Submitted: 14 July, 2023 Accepted: 14 August, 2023 Published: 15 December, 2023 DOI:10.22514/ejgo.2023.105

ORIGINAL RESEARCH

European Journal of Gynaecological Oncology

# Comprehensive analysis of hypoxia-related gene signature in cervical cancer

Tingting He[1,2,†], Xiaoyu Tang[1,†], Siru Chen[3], Xin Chen[1], Fuye Lin[4], Minmin Yu[5,*], Changsong Lin[4,*]

[1]Nanjing University of Chinese Medicine, 210023 Nanjing, Jiangsu, China
[2]Department of Gynecology, Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine, 210022 Nanjing, Jiangsu, China
[3]Department of Traditional Chinese Medicine, The Affiliated Huai'an Hospital of Xuzhou Medical University and The Second People's Hospital of Huai'an, 223002 Huai'an, Jiangsu, China
[4]Department of Bioinformatics, Nanjing Medical University, 211166 Nanjing, Jiangsu, China
[5]Department of Gynecology, Nanjing Hospital Affiliated to Nanjing University of Chinese Medicine, 210003 Nanjing, Jiangsu, China

***Correspondence**
njyy022@njucm.edu.cn (Minmin Yu);
lcs04bio@njmu.edu.cn (Changsong Lin)

† These authors contributed equally.

## Abstract

Hypoxia significantly influences the growth, metastasis and treatment resistance of cervical cancer (CC), thereby affecting patient prognosis. However, accurately predicting CC survival remains challenging, and the potential of hypoxia-related genes as prognostic markers remains uncertain. In this study, using CC single-cell transcriptional data from the Gene Expression Omnibus database, we employed the InferCNV package to identify tumor cells and used CellChat to confirm stronger intercellular interactions in tumor cells with high-hypoxia status. Next, we identified differentially expressed hypoxia-related genes (DEHLGs) by analyzing data from the Cancer Genome Atlas (TCGA), Genotype-Tissue Expression, and Molecular Signature Database, which were further screened using univariate Cox regression and lasso regression analyses, based on which we constructed a hypoxia prognosis model comprising nine prognosis-related genes. Risk scores were generated using multivariate Cox regression analysis. The prognosis model revealed that the overall survival rate was higher in the low-risk than in the high-risk group. The model's performance was assessed using the area under the time-dependent receiver operator characteristic curve, which yielded values of 0.836 and 0.804 for the training and test groups, respectively, indicating a robust prognostic capability of the model. A nomogram based on the nine hypoxia-related genes and training groups exhibited a favorable discriminatory ability for CC. Additionally, using CIBERSORT, we estimated the proportion of immune cells in patients with high- and low-hypoxia risk, revealing a higher proportion of macrophages (M0) and activated mast cells in the high-risk group. We successfully established a prognostic model for CC based on nine hypoxia-related genes to accurately predict the prognosis of affected patients.

## Keywords

Hypoxia; Cervical cancer; Prognosis; Prediction; Gene signature

## 1. Introduction

Cervical cancer (CC) accounts for the majority of cancer-related deaths in women [1], despite the widespread increase in screening and treatment [2]. Unfortunately, CC is highly aggressive and patients with metastases and recurrence have limited treatment options, and a poor overall prognosis [3]. Furthermore, CC is also highly heterogeneous and there is a lack of tools to address this biological heterogeneity [4]. Therefore, the identification of prognostic biomarkers for CC is crucial for achieving precise treatment, especially for patients with advanced illness.

Malignant tumors often exhibit hypoxia, which significantly contributes to their progression. In hypoxic tumors, genes related to tumor growth, invasiveness, angiogenesis and evasion from apoptosis are upregulated, rendering tumors more aggressive [5]. Consequently, hypoxia is considered a factor that adversely affects prognosis in various cancers, including bladder, breast, colon, gastric, liver and uterine cancers [6]. According to a study that directly assessed the oxygenation status in tumors, patients with hypoxic cervical cancer demonstrated poorer rates of disease-free survival [7]. Hypoxia promotes CC cell migration through the Rab11 trafficking integrins $\alpha v\beta 3$/focal adhesion kinase (FAK)/phosphoinositide 3-kinases (PI3K) signaling pathway [8]. However, there is no clear evidence of a relationship between hypoxia and CC cells. It remains unclear how hypoxia drives CC progression and affects the expression of related genes. Therefore, exploring hypoxia-related genes may contribute to survival prediction and treatment of CC.

The widespread application of second-generation and single-cell sequencing has provided a deeper understanding of tumor gene mutations and microenvironments. An increasing number of prognostic gene markers have been identified, shedding light on the characteristics of tumor occurrence and development. Single-cell sequencing enables specific analysis of cell groups at the single-cell level, revealing differences in transcriptional expression among different tumor cell subsets

[9].

Hypoxia plays a significant role in tumor occurrence, metastasis, immune escape and treatment resistance. CC is a highly heterogeneous, invasive and malignant disease. However, whether tumor cells in CC are in a hypoxic state or whether hypoxia-related genes can help predict patient prognosis remains unclear. Therefore, we employed single-cell sequencing technology to explore hypoxia in CC cells and the interaction between tumor cells under varying hypoxic conditions in different CC cells. We obtained clinicopathological and transcriptomic data from the Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) databases. Using the Molecular Signatures Database (MsigDB) and high-throughput bioinformatics analysis, we identified hypoxia-related genes abnormally expressed in both cancer and healthy tissues. We carefully discuss their possible functions and molecular mechanisms. The identified hypoxia-related genes associated with the prognosis of CC serve as biomarkers clinical practice. Subsequently, we developed a prognostic model of CC based on differentially expressed hypoxia-related genes (DEHLGs). Our model can assist in risk stratification and prognosis assessment of patients with CC, and the identified genes may become potential new therapeutic targets.

## 2. Materials and methods

### 2.1 Data processing

We downloaded single-cell transcriptome data from the Gene Expression Omnibus (GEO) database (GEO accession: GSE168652), including CC and adjacent tissue data. Data analysis was conducted using the R package Seurat (v4.0, https://github.com/satijalab/seurat) [10]. Various analyses, such as principal component analysis, uniform manifold approximation and projection (UMAP), dimensionality reduction and visualization were conducted on the data. To identify cell types, we employed SingleR (v1.0.5) [11] and to examine the copy number alterations in tumor cells, we used InferCNV (v1.16.0, Trinity CTAT Project, https://github.com/broadinstitute/inferCNV) [11]. Gene regulatory network analysis was performed using pySCENIC (v0.11.2, https://packages.guix.gnu.org/packages/pyscenic/0.11.2/) with default parameters. The AUCell algorithm was used to calculate the regulon activity score (RAS) [12]. The differential expression analysis of important pathways in each cell subgroup was conducted using the gene set variation analysis (GSVA) package [13], which is a gene set enrichment (GSE) method that estimates variation of pathway activity across a sample population in an unsupervised manner. To identify and visualize hypoxia-inducible factor 1 subunit alpha (*HIF-1A*) expression in each cell subgroup, we utilized the FeaturePlot function in Seurat. The hypoxia pathway activity scores for each cell subgroup were calculated and visualized using the AUCell package (v3.17, https://github.com/aertslab/AUCell) [14]. Based on the hypoxia score, which served as an independent prognostic factor for tumor progression, tumor cells were divided into high- and low-hypoxia groups using median value as cutoff. The Gene Set Enrichment Analysis (GSEA) function embedded in the ClusterProfiler (v4.0 https://github.com/YuLab-SMU/clusterProfiler) software package was employed to identify gene sets that were significantly enriched between the two groups [15]. Finally, to study the molecular interactions between different cell types and between high- and low-hypoxia tumor cells and other cells, we utilized the cell communication tool [16].

For further analyses, we retrieved the RNA-sequencing dataset consisting of 13 normal cervical tissue samples and 305 CC samples along with their corresponding clinical data from the GTEx (https://gtexportal.org/home/) and TCGA (https://www.cancer.gov/tcga) databases. The MSigDB contains a hypoxia-related gene set called HALLMARK_HYPOXIA. Using the limma package, DEHLGs were identified based on $|\log_2 \text{fold change}| \geq 0.5$ and false discovery rate $<0.01$ [17].

### 2.2 Functional enrichment analysis based on gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG)

To gain a comprehensive understanding of the roles of the DEHLGs, we conducted GO and KEGG enrichment analyses using the clusterProfiler package. GO analysis terms included cellular components, molecular functions and biological processes. Values of $p$ and false discovery rate $<0.05$ were set as the cut-off criteria.

### 2.3 Protein-protein interaction (PPI) network construction and subnetwork screening

The DEHLGs were uploaded to the STRING database (http://www.string-db.org/) to construct a PPI network. This network was visualized using Cytoscape (v3.8.2), and each gene node was identified using a circular icon. The molecular complex detection (MCODE) plug-in in Cytoscape was used to identify subnetworks in the PPI network. The identified subnetworks were then subjected to further functional enrichment analysis. The significance level was set at $p \leq 0.05$.

### 2.4 Prognostic model construction

The samples of patients with CC with reliable prognostic information in the TCGA database were randomly divided into training and test groups. Using the survival package, we conducted univariate Cox proportional hazards regression analyses with a $p < 0.05$ cut-off for the prognostic value of each gene. LASSO regression analysis was then applied to identify significant candidate genes. Subsequently, we constructed a prognostic model using the training group. For each patient, we calculated a risk score based on the gene expression level ($Exp_i$) and the coefficient of multivariate Cox regression analysis ($\beta_i$) in the training group, according to the following formula:

$$Risk\ score = \beta_1 \times Exp_1 + \beta_2 \times Exp_2 + \beta_i \times Exp_i$$

Using the median risk score as the threshold, patients in both cohorts were categorized into high- and low-risk groups. The prognostic performance of the model was evaluated using a receiver operating characteristic (ROC) curve generated with the SurvivalROC software package (v1.0.3). The predictive capability of the prognostic model was validated using the results from the test group.

### 2.5 Survival and prognostic analysis

We used pheatmap package to draw the risk curve, survival state diagram, and risk heat diagram for the training and test groups. To assess the impact of clinical parameters and risk scores on the overall survival (OS) of patients with CC, we conducted univariate Cox analysis. Additionally, multivariate Cox analysis was performed to identify independent prognostic factors. Survival analysis was conducted using the survival and survminer packages in R, while the survival package was also used both for univariate and multivariate Cox analyses.

### 2.6 Nomogram construction

The likelihood of OS was predicted using a nomogram with calibration plots, created using the rms package for R software. A significance level of $p < 0.05$ was considered statistically significant.

