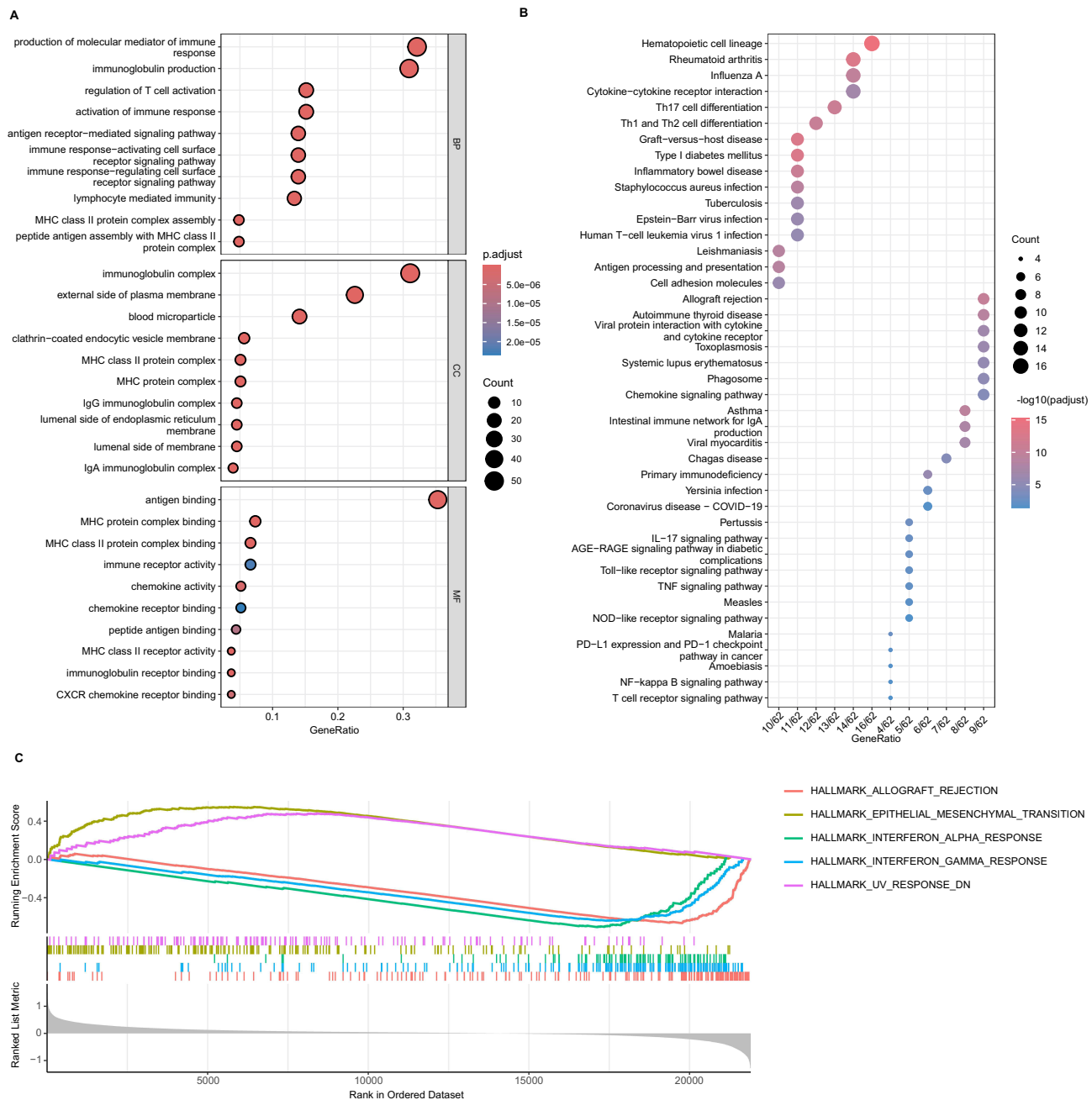
Figure 6 Biological functional and pathway enrichment analysis of high-risk group and low-risk group based on the risk signature. (**A**) Top 30 most enriched GO terms according to differential genes between high- and low- risk populations. (**B**) Top 30 most enriched KEGG pathways of the common differently genes between high- and low-risk populations (**C**) GSEA enrichment analysis based on hallmark gene set showing the activation states of biological pathways in high- and low-risk populations.

Supplementary Table 9). These processes have established associations with immune inflammation and cancer progression. The results provide insights into the biological mechanisms represented by the differential gene expression profile.

Immunosuppressive Microenvironment in the High-Risk Group

To explore the distinctions in the tumor immune microenvironment between the high-risk and low-risk groups, we utilized the ESTIMATE algorithm to assess stromal scores, immune scores, and tumor purity in cervical cancer patients (Figure 7A). The results revealed significant differences in the tumor microenvironment characteristics between the two groups. Specifically, the low-risk group exhibited significantly higher Immune Scores and Stromal Scores (indicating greater non-tumor cell infiltration) compared to the high-risk group, while tumor purity showed the inverse trend. Given
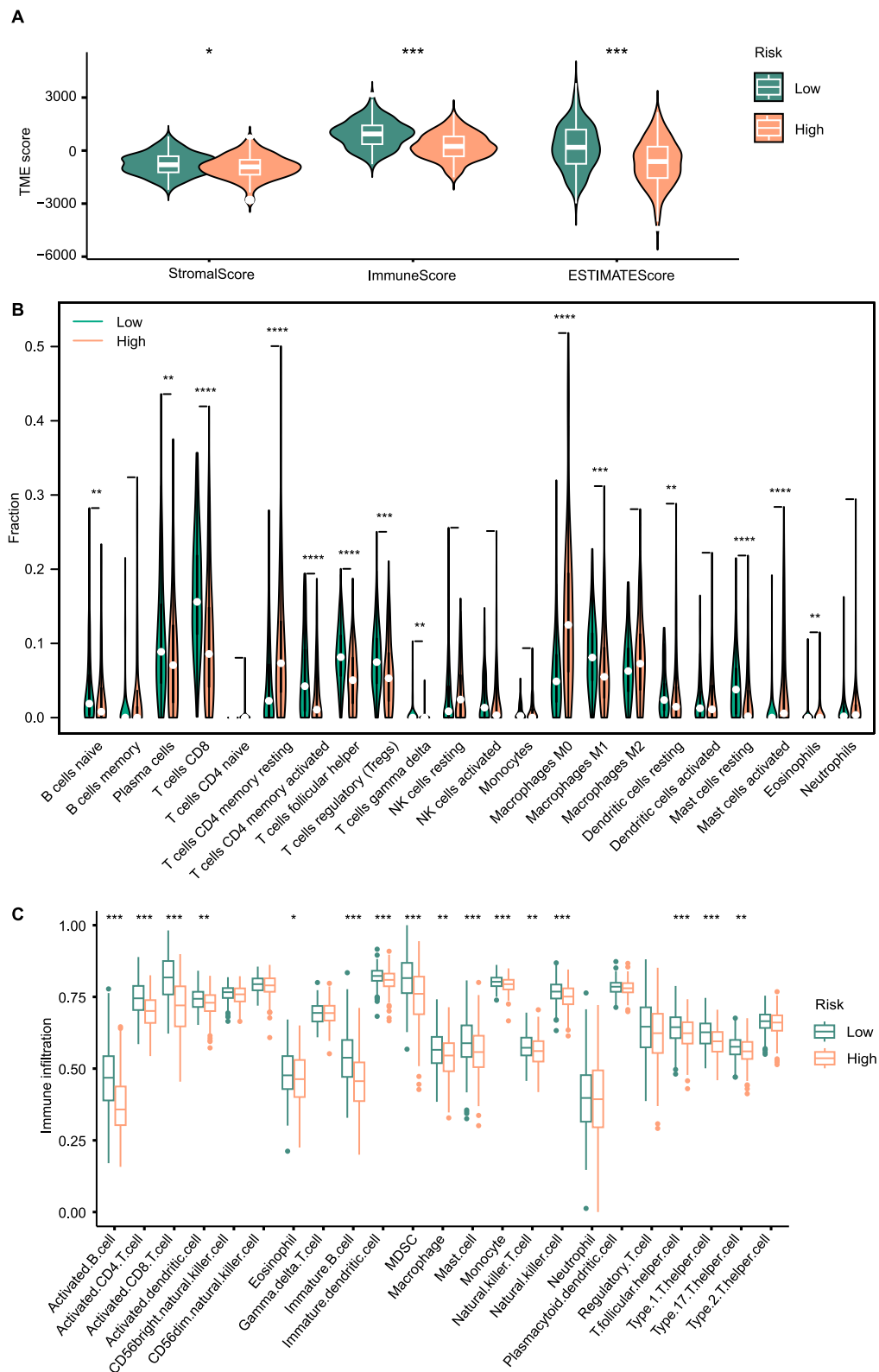
Figure 7 Analysis of tumor immune microenvironment in cervical cancer patients. (**A**) Stroma, immune, and ESTIMATE scores based on ESTIMATE algorithm. (**B**) The violin plots for the infiltration of 22 immune cells according to CIBERSORT analysis. (**C**) The boxplots for the tumor infiltrating lymphocytes assessments by the ssGSEA algorithm. *p < 0.05, **p < 0.01, and ***p < 0.001.