### 2.7 Verification of the expression level

Immunohistochemical staining images of proteins coding for 7 genes were downloaded from the Human Protein Atlas (HPA) database (http://www.proteinatlas.org/) to validate at the translational level in CC tissues compared to normal tissues.

### 2.8 Evaluation of tumor-infiltrating immune cells

We used CIBERSORT to evaluate the tumor immune microenvironment and its relationship with the hypoxia risk score. This tool allowed the estimation of the proportions of 22 immune cell types between the low- and high-risk groups.

## 3. Results

### 3.1 Hypoxia in CC cells

We followed the process shown in Fig. 1 to analyze the prognostic value of hypoxia-related genes in CC. Single-cell RNA-sequencing data of CC in the GSE168652 dataset were analyzed. Using the SingleR package, we identified five different cell types depicted in the UMAP diagrams, epithelial cells, endothelial cells, smooth muscle cells, monocytes and T cells (Fig. 2A,B). Using the InferCNV package to detect chromosomal amplifications and deletions, we observed copy number alterations in multiple chromosomes specifically in epithelial cells, indicating their cancerous nature (Fig. 2C). Furthermore, GSVA revealed significant variations in hypoxia across different cell types (Fig. 2D).

To further investigate hypoxia in CC cells, we extracted the expression intensity of *HIF-1A* from each cell's subpopulation and found it to be significantly higher in tumor cells, confirming the presence of significant hypoxia (Fig. 2E–I). Notably, our regulon activity score (RAS) analysis, which evaluated the activities of transcriptional regulators and their targeted genes, highlighted the correlation between hypoxia and malignancy.

### 3.2 Strong intercellular interactions of tumor cells with highly hypoxic status

The hypoxia pathway activity scores of individual CC cells were calculated and tumor cells exhibited significantly higher hypoxia scores compared to other cell types (Fig. 3A). Based on these scores, tumor cells were classified into high- and low-hypoxia groups. Validation using GSEA revealed significant enrichment of differentially expressed genes in the hypoxic gene set between the two groups (Fig. 3B). The GSEA diagram depicted variations in pathway activity between the high- and low-hypoxia groups. Furthermore, analyses of cellular interactions (Fig. 3C,D) revealed that tumor cells exhibited enriched cellular interactions with other cell types, with the dot size representing the number of cell groups and the thickness of connecting lines indicating the number of interactions. Notably, cells with high-hypoxia status exhibited stronger interactions with other cells than those of cells with a low-hypoxia status.

### 3.3 Identification of DEHLGs in CC

A total of 177 differentially expressed hypoxia-related genes (DEHLGs) were identified, of which 78 genes met the screening criteria ($p < 0.01$, $|\log_2 \text{fold change}| > 0.5$). Among these genes, 39 were upregulated and 39 were downregulated. The expression distribution of these genes is depicted in Fig. 4A,B.

### 3.4 GO and KEGG pathway enrichment analysis of DEHLGs

To explore the underlying mechanisms and functions of DEHLGs, we performed functional enrichment analysis by categorizing them into upregulated and downregulated groups. GO terms and KEGG pathways revealed several key findings mostly related to metabolic processes (Fig. 5A,C).

The downregulated DEHLGs were significantly enriched in biological processes associated with catabolic processing of carbohydrates, canonical glycolysis and catabolic processing of glucose to pyruvate. In contrast, the upregulated DEHLGs were significantly enriched in the biosynthetic process for purine nucleotides and purine-containing compounds. Molecular function analysis revealed that downregulated DEHLGs were significantly enriched in acetylglucosaminyltransferase activity and monosaccharide binding, whereas the upregulated DEHLGs were significantly enriched in monosaccharide binding, uridine diphosphate (UDP)-glycosyltransferase activity, and cell adhesion molecules. In terms of cellular component analysis, the downregulated DEHLGs exhibited enrichment in the collagen-containing extracellular matrix, vacuolar lumen and lysosomal lumen.

Moreover, KEGG pathway enrichment analyses revealed that the upregulated DEHLGs were enriched in glycolysis/gluconeogenesis, HIF-1 signaling pathway, and central carbon metabolism in cancer. Downregulated DEHLGs were enriched not only in the same processes but also in amino acid biosynthesis (Fig. 5B,D).

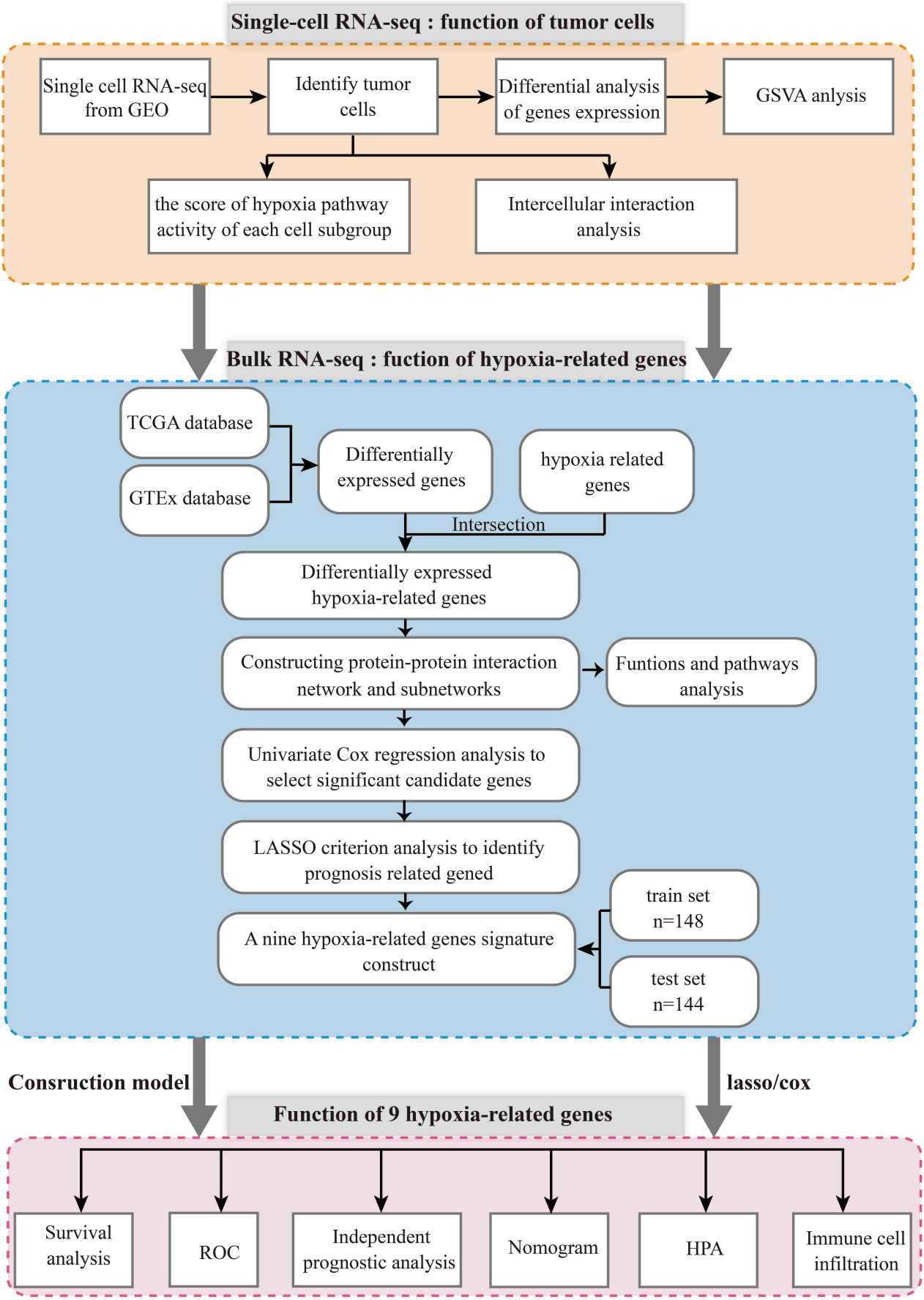


**FIGURE 1. Study design for analyzing hypoxia-related genes in cervical cancer (CC).** Abbreviations: GEO: gene expression omnibus; GSVA: gene set variation analysis; GTEx: Genotype-Tissue Expression; HPA: Human Protein Atlas; RNA-seq: RNA-sequencing; ROC: receiver operator characteristic; TCGA: The Cancer Genome Atlas.