the pronounced differences in immune scores, we further evaluated the immune infiltration patterns of 22 immune cell types in cervical cancer patients using the CIBERSORT algorithm (Figure 7B). The high-risk group showed higher proportions of T cells CD4 memory resting, macrophages M0, activated mast cells and eosinophils, whereas lower proportions of B cells naive, plasma cells, CD8 T cells, activated CD4 memory T cells, T cell follicular helpers, regulatory T cells, macrophage M1, resting dendritic cells, and resting mast cells were noted. Additionally, further immune cell enrichment analysis using the ssGSEA algorithm indicated generally lower immune cell infiltration in the high-risk group, aligning with the lower immune scores observed in this group (Figure 7C).

Discussion

Cervical cancer stands as one of the most prevalent malignant tumors affecting the female reproductive system, with a persistently high global incidence and mortality rate.[21] Early detection and diagnosis are paramount for improving the prognosis of cervical cancer patients. While widely utilized screening methods such as cytology and HPV testing have achieved some degree of success, their suffer from limited specificity and sensitivity, underscoring the need for more effective biomarkers to enhance cervical cancer diagnosis and patient risk stratification.[22,23] In this study, we identified and validated a novel anoikis-related gene signature, which has significant prognostic value in cervical cancer and effectively stratifies patients into distinct risk groups, potentially aiding in personalized risk assessment. These findings suggest that anoikis-related genes could serve as promising prognostic biomarkers and provide valuable insights into the biological mechanisms underlying cervical cancer progression. Anoikis, a critical form of programmed cell death, regulates cell number and tissue structure by inducing cells to detach from extracellular matrix connections, thus preventing abnormal proliferation and inhibiting detached cells from reattaching to unsuitable matrices. It plays a pivotal role in tumorigenesis, progression, and metastasis.[24] In recent years, the scientific community has increasingly recognized the significance of anoikis, particularly its links to properties such as anchorage-dependent growth[25] and epithelial-mesenchymal transition,[26] both critical in tumor progression and cancer cell dissemination. Furthermore, studies in various malignancies, including osteosarcoma, lung adenocarcinoma, renal clear cell carcinoma, colon adenocarcinoma, hepatocellular carcinoma, and bladder cancer, have demonstrated the pivotal roles of anoikis regulators.[27–32] However, the full extent of the mechanisms and prognostic value of ARGs in cervical cancer remains to be elucidated.

Previous studies have indicated that ARGs possess potential predictive value in the prognosis of various cancers, and their expression patterns are closely associated with tumor aggressiveness, metastasis, and patient survival. For instance, a study by Frisch et al (2016) demonstrated that ARG expression patterns could predict disease-free survival and overall survival in breast cancer patients, with patients expressing specific ARG sets exhibiting poorer responses to chemotherapy.[26] Tan et al (2021) developed an ARG-based risk scoring model that effectively predicted the prognosis of lung adenocarcinoma patients and was significantly correlated with tumor mutation burden (TMB) and immune infiltration.[33] Huang et al (2022) discovered that ARG expression profiles could serve as independent predictors of survival and metastasis in pancreatic cancer patients, with EMT-related ARGs being particularly important.[34]

By integrating expression profiles of cervical cancer tissue and survival data from the TCGA-CESC and CGCI-HTMCP-CC datasets, we initially screened for ARGs sourced from the Genecards website and the Harmonizome portal. Through univariate Cox prognostic analysis, we identified a set of prognostic relevant ARGs, based on which we divided samples into subtype A and subtype B. The significantly poorer prognosis of subtype A suggested that ARGs may play a substantial role in cervical cancer prognosis. Using the identified prognostically relevant ARGs, we developed a robust risk score model featuring 10 key genes to predict the prognosis of cervical cancer patients. In selecting our final prognostic model, we rigorously compared the performance of the LASSO-Cox derived 10-gene signature against several machine learning approaches using repeated cross-validation (Supplementary Table 5). While machine learning models hold promise for capturing complex interactions, our analysis indicated they were prone to overfitting on this dataset, showing high performance on training data but diminished accuracy on unseen test data. The Cox model provided a more robust and generalizable performance, while also offering clear interpretability via hazard ratios for each gene. This balance justified its selection as our primary model. Our subsequent analyses demonstrated that this model exhibits excellent stratification and predictive performance, effectively categorizing cervical cancer patients into high-risk and

low-risk groups. Patients in the high-risk group exhibited a significantly poorer prognosis. We found a strong association between high-risk scores and T stage and FIGO stage, and the nomogram proved that our risk model had a better performance than other clinical factors alone in prognosis prediction, underscoring itspotential clinical utility in predicting cervical cancer outcomes. Building on these findings, we further demonstrated the predictive power of the ARG-based model in both training and validation cohorts. Specifically, our analysis revealed that patients in the low-risk group had significantly higher survival rates compared to the high-risk group, further confirming the close association between ARG expression patterns and the malignant progression of cervical cancer. Furthermore, GO and KEGG enrichment analyses indicated that these ARGs are primarily involved in crucial biological processes and signaling pathways such as cell apoptosis, cell adhesion, tumor microenvironment remodeling, and immune evasion. The aberrant activation of these pathways may enhance invasiveness and metastatic potential, enabling cancer cell to evade anoikis and drive cervical cancer progression.

Previous studies have shown that ITGA5 is closely associated with the survival and progression of cervical cancer patients and effectively promotes angiogenesis via the AKT/VEGFA pathway in cervical cancer.[35] Researchers have also reported that E6/E7 of HPV16 and 18 regulates the overexpression of CXCL8 in cervical cancer tissues, and that CXCL8 is associated with malignant status and prognosis in cervical cancer patients.[36,37] TP73 was overexpressed in cervical cancer and associated with favorable patient prognosis.[38] GZMB is involved in programmed cell death and natural killer cell mediated cytotoxicity, and plays a key role in LINC02474 inhibiting apoptosis in colorectal cancer.[39] And as expected, GZMB, overexpressed in cervical cancer tissue, was a significantly protective prognosis marker in our experimental validation.