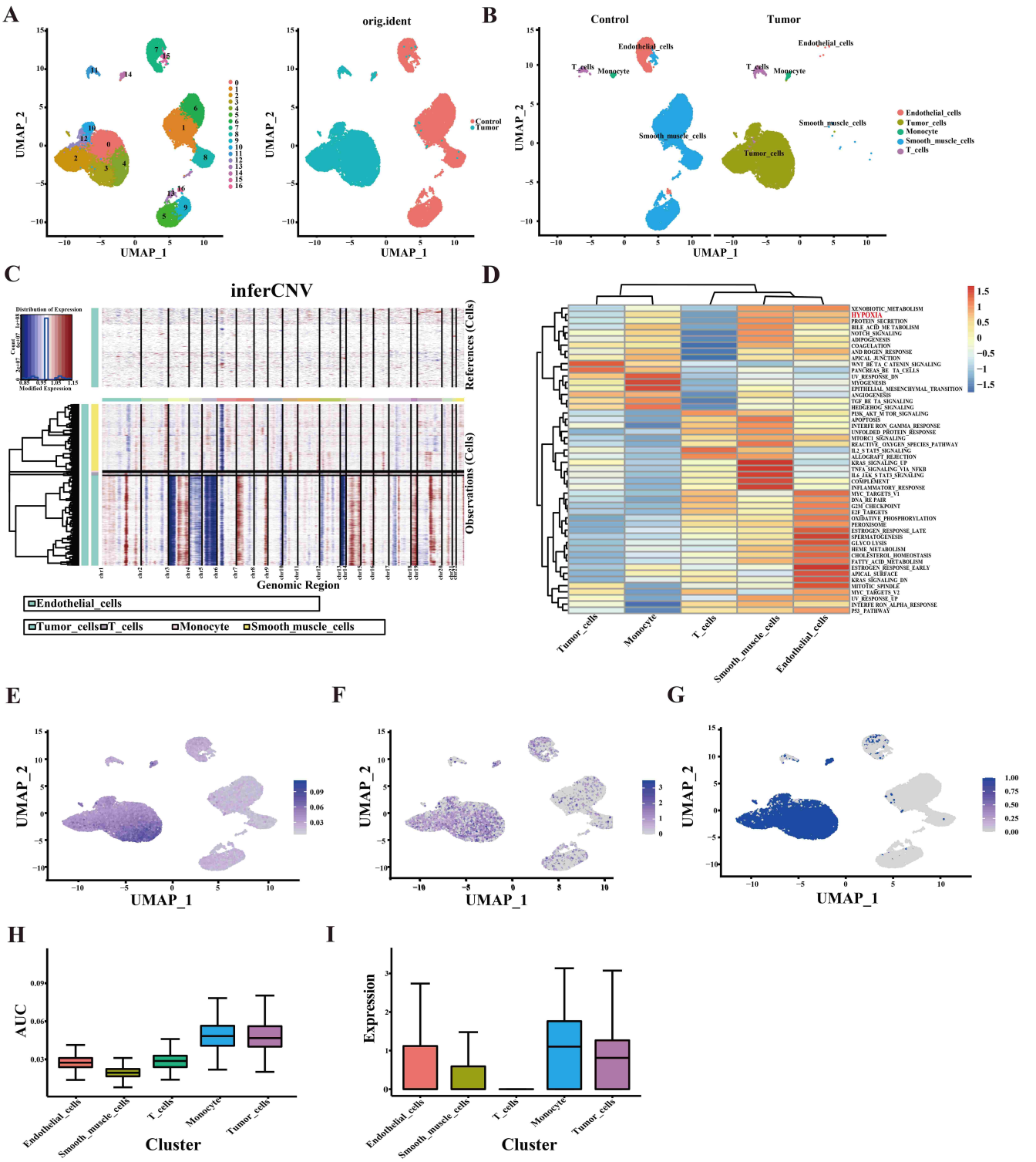


**FIGURE 2. Analysis of hypoxia in cervical cancer (CC) tumor cells using single-cell RNA-sequencing data.** (A) Uniform manifold approximation and projection (UMAP) diagram of all 17 clusters in the GSE168652 dataset. (B) UMAP diagram displaying five different cell types identified using the SingleR package. (C) Analysis of chromosomal copy number variation in tumor cells and cervical endothelial cells as reference data using InferCNV. (D) Heatmap demonstrating gene set enrichment analysis for the five cell types based on gene set variation analysis. (E–G) UMAP diagram showing (E) the RAS (regulon activity score), (F) gene expression and (G) the binarized RAS (all samples were Z score-normalized and converted to 0 and 1) of the hypoxia-inducible factor HIF-1A across different cell subpopulations. (H and I) Box plots presenting (H) the average RAS, and (I) average gene expression of *HIF-1A* in different cell subpopulations.

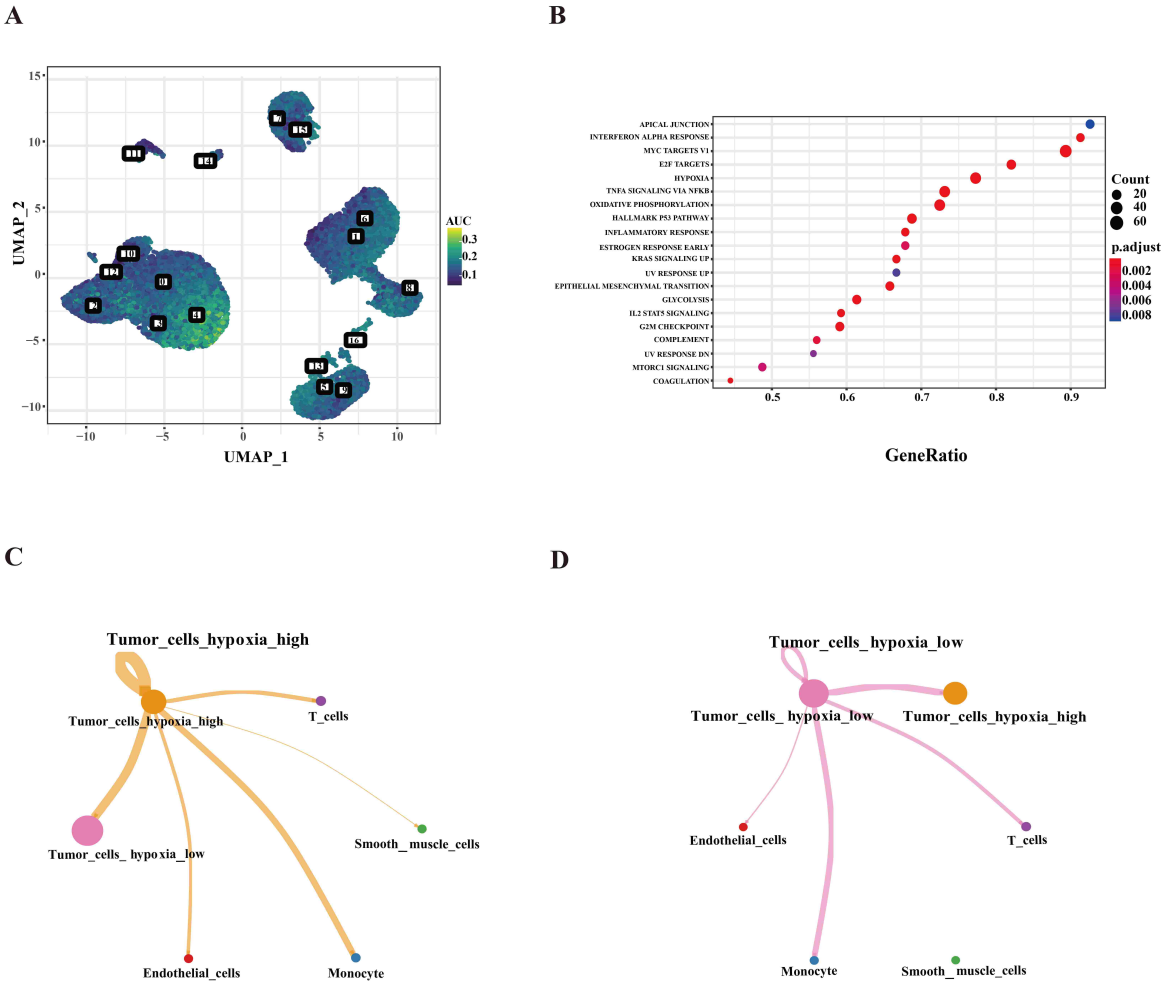


**FIGURE 3. Analysis of interactions between tumor cells in different hypoxic states and other cells in cervical cancer.** (A) Uniform manifold approximation and projection (UMAP) diagram illustrating the fraction of hypoxic pathway activity in each cell cluster. The color spectrum ranging from blue to yellow represents the fraction from low to high. (B) Gene set enrichment analysis (GSEA) diagram displaying differences in pathway activity between the high- and low-hypoxia groups. (C,D) Intercellular interactions depicted through a dot plot, where the size of the dots corresponds to the number of cell groups, and thickness of connecting lines between cell groups represents the number of interactions. AUC: area under the curve.

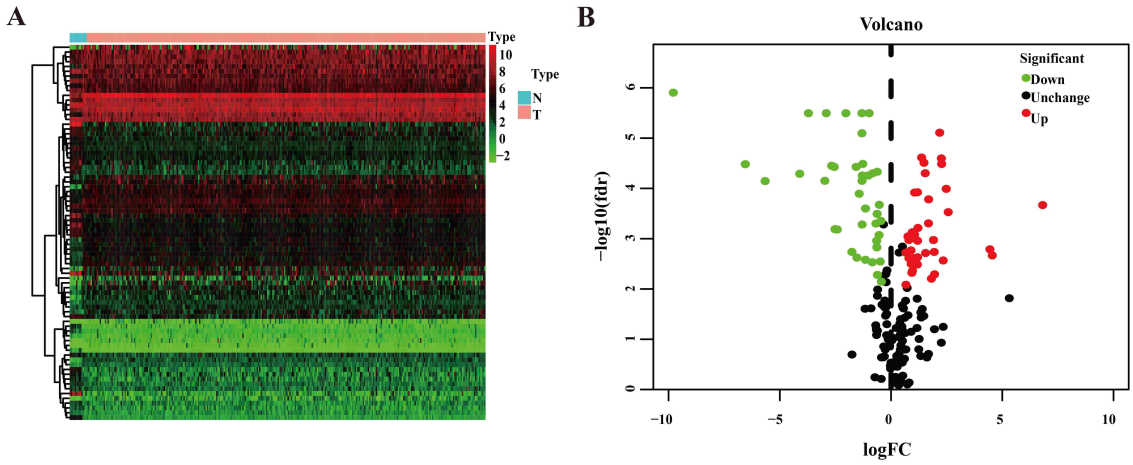


**FIGURE 4. Differentially expressed hypoxia-related genes (DEHLGs) in cervical cancer.** (A) Heatmap illustrates the expression levels of DEHLGs. (B) The volcano plot displays the distribution of DEHLGs, with upregulated genes shown in red, downregulated genes shown in green, and genes with unchanged expression shown in black.

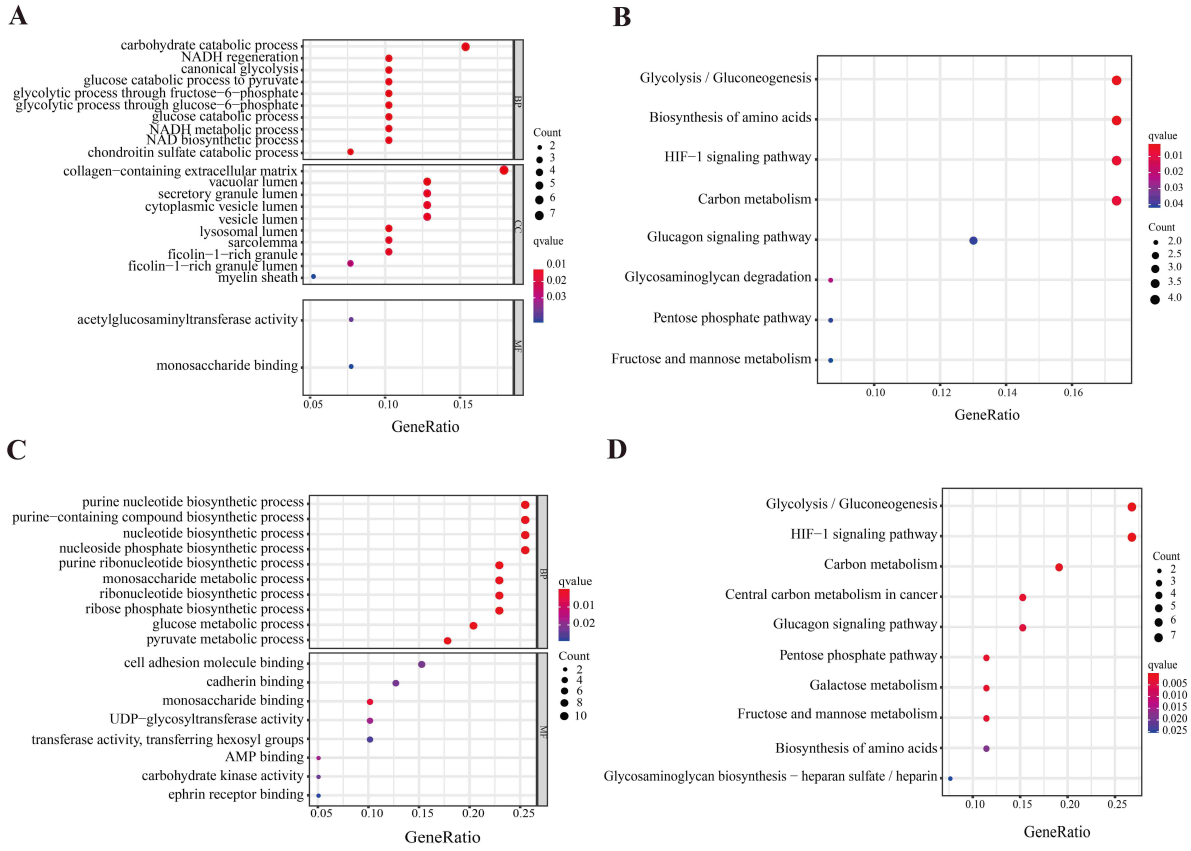


**FIGURE 5. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of differentially expressed hypoxia-related genes (DEHLGs).** (A) GO and (B) KEGG pathway enrichment analyses of downregulated DEHLGs. (C) GO and (D) KEGG pathway enrichment analysis of upregulated DEHLGs.

## 3.5 PPI network and subnetworks

To gain further insights into the roles of DEHLGs in CC, we constructed a PPI network using data from the STRING database. The networks consisted of 93 nodes and 198 edges (Fig. 6A). To identify densely connected regions within the network, we applied the MCODE plug-in, which led to the identification of three distinct subnetworks. Subnetwork 1 (Fig. 6B) comprised genes associated with key metabolic processes such as glycolysis/gluconeogenesis, carbon metabolism, amino acid biosynthesis, the HIF-1 signaling pathway, and fructose and mannose metabolism.

## 3.6 Selection of prognostic hypoxia-related genes

In order to identify hypoxia-related genes with prognostic significance, we conducted univariate Cox regression analysis on the hypoxia-related genes within the PPI network. As a result, we identified 14 candidate genes that showed association with prognosis (Fig. 7). Subsequently, we applied the LASSO criterion for further gene selection, leading to the retention of 9 tumor hypoxia-related genes exhibiting high prognostic value (Fig. 8 and Table 1).

## 3.7 Prognostic signature construction and evaluation

We used the nine hypoxia-related genes identified through the multivariate stepwise Cox regression analysis to construct a prognostic signature. The following formula was used to calculate patient risk scores.

$$\text{Risk score} = (0.0163 \times \text{Exp}_{P4HA1}) + (0.0172 \times \text{Exp}_{PFKP}) + (0.0319 \times \text{Exp}_{PLIN2}) + (-0.2174 \times \text{Exp}_{PDK3}) + (0.0435 \times \text{Exp}_{CAVIN3}) + (-0.0185 \times \text{Exp}_{SIAH2}) + (-0.1071 \times \text{Exp}_{GRHPR}) + (-0.4070 \times \text{Exp}_{LDHC}) + (0.0035 \times \text{Exp}_{TGFBI})$$

Survival analysis was performed to assess the predictive ability of this signature. Based on the median risk score, 148 patients with CC in the training group were split into high- and low-risk subgroups. Patients in the high-risk subgroup exhibited a lower OS than those in low-risk subgroup (Fig. 9A). Time-dependent ROC analyses were conducted to assess the prognostic utility of the signature based on the nine hypoxia-related genes. The area under the ROC curve for the hypoxia gene risk-scoring model was 0.836 (Fig. 9B), indicating a good prognostic performance. A heatmap of the expression signatures of the nine hypoxia-related genes in the low- and high-risk subgroups, patient survival status, and risk score are shown in Fig. 9C–E.

To validate the prognostic value, we applied the same formula to TCGA test group. The results showed an area under the

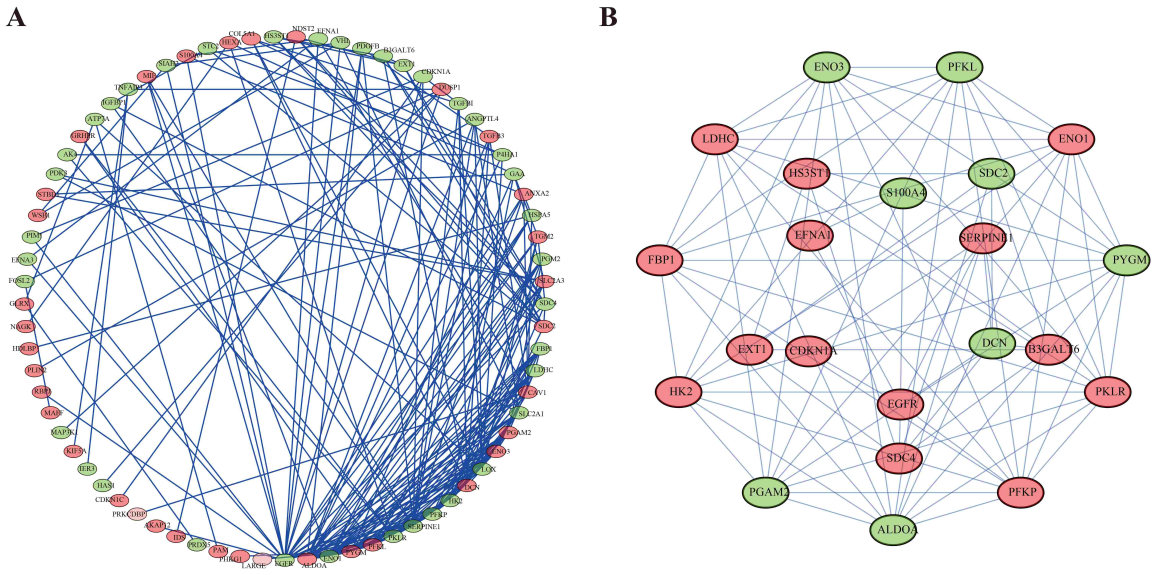


**FIGURE 6. Protein–protein interaction (PPI) network and subnetwork analyses.** (A) PPI network of differentially expressed hypoxia-related genes (DEHLGs). (B) Subnetwork 1 of the PPI network. Green circles represent downregulated genes and red circles represent upregulated genes.

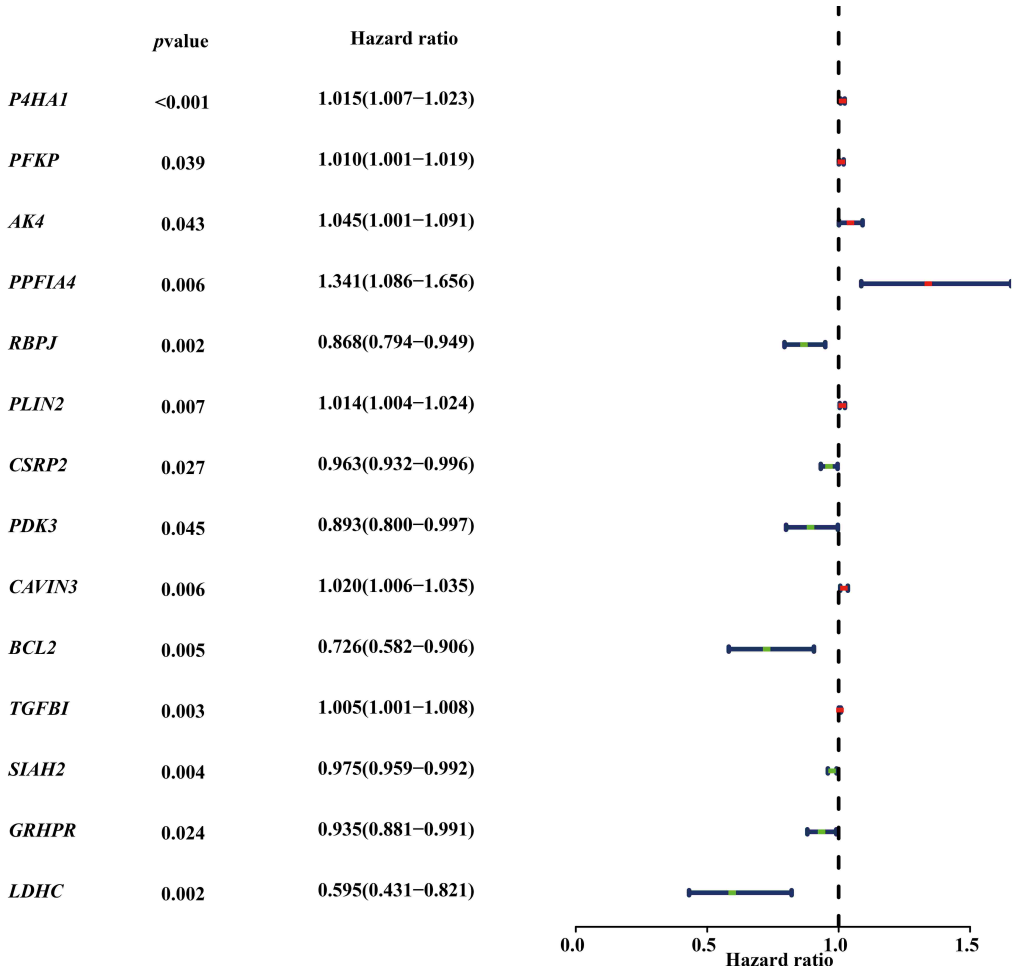


**FIGURE 7. Results of univariate Cox proportional hazards regression (Cox regression) analysis of 14 candidate genes associated with disease prognosis and hypoxia in the training group.** The Nomogram displays hazard ratio scores and 95% confidence intervals. Genes depicted in green represent downregulated genes, while genes depicted in red represent upregulated genes.