This study employed univariate Cox regression as an initial screening method for ARGs, a commonly used approach in biomarker studies.[40] Although univariate analysis has inherent limitations, such as its inability to account for confounding factors and interactions between variables, it offers the advantages of being intuitive and computationally efficient as an initial screening tool. When screening for potential prognostic biomarkers from high-dimensional gene data, the strategy of combining univariate Cox regression with subsequent multivariate analysis is widely adopted.[41] Using this analytical strategy, we successfully identified 83 ARGs significantly associated with cervical cancer prognosis, laying the foundation for subsequent model construction. To construct the prognostic model, we systematically evaluated the applicability of various statistical and machine learning methods. Among them, traditional Cox proportional hazards regression, as a standard method for survival analysis, offers direct clinical interpretability, providing hazard ratios (HRs) and confidence intervals (CIs), while effectively handling censored data.[42] However, the Cox model faces significant limitations when dealing with high-dimensional data and is strictly dependent on the proportional hazards assumption. To overcome these limitations, we employed the LASSO regression method. LASSO (Least Absolute Shrinkage and Selection Operator) effectively enhances model performance with high-dimensional data and in the presence of collinearity by introducing an L1 regularization term to simultaneously achieve feature selection and coefficient estimation.[43,44] We also assessed the potential of various machine learning approaches for prognostic modeling. Random Forest, as a tree-based ensemble method, can effectively capture non-linear relationships and complex interactions, is not restricted by the proportional hazards assumption, and exhibits strong robustness to outliers.[15] Gradient Boosting Machine (GBM) and its optimized version XGBoost optimize predictive performance by sequentially building weak learners, and are particularly suitable for handling mixed-type data.[45] However, these machine learning methods generally suffer from "black box" characteristics, and limited clinical interpretability, and often require larger sample sizes to avoid overfitting. After comprehensive evaluation, we selected LASSO-Cox regression as the primary modeling method for this study, mainly based on the following considerations: (1) LASSO-Cox can effectively control the risk of overfitting while maintaining a degree of clinical interpretability; (2) Through its built-in feature selection process, LASSO can identify the most prognostically valuable gene biomarkers, providing clues for subsequent biological mechanism research; (3) Under the current sample size conditions, the machine learning models we implemented (Random Forest, XGBoost, GBM) did not significantly outperform the LASSO-Cox model, possibly due to the relatively limited sample size.[46,47] It is worth noting that when the sample size increases and multi-omics data are incorporated, methods such as deep learning may demonstrate greater advantages.[48]

While our risk scoring model demonstrates promising predictive performance, it is currently a research tool and not yet poised for direct clinical application. Further research is needed to translate these findings into clinical utility. Future studies should focus on prospective validation of our gene signature in independent patient cohorts, ideally in a clinical trial setting. Moreover, exploring the development of a clinically applicable diagnostic assay based on our risk scoring model could be a valuable next step. Such an assay could potentially be used to identify high-risk patients who may benefit from more intensive treatment or closer monitoring, and to guide personalized treatment strategies based on individual risk profiles. In addition, functional validation of the identified anoikis-related genes and further investigation of their role in cervical cancer biology are warranted to fully understand their clinical implications.

Our ARG prognostic model demonstrated superior predictive performance compared to both a clinical-factor-only model and previously published multi-gene prognostic models for cervical cancer.[27,32,41] This suggests that anoikis-related genes may capture unique molecular characteristics associated with cervical cancer prognosis. Notably, the nomogram integrating the ARG model with clinical factors further improved prediction accuracy (AUC=0.89), highlighting the value of integrating molecular biomarkers with traditional clinical factors, which aligns with the principles of predictive model development proposed by Steyerberg et al.[49]

Nevertheless, translating biomarkers into clinical applications still faces multiple challenges, including standardization of techniques, cost-effectiveness analysis, and prospective validation.[50] In the future, we plan to validate the predictive value of the ARG model through multi-center clinical studies and explore its potential applications in treatment decisions and patient stratification, such as identifying high-risk patients who may benefit from intensified treatment or low-risk patients suitable for de-escalated therapy. Furthermore, combining the ARG model with emerging liquid biopsy and circulating tumor DNA technologies may offer new avenues for non-invasive monitoring and personalized treatment of cervical cancer.[51]

However, several limitations must be acknowledged. First, our sample size was relatively small, potentially affecting the statistical robustness of the results. Future studies with larger sample sizes are needed to enhance the reliability of the findings. Second, while our study employed an observational design involving survival analysis to evaluate the prognostic value of the markers, experimental studies are needed to validate the functional roles of these markers. Third, this study represents a bioinformatics exploration, and the specific mechanisms require further elucidation. Fourth, it is important to note that this study was retrospective and based on publicly available databases, which could introduce sample selection bias. Future research should aim to integrate data from multiple sources to improve result reliability. Fifth, considering the multifactorial nature of cervical cancer, this study did not comprehensively incorporate genetic and environmental influences, which presents a notable limitation. Finally, the markers identified in this study are still in the early stages of clinical translation, warranting further validation studies with larger sample sizes and the development of precise detection technologies.

In summary, our study systematically constructed and evaluated a risk scoring model for cervical cancer based on 10 ARGs, demonstrating high accuracy in predicting the prognosis of cervical cancer patients within the analyzed datasets. Moreover, we identified potential novel biomarkers and provided methodological foundations that may facilitate future personalized screening, targeted therapy, and prognostic assessment for cervical cancer. We also revealed differences in immune phenotypes and tumor microenvironments between high- and low-risk cervical cancers, which may provide insights for the development of immunotherapies and personalized treatment strategies. Our study enhances the understanding of the role of ARGs in cervical cancer prognosis and provides a foundation for future research aimed at reducing mortality rates and improving patient outcomes.

Institutional Review Board Statement

The tissue microarray were obtained from Outdo Biotech. Co., Ltd. (Shanghai, China; Product Code: HUteS168Su01). The study involving human participants were reviewed and approved by Shanghai Outdo biotech Co. Ltd. Ethics Committee.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Sharing Statement

Publicly available datasets were analyzed in this study. The names of the repositories and accession numbers are contained within the article.

Acknowledgments

We acknowledged TCGA and CGCI database for providing their platform and contributors for uploading their meaningful datasets.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Disclosure

The authors declare no conflicts of interest in this work.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249. doi:10.3322/caac.21660
2. Macios A, Nowakowski A. False negative results in cervical cancer screening-risks, reasons and implications for clinical practice and public health. *Diagnostics* 2022;12(6):1508. doi:10.3390/diagnostics12061508
3. Güzel C, van Sten-Van't Hoff J, de Kok IMCM, et al. Molecular markers for cervical cancer screening. *Expert Rev Proteomics*. 2021;18 (8):675–691. doi:10.1080/14789450.2021.1980387
4. Khan S, Fatima F, Malik F, et al. Understanding the cell survival mechanism of anoikis-resistant cancer cells during different steps of metastasis. *Clin Exp Metastasis*. 2022;39(5):715–726. doi:10.1007/s10585-022-10172-9
5. Han H, Sung J, Kim S, et al. Fibronectin regulates anoikis resistance via cell aggregate formation. *Cancer Letters*. 2021; 508:59–72. doi:10.1016/j.canlet.2021.03.011
6. Zhong X, Rescorla FJ. Cell surface adhesion molecules and adhesion-initiated signaling: understanding of anoikis resistance mechanisms and therapeutic opportunities. *Cell Signal*. 2012;24(2):393–401. doi:10.1016/j.cellsig.2011.10.005
7. Tian T, Lu Y, Lin J, et al. CPT1A promotes anoikis resistance in esophageal squamous cell carcinoma via redox homeostasis. *Redox Biol*. 2022;58:102544. doi:10.1016/j.redox.2022.102544
8. Zhang T, Wang B, Su F, et al. TCF7L2 promotes anoikis resistance and metastasis of gastric cancer by transcriptionally activating PLAUR. *Int J Biol Sci*. 2022;18(11):4560–4577. doi:10.7150/ijbs.69933
9. Sharma R, Gogoi G, Saikia S, et al. BMP4 enhances anoikis resistance and chemoresistance of breast cancer cells through canonical BMP signaling. *J Cell Commun Signal*. 2022;16(2):191–205. doi:10.1007/s12079-021-00649-9
10. González-Llorente L, Santacatterina F, García-Aguilar A, et al. Overexpression of mitochondrial IF1 prevents metastatic disease of colorectal cancer by enhancing anoikis and tumor infiltration of NK cells. *Cancers* 2019;12(1):22. doi:10.3390/cancers12010022
11. Liao C, Li M, Chen X, et al. Anoikis resistance and immune escape mediated by Epstein-Barr virus-encoded latent membrane protein 1-induced stabilization of PGC-1α promotes invasion and metastasis of nasopharyngeal carcinoma. *J Exp Clin Cancer Res*. 2023;42(1):261. doi:10.1186/s13046-023-02835-6
12. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis – an introduction to concepts and methods. *Br J Cancer*. 2003;89(3):431–436. doi:10.1038/sj.bjc.6601119
13. Zhao Q, Shi X, Xie Y, Huang J, Shia B, Ma S. Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA. *Briefings in Bioinformatics*. 2015;16(2):291–303. doi:10.1093/bib/bbu003
14. Cox DR. Regression models and life-tables. *J Royal Statistical Soc*. 1972;34(2):187–202. doi:10.1111/j.2517-6161.1972.tb00899.x
15. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Statist*. 2008;2(3):841–860. doi:10.1214/08-AOAS169
16. Xgboost: extreme gradient boosting. *R package version 04-2*. 2015;1(4):1-4.
17. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Statist*. 2001;29(5):1189–1232.
18. Du H, Tang Y, Ren X, et al. A prognostic model for cervical cancer based on ferroptosis-related genes. *Front Endocrinol*. 2022:13. doi:10.3389/fendo.2022.991178.
19. Shi H, Zhong F, Yi X, et al. Application of an autophagy-related gene prognostic risk model based on TCGA database in cervical cancer. *Front Genet*. 2021:11. doi:10.3389/fgene.2020.616998.