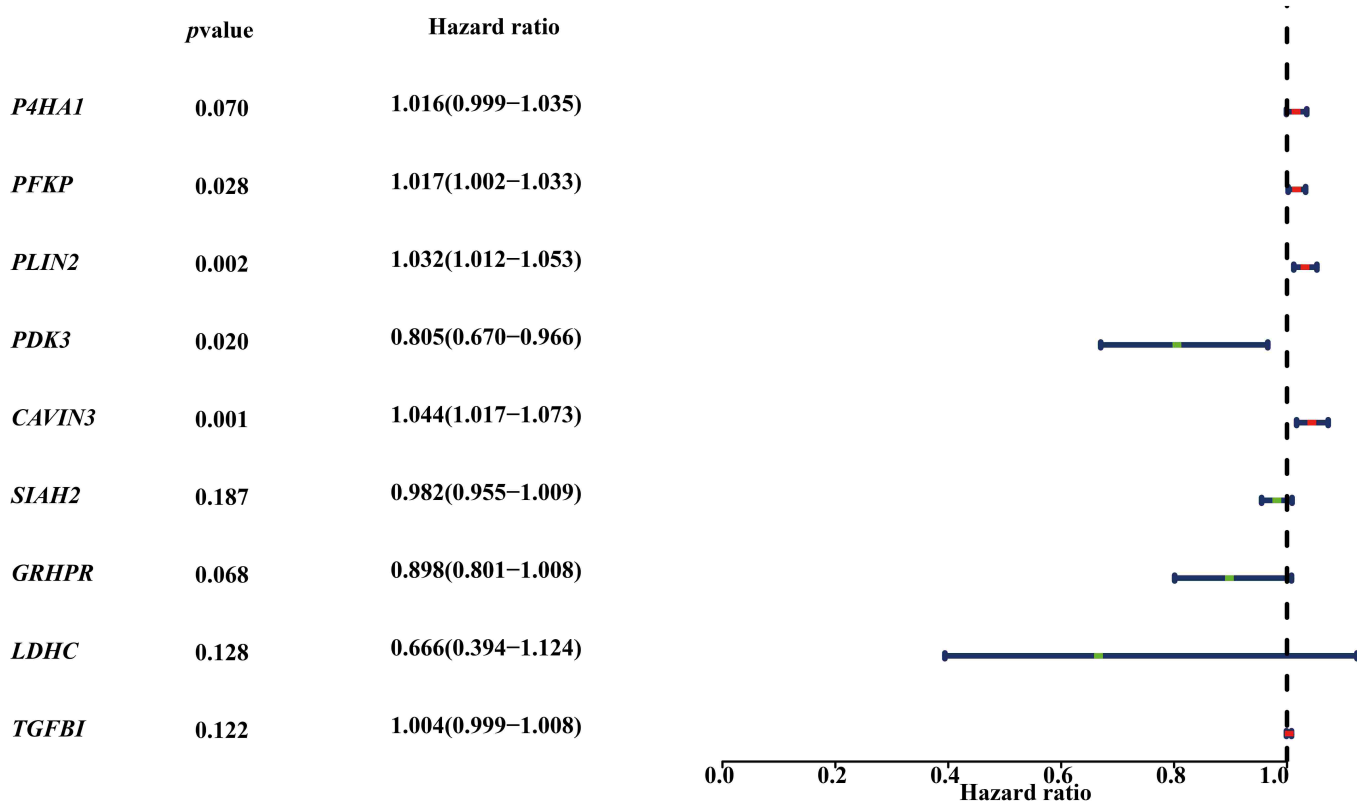


**FIGURE 8. Multivariate Cox proportional hazards regression (Cox regression) analysis to identify prognosis-related genes.**

**TABLE 1. Prognosis-related hypoxia genes identified using multivariate Cox regression analysis.**

| Gene regulation | Gene name | Coeff. | Hazard ratio | Lower 95% CI[a] | Upper 95% CI | *p*-value |
|---|---|---|---|---|---|---|
| Upregulated | | | | | | |
| | *P4HA1* | 0.0163 | 1.0165 | 0.9987 | 1.0346 | 0.0701 |
| | *PFKP* | 0.0173 | 1.0174 | 1.0019 | 1.0332 | 0.0281* |
| | *PLIN2* | 0.0320 | 1.0325 | 1.0123 | 1.0531 | 0.0015* |
| | *CAVIN3* | 0.0435 | 1.0445 | 1.0169 | 1.0728 | 0.0014* |
| | *TGFBI* | 0.0036 | 1.0036 | 0.9990 | 1.0081 | 0.1223 |
| Downregulated | | | | | | |
| | *PDK3* | −0.2174 | 0.8046 | 0.6703 | 0.9658 | 0.0196* |
| | *SIAH2* | −0.0185 | 0.9817 | 0.9551 | 1.0090 | 0.1866 |
| | *GRHPR* | −0.1071 | 0.8984 | 0.8008 | 1.0079 | 0.0678 |
| | *LDHC* | −0.4070 | 0.6656 | 0.3943 | 1.1237 | 0.1277 |

[a] *CI: confidence interval; *p < 0.05. P4HA1: Prolyl 4-Hydroxylase Subunit Alpha 1; PFKP: Phosphofructokinase Platelet; PLIN2: Perilipin 2; CAVIN3: Caveolae Associated Protein 3; TGFBI: Transforming Growth Factor Beta 1; PDK3: Pyruvate Dehydrogenase Kinase 3; SIAH2: Siah E3 Ubiquitin Protein Ligase 2; GRHPR: Glyoxylate And Hydroxypyruvate Reductase; LDHC: Lactate Dehydrogenase C.*

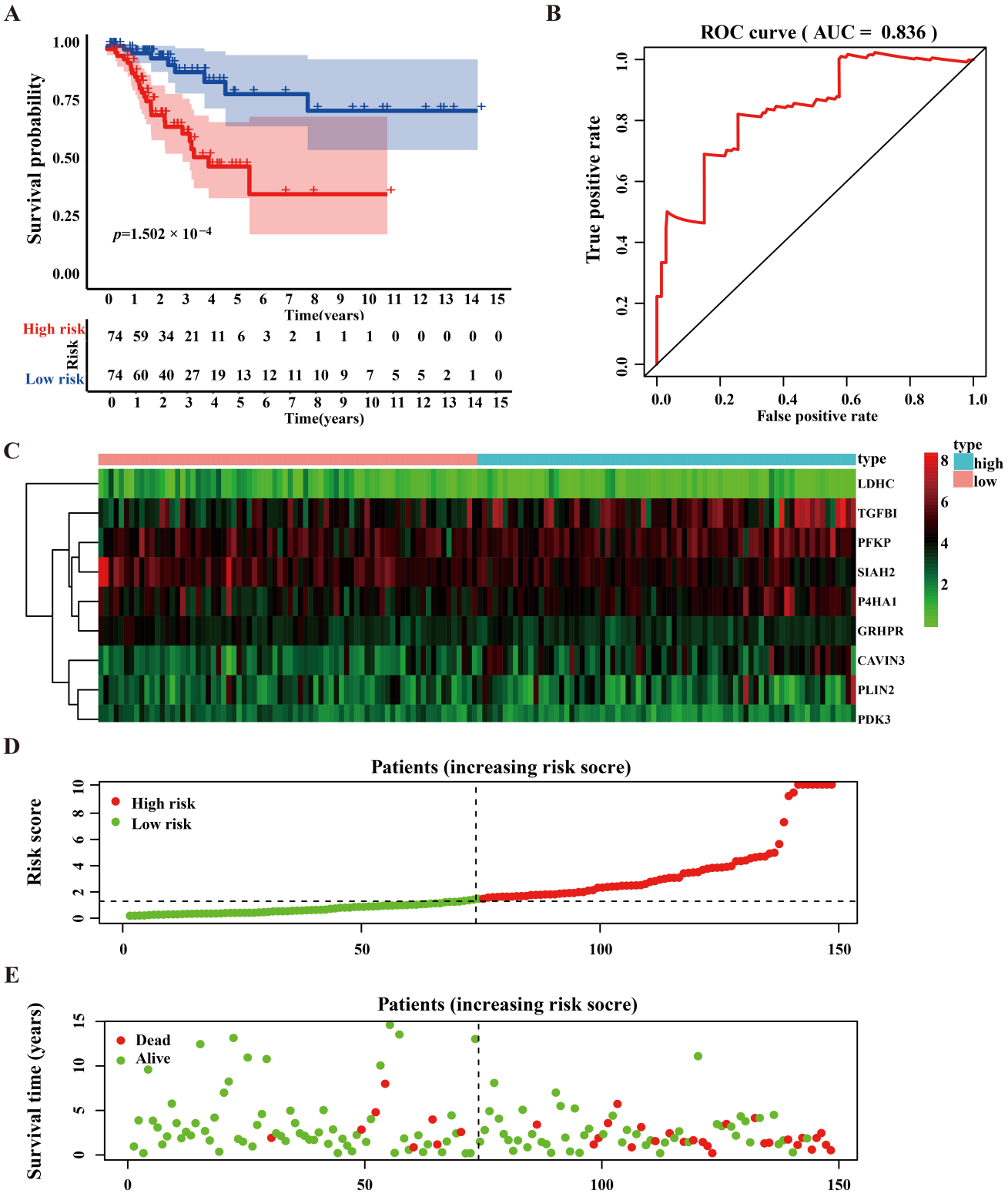


**FIGURE 9. Results of the risk score analysis of the nine-genes prognostic model in the training group.** (A) Survival curves for the low- (blue) and high-risk (red) subgroups; the lower graph depicts absolute patient counts. (B) Receiver operating characteristic (ROC) curve illustrating the predictive overall survival based on a risk scores. (C) Heatmap displaying the expression of the nine hypoxia-related genes, with high expression shown in red and low expression shown in green. (D) Distribution of patient risk scores, ranging from low (green) to high (red). (E) Survival status of test patients categorized by low (green) and high (red) risk scores.

curve score of 0.804 for this hypoxic genetic risk score model (Fig. 10B). Moreover, patients in the high-risk group exhibited worse OS (Fig. 10A,C–E), indicating a good sensitivity and specificity of the prognostic model.

### 3.8 Prognostic value of the hypoxia risk signature in CC

To determine the prognostic significance of different clinical features in patients with CC from the TCGA database, we conducted univariate and multivariate Cox regression analyses. The results revealed risk scores derived from the hypoxia risk signature as independent prognostic factors associated with OS ($p < 0.001$, Fig. 11A–D). These findings highlight the critical role of hypoxia signaling pathway in tumor progression.