https://doi.org/10.2147/IJGM.S508059 **2877**

20. Wang N, Nanding A, Jia X, et al. Mining of immunological and prognostic-related biomarker for cervical cancer based on immune cell signatures. *Front Immunol*. 2022;13. doi:10.3389/fimmu.2022.993118.

21. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424. doi:10.3322/caac.21492

22. Xiao FY, Xie F, Sui L. Diagnostic accuracy of colposcopically directed biopsy and loop electrosurgical excision procedure for cervical lesions. *Reprod Develop Med*. 2018;2(3):137. doi:10.4103/2096-2924.248448

23. Wang Q, Zhu CY, Chen LM, et al. Clinical value of human papillomavirus E6/E7 mRNA testing in patients with atypical squamous cells of undetermined significance and low-grade squamous intraepithelial lesion. *Reprod Develop Med*. 2018;2(3):157. doi:10.4103/2096-2924.248485

24. Paoli P, Giannoni E, Chiarugi P. Anoikis molecular pathways and its role in cancer progression. *Biochimica Et Biophysica Acta*. 2013;1833 (12):3481–3498. doi:10.1016/j.bbamcr.2013.06.026

25. Dai Y, Zhang X, Ou Y, et al. Anoikis resistance—protagonists of breast cancer cells survive and metastasize after ECM detachment. *Cell Commun Signal*. 2023;21(1):190. doi:10.1186/s12964-023-01183-4

26. Frisch SM, Schaller M, Cieply B. Mechanisms that link the oncogenic epithelial–mesenchymal transition to suppression of anoikis. *J Cell Sci*. 2013;126(1):21–29. doi:10.1242/jcs.120907

27. Liu Y, Shi Z, Zheng J, et al. Establishment and validation of a novel anoikis-related prognostic signature of clear cell renal cell carcinoma. *Front Immunol*. 2023;14. doi:10.3389/fimmu.2023.1171883

28. Zhang Z, Zhu Z, Fu J, et al. Anoikis patterns exhibit distinct prognostic and immune landscapes in Osteosarcoma. *Inter Immunopharmacol*. 2023;115:109684. doi:10.1016/j.intimp.2023.109684

29. Wang Q, Sun N, Zhang M. Identification and validation of anoikis-related signatures for predicting prognosis in lung adenocarcinoma with machine learning. *IJGM*. 2023;16:1833–1844. doi:10.2147/IJGM.S409006

30. Hu G, Li J, Zeng Y, et al. The anoikis-related gene signature predicts survival accurately in colon adenocarcinoma. *Sci Rep*. 2023;13(1):13919. doi:10.1038/s41598-023-40907-x