### 3.9 Construction of a nomogram based on the nine hypoxia-related genes

We integrated the signature of the nine hypoxia-related genes derived from multivariate Cox analysis to develop a nomogram, a quantitative tool to predict the prognosis of patients with CC. By drawing a vertical line between the total point axis and each pre- and post-axis, we can estimate the anticipated survival rates for patients with CC at 1, 2 and 3 years. This information may be useful for clinical decision-making (Fig. 12).

### 3.10 Validating the expression of the hypoxia-related genes

We used immunohistochemistry (IHC) results from the HPA database to further validate these hypoxia-related genes in CC. Among the nine hypoxia-related genes, seven had available antibody staining data. The analysis revealed that the levels of the proteins encoded by these seven genes, namely P4HA1 (Prolyl 4-Hydroxylase Subunit Alpha), PKFP (Phosphofructokinase Platelet), PDK3 (Pyruvate Dehydrogenase Kinase 3), TGFB1 (Transforming Growth Factor Beta 1), CAVIN3 (Caveolae Associated Protein), SIAH2 (Siah E3 Ubiquitin Protein Ligase) and GRHPR (Glyoxylate and Hydroxypyruvate Reductase) were significantly higher in cancerous tissues compared to healthy tissues (Fig. 13). The IHC results confirmed the significant upregulation of these hypoxia-related genes in tumor tissues compared to normal controls.

### 3.11 Immune cell analysis between high- and low-hypoxia risk groups in CC

Using the CIBERSORT tool, we compared the differences in immune cell infiltration of 22 immune cell types in patients with high and low risk of hypoxia in the TCGA training group (Fig. 14). Our findings revealed notable differences between groups. Specifically, the high-risk group exhibited higher proportions of macrophages (M0) and activated mast cells compared to the low-risk group ($p < 0.05$) (Fig. 14A,B). In contrast, the proportion of resting mast cells was higher in the low-risk group than that in the high-risk group ($p < 0.05$; Fig. 14C).

## 4. Discussion

Single-cell sequencing is a high-throughput technology that enables gene expression profiling at the individual cell level, allowing for the identification of distinct cell subtypes and providing insights into their functional characteristics and intercellular interactions. This approach overcomes the limitations of traditional multicellular transcriptome sequencing technology, which often overlook tumor tissue heterogeneity [18]. In our study, we confirmed the hypoxic status of CC and observed that highly hypoxic cells exhibited stronger intercellular interactions. Using single-cell transcriptomic data of CC from the GEO database, we identified different cell groups and tumor cells. Moreover, GSEA revealed significant variation in hypoxia across cells. Furthermore, by calculating and visualizing the hypoxia pathway activity score for each cell subgroup, we confirmed prominent hypoxia in tumor cells. In particular, tumor cells with a highly hypoxic status showed stronger cell–cell connections with other cells.

The biological functions of the 177 identified DEHLGs were explored with a co-expression and PPI networks. Subsequently, we developed a risk model based on nine DEHLGs to predict prognosis of CC. By analyzing the correlation between the risk model and tumor immune infiltration, we further clarified the relationship between tumor hypoxia and immunity. Our study contributes to a better understanding of the tumor microenvironment and provides potential tools for evaluating the prognosis of CC and identifying therapeutic targets, shedding light on the pathogenesis of CC.

Functional enrichment analyses revealed that the DEHLGs were associated with various metabolic processes, namely canonical glycolysis, glucose catabolic processing to pyruvate, and glycolytic processing through fructose-6-phosphate. KEGG pathway analysis indicated that the DEHLGs play regulatory roles in the occurrence and progression of CC through pathways such as glycolysis/gluconeogenesis and the HIF-1 signaling pathway. Over the past decade, studies have demonstrated that aerobic glycolysis is a hallmark of abnormal tumor energy metabolism [19, 20]. Tumor cells exhibit enhanced glycolytic metabolism under aerobic conditions compared to normal cells, which is known as the Warburg effect. This metabolic switch provides an advantageous supply of material and energy substrate for tumor cell growth, enabling them to survive hypoxic environments. Studies have shown that transketolase like 1 (TKTL1) accelerates the malignant process of CC cells by participating in glycolysis [21]. Myosin 1B (MYO1B) activates the extracellular regulated protein kinases (ERK)/HIF-1A pathway and its downstream glycolysis-related genes, promoting glycolysis, migration and invasion, thus promoting the malignant progression of CC [22].

Through the interaction network of the DEHLGs, we identified 14 hypoxia-related genes with prognostic value following a preliminary screening. Several of these genes were shown to be associated with tumor progression. Further analyses allowed us to identify nine hypoxia-related genes as the most relevant ones with high prognostic value. According to the HPA database, prolyl 4-hydroxylase subunit $\alpha$ 1 (P4HA1), transforming growth factor $\beta$ (TGF$\beta$), and caveolae associated protein 3 (CAVIN3) have been associated with prognostic roles

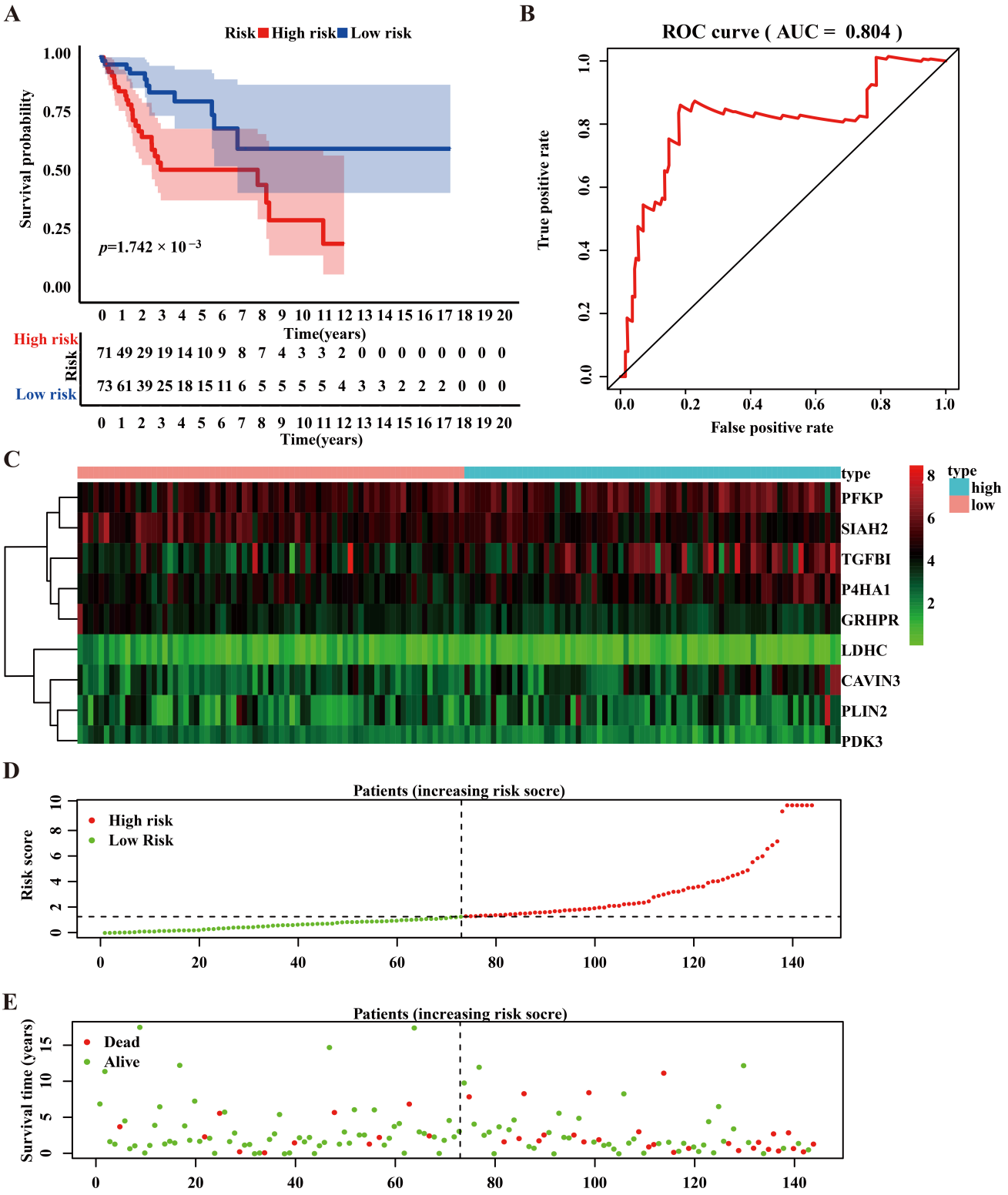


**FIGURE 10. Results of the risk score analysis of the nine-gene prognostic model in the test group.** (A) Survival curves for low- (blue) and high-risk (red) subgroups; the lower graph shows absolute patient counts. (B) Receiver operating characteristic (ROC) curve demonstrating the predictive overall survival based on risk scores; (C) Heatmap illustrating the gene expression of the nine hypoxia-related genes, with high expression depicted in red and low expression shown in green. (D) Distribution of patient risk scores, ranging from low (green) to high (red). (E) Survival status of test patients classified by low (green) and high (red) risk scores. AUC: area under the curve.

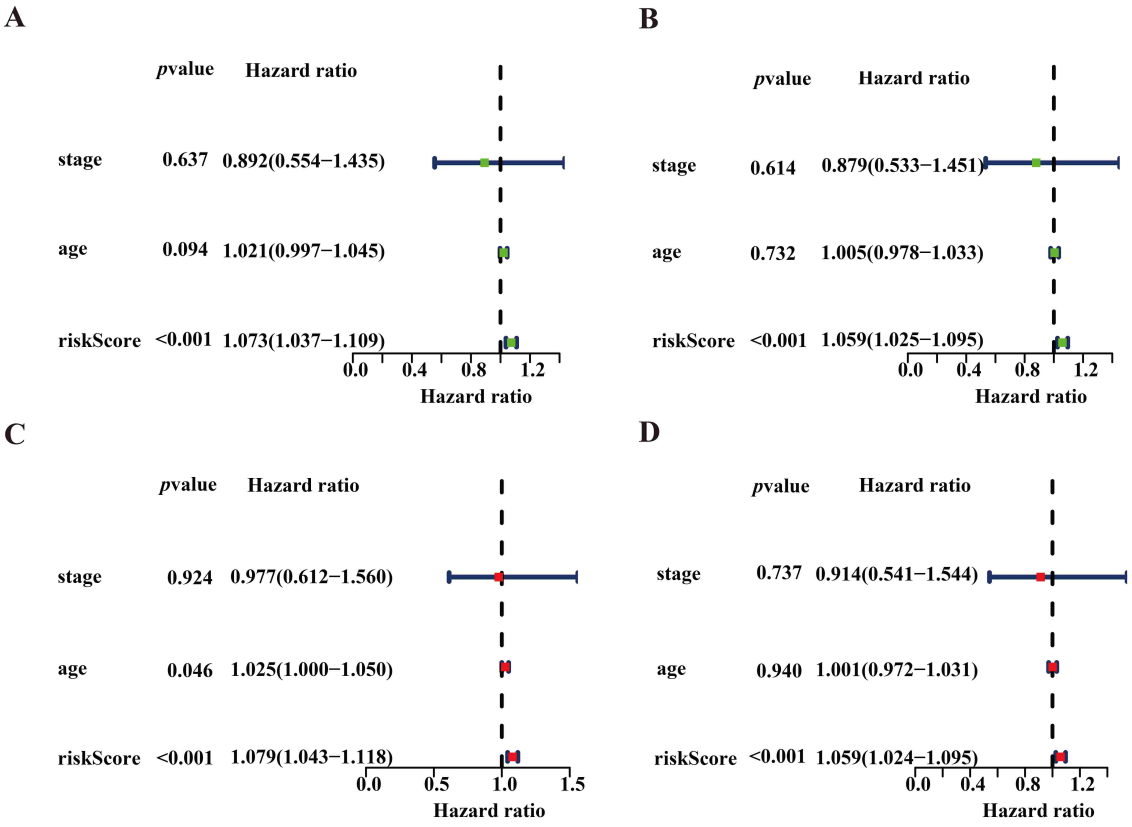


**FIGURE 11. Assessment of the prognostic significance of clinical features including disease stage, age and risk score in patients with cervical cancer.** (A,B) Univariate analysis of the prognostic values of different clinical parameters in the (A) training group and (B) test group. (C,D) Multivariate analysis of the prognostic values of different clinical parameters in the (C) training group and (D) test group.

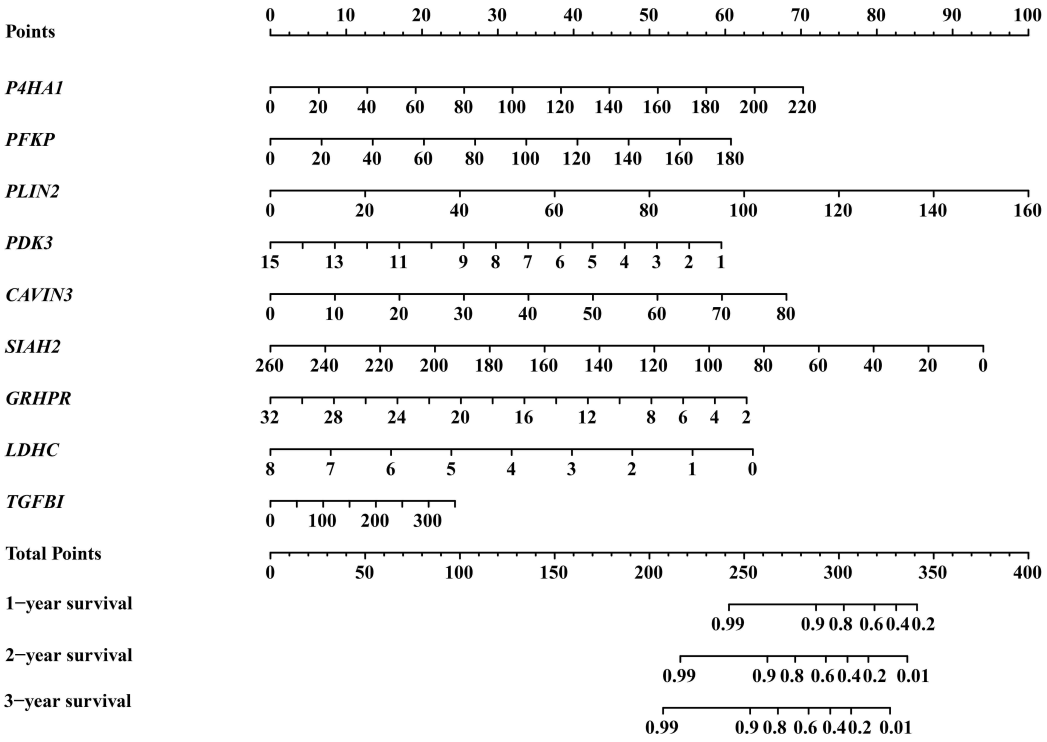


**FIGURE 12. Nomogram for predicting 1-, 2- and 3-year overall survival of patients with cervical cancer in the Cancer Genome Atlas training group.**

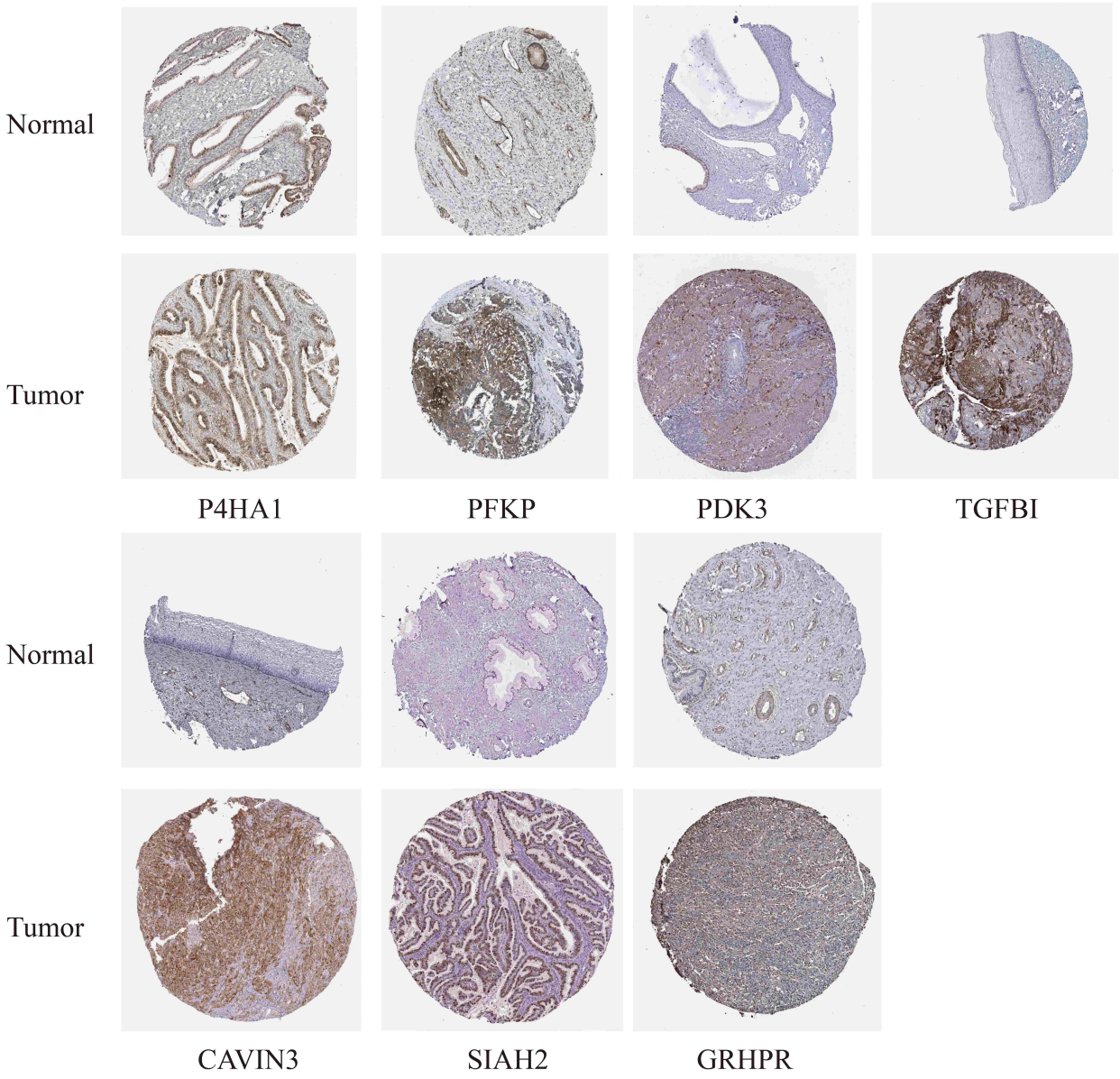


**FIGURE 13. Verification of differentially expressed hypoxia-related genes in cervical cancer (tumor) and normal cervical tissues using data from the Human Protein Atlas (HPA) database.**

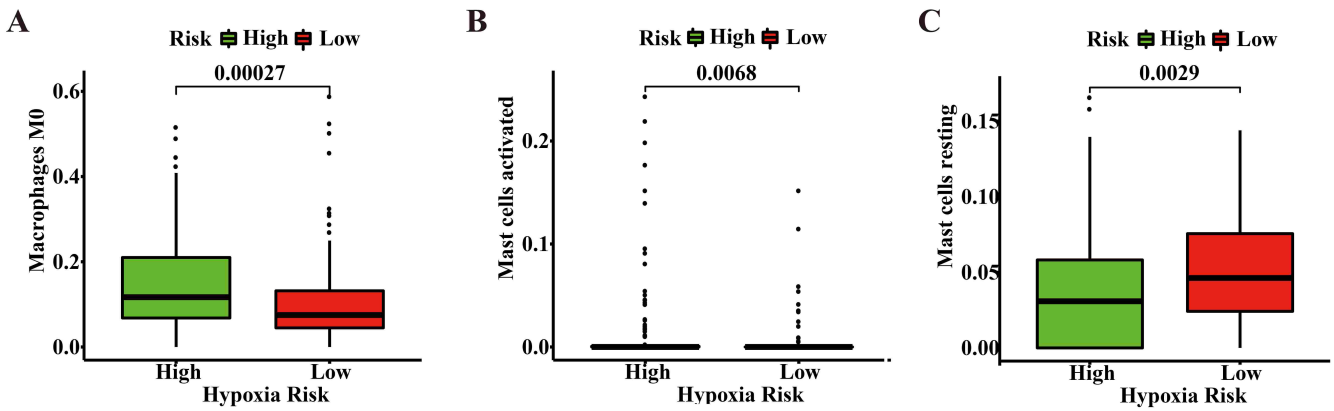


**FIGURE 14. Infiltration rates of immune cells in high- and low-risk groups in the TCGA training group.** Box plots depict the distribution of immune cell infiltration rates, with the high-risk group represented in green and the low-risk group in red. (A) Macrophages (M0). (B) Activated mast cells. (C) Resting mast cells.

and are not highly expressed in CC.