31. Guizhen Z, Weiwei Z, Yun W, Guangying C, Yize Z, Zujiang Y. An anoikis-based signature for predicting prognosis in hepatocellular carcinoma with machine learning. *Front Pharmacol*. 2023;13. doi:10.3389/fphar.2022.1096472

32. Dong Y, Xu C, Su G, et al. Clinical value of anoikis-related genes and molecular subtypes identification in bladder urothelial carcinoma and in vitro validation. *Front Immunol*. 2023;14:1122570. doi:10.3389/fimmu.2023.1122570

33. Tan BS, Yang MC, Singh S, et al. LncRNA NORAD is repressed by the YAP pathway and suppresses lung and breast cancer metastasis by sequestering S100P. *Oncogene*. 2019;38(28):5612–5626. doi:10.1038/s41388-019-0812-2

34. Song W, Hu H, Yuan Z, Yao H. A prognostic model for anoikis-related genes in pancreatic cancer. *Sci Rep*. 2024;14(1):15200. doi:10.1038/s41598-024-65981-7

35. Xu X, Shen L, Li W, Liu X, Yang P, Cai J. ITGA5 promotes tumor angiogenesis in cervical cancer. *Cancer Med*. 2023;12(10):11983–11999. doi:10.1002/cam4.5873

36. Fernandez-Avila L, Castro-Amaya AM, Molina-Pineda A, Hernández-Gutiérrez R, Jave-Suarez LF, Aguilar-Lemarroy A. The value of CXCL1, CXCL2, CXCL3, and CXCL8 as potential prognosis markers in cervical cancer: evidence of E6/E7 from HPV16 and 18 in chemokines regulation. *Biomedicines*. 2023;11(10):2655. doi:10.3390/biomedicines11102655

37. Yan R, Shuai H, Luo X, Wang X, Guan B. The clinical and prognostic value of CXCL8 in cervical carcinoma patients: immunohistochemical analysis. *Biosci Rep*. 2017;37(5):BSR20171021. doi:10.1042/BSR20171021

38. Ye H, Guo X. TP73 is a credible biomarker for predicting clinical progression and prognosis in cervical cancer patients. *Biosci Rep*. 2019;39(8): BSR20190095. doi:10.1042/BSR20190095

39. Du T, Gao Q, Zhao Y, et al. Long non-coding RNA LINC02474 affects metastasis and apoptosis of colorectal cancer by inhibiting the expression of GZMB. *Front Oncol*. 2021;11. doi:10.3389/fonc.2021.651796

40. Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 2: prognostic factor research. *PLOS Medicine*. 2013;10(2): e1001380. doi:10.1371/journal.pmed.1001380

41. Wu J, Zhao Y, Zhang J, Wu Q, Wang W. Development and validation of an immune-related gene pairs signature in colorectal cancer. *OncoImmunology*. 2019. doi:10.1080/2162402X.2019.1596715

42. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Available from: https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4. Accessed March 23, 2025.

43. Tibshirani R. Regression shrinkage and selection via the lasso. *J Royal Statist Soc Series B* 1996;58(1):267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

44. Simon N, Friedman JH, Hastie T, Tibshirani R. Regularization paths for cox's proportional hazards model via coordinate descent. *J Statist Software*. 2011;39(5):1–13. doi:10.18637/jss.v039.i05

45. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM; 2016:785–794. doi:10.1145/2939672.2939785.

46. Boulesteix AL, Schmid M. Machine learning versus statistical modeling. *Biometric J*. 2014;56(4):588–593. doi:10.1002/bimj.201300226

47. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22. doi:10.1016/j.jclinepi.2019.02.004

48. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface*. 2018;15(141):20170387. doi:10.1098/rsif.2017.0387

49. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Europ Heart J*. 2014;35(29):1925–1931. doi:10.1093/eurheartj/ehu207

50. Hayes DF. Biomarker validation and testing. *Molecular Oncol*. 2015;9(5):960–966. doi:10.1016/j.molonc.2014.10.004

51. Kinde I, Bettegowda C, Wang Y, et al. Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. *Sci Transl Med* 2013;5(167). doi:10.1126/scitranslmed.3004952