Recent research has highlighted the importance of P4HA1 in the synthesis of various types of collagens. P4HA1 is overexpressed in multiple tumors, including breast, prostate, glioma, colorectal, head and neck cancer. Thus, it may be a potential prognostic biomarker for CC. In breast cancer, *P4HA1* enhances the stemness of cancer cells, reduces the level of reactive oxygen species and oxidative phosphorylation, and promotes chemoresistance by reducing proline hydroxylation on HIF-1A and activating the P4HA/HIF-1 axis [23]. Additionally, in lung adenocarcinoma, P4HA1 has been associated with a poor prognosis, and its abnormal expression related to different numbers of immune infiltrating cells ($CD4^+$ T cells and B cells) [24].

TGF$\beta$ is known to exhibit both tumor-inhibiting and tumor-promoting effects, with differential expression in various types of tumors. Specifically, *TGFBI* is downregulated in breast, ovarian and lung cancers and overexpressed in renal clear cell carcinoma, colorectal, and pancreatic cancer [25]. The role of TGF$\beta$ protein in tumorigenesis depends on the tissue type and tumor microenvironment. During early carcinogenesis, TGFBI exerts a tumor-suppressive effect. Cancer cells use regulatory mechanisms to inhibit *TGFBI* expression, including epigenetic regulation through microRNA and *TGFBI* promoter methylation [26]. However, in most identified tumors, TGFBI promotes cancer progression. For instance, in breast cancer, TGF$\beta$ regulates tumor hypoxia and promotes metastasis by influencing tumor vasculature, blood perfusion improvement, and reducing hypoxia [27]. This is the main driver of cancer stem cells and metastasis. Other studies have demonstrated that *TGFBI* is abundantly expressed in CC tissues and is associated with hypoxia and poor prognosis [28]. Therefore, TGFBI exhibits oncogenic features and has the potential to serve as a marker and potential therapeutic target for CC.

Phosphofructokinase (PFK) is the main rate-limiting enzyme in glucose metabolism, and phosphofructokinase platelet (PFKP) represents the platelet subtype of PFK that contributes to the reprogramming of cancer metabolic pathways. Hypoxia-induced oxidation of ataxia-telangiectasia mutated kinase enhances breast cancer cell invasion by upregulating the protein levels of PFKP and citrate synthase in hypoxic breast cancer cells, redirecting the glycolytic flow to pyruvate and citrate [29]. High *PFKP* expression has been associated with increased mortality in patients with lung cancer [30]. *PFKP* expression was further found to be associated with glycolysis in lung cancer cells, which may influence the development and progression of lung cancer. Silencing *PFKP* can reduce the proliferation of hepatocellular carcinoma cells, their colony-forming ability, stem cell markers, and $\beta$-catenin levels, whereas PFKP overexpression shows the opposite effects. Therefore, PFKP promotes liver cancer cell proliferation and maintains the stability of liver cancer cells [31].

In different studies, *CAVIN3*, perilipin 2 (*PLIN2*), pyruvate dehydrogenase kinase 3 (PDK3), lactate dehydrogenase C (*LDHC*), and E3 ubiquitin-protein ligase 2 (*SIAH2*) have been reported to be associated with the occurrence and development of various tumors and poor patient prognosis [32–36]. These markers are expected to be used as markers for targeted tumor therapy and prognostic evaluation.

The risk model based on the nine hypoxia-related genes, demonstrated high accuracy in predicting the prognosis of CC, making it a valuable tool to select patients with poor prognoses. Furthermore, the construction of a nomogram provides an intuitive way to predict overall survival rates. Analyzing the proportion of immune cells in the high- and low-risk groups, suggested a possible interaction between tumor hypoxia and the immune system. Hypoxia is a prominent feature of the tumor immune microenvironment, recruiting immunosuppressive cells and inhibiting the function of immune cells, thereby promoting tumor progression. Our results suggest that this prognostic model can be advantageous in assisting the clinical treatment of CC.

Several studies have demonstrated the crucial role of multigene signatures in predicting the prognosis of CC [37, 38]. For instance, one study constructed a four immune-related gene model for the prediction of CC prognosis, Xie *et al.* [39] constructed an eight-gene model to predict the prognosis of CC after radiotherapy, while another study established a model based on 70 genes to predict the prognosis of advanced CC [39–41]. Despite these efforts, there is a limited number of CC prediction models based on hypoxia-related genes in the literature. In our study, we sought to optimize the prognosis prediction for CC by combining multiple algorithms to analyze DEHLGs. Moreover, predictions based on only nine genes will greatly reduce sequencing costs and medical resources. The identified genes exhibit important biological functions, highlighting their potential for clinical application in adjuvant therapy. However, it is important to acknowledge that our signature is limited to TCGA database, and its accuracy should be thoroughly validated in clinical cohorts. In addition, due to the small number of single cell samples, there was no good connection between single cell results and bulk RNA results in our study. Furthermore, the biological functions of the prognostic genes need to be confirmed by more prospective experiments.

## 5. Conclusions

In the study, we used single-cell sequencing data to analyze the expression of hypoxia in CC cells and investigated the expression and prognostic value of these DEHLGs in CC. We successfully constructed a prognostic model of nine hypoxia genes, with the potential to be used as an independent prognostic factor for CC. Our findings contribute to a better understanding of the pathogenesis of CC and provide valuable insights for survival prediction and treatment strategies.

### AVAILABILITY OF DATA AND MATERIALS

The datasets analyzed in the present paper are all available on NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi) and TCGA (https://portal.gdc.cancer.gov/). Single-cell transcriptome data from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168652). RNA-sequencing dataset of 13 normal cervical tissue samples and 305 CC samples with their corresponding clinical data from the GTEx (https://gtexportal.org/home/) and TCGA (https://portal.gdc.cancer.gov/) databases.

The MSigDB database contains the hypoxia-related gene set, Hallmark-hypoxia (https://www.gsea-msigdb.org/gsea/index.jsp).

## AUTHOR CONTRIBUTIONS

MMY and CSL—designed the study. TTH and FYL—analyzed the data and wrote the paper. SRC—answered the response. XC and XYT—assisted in the completion of the paper.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

## ACKNOWLEDGMENT

Not applicable.

## FUNDING

This work was supported by the National Natural Science Foundation of China-Youth Science Foundation Project (82002640); the Natural Science Research Project of Jiangsu Province University (20KJB320028); and the Natural Science Foundation of Nanjing University of Traditional Chinese Medicine (XZR2020070).

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

[1] Small W, Bacon MA, Bajaj A, Chuang LT, Fisher BJ, Harkenrider MM, *et al*. Cervical cancer: a global health crisis. Cancer. 2017; 123: 2404–2412.

[2] Zhang XR, Li ZQ, Sun LX, Liu P, Li ZH, Li PF, *et al*. Cohort profile: Chinese cervical cancer clinical study. Frontiers in Oncology. 2021; 11: 690275.

[3] Belli C, Trapani D, Viale G, D'Amico P, Duso BA, Della Vigna P, *et al.* Targeting the microenvironment in solid tumors. Cancer Treatment Reviews. 2018; 65: 22–32.

[4] Cao G, Yue J, Ruan Y, Han Y, Zhi Y, Lu J, *et al*. Single-cell dissection of cervical cancer reveals key subsets of the tumor immune microenvironment. The EMBO Journal. 2023; 42: e110757.

[5] Mennerich D, Kubaichuk K, Kietzmann T. DUBs, hypoxia, and cancer. Trends in Cancer. 2019; 5: 632–653.

[6] Linge A, Löck S, Krenn C, Appold S, Lohaus F, Nowak A, *et al*. Independent validation of the prognostic value of cancer stem cell marker expression and hypoxia-induced gene expression for patients with locally advanced HNSCC after postoperative radiotherapy. Clinical and Translational Radiation Oncology. 2016; 1: 19–26.

[7] Xu H, Yuan Y, Wu W, Zhou M, Jiang Q, Niu L, *et al*. Hypoxia stimulates invasion and migration of human cervical cancer cell lines HeLa/SiHa through the Rab11 trafficking of integrin $\alpha v\beta 3$/FAK/PI3K pathway-mediated Rac1 activation. Journal of Biosciences. 2017; 42: 491–499.

[8] Mustachio LM, Roszik J. Single-cell sequencing: current applications in precision onco-genomics and cancer therapeutics. Cancers. 2022; 14: 657.

[9] Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, *et al*. Integrated analysis of multimodal single-cell data. Cell. 2021; 184: 3573–3587.e29.

[10] Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, *et al*. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nature Immunology. 2019; 20: 163–172.

[11] Huang Q, Liu Y, Du Y, Garmire LX. Evaluation of cell type annotation R packages on single-cell RNA-seq data. Genomics, Proteomics & Bioinformatics. 2021; 19: 267–281.

[12] Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, *et al*. SCENIC: single-cell regulatory network inference and clustering. Nature Methods. 2017; 14: 1083–1086.

[13] Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics. 2013; 14: 7.

[14] Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, *et al*. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nature Protocols. 2020; 15: 2247–2276.

[15] Yu G, Wang L, Han Y, He Q. ClusterProfiler: an R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology. 2012; 16: 284–287.

[16] Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, *et al*. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021; 12: 1088.

[17] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al*. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 2015; 43: e47.

[18] Ni J, Wang X, Stojanovic A, Zhang Q, Wincher M, Bühler L, *et al.* Single-Cell RNA sequencing of tumor-infiltrating NK cells reveals that inhibition of transcription factor HIF-1$\alpha$ unleashes NK cell activity. Immunity. 2020; 52: 1075–1087.e8.

[19] Wang Q, Guo X, Li L, Gao Z, Su X, Ji M, *et al.* $N^6$-methyladenosine METTL3 promotes cervical cancer tumorigenesis and Warburg effect through YTHDF1/HK2 modification. Cell Death & Disease. 2020; 11: 911.

[20] Castelli S, Ciccarone F, Tavian D, Ciriolo MR. ROS-dependent HIF1$\alpha$ activation under forced lipid catabolism entails glycolysis and mitophagy as mediators of higher proliferation rate in cervical cancer cells. Journal of Experimental & Clinical Cancer Research. 2021; 40: 94.

[21] Zhu Y, Qiu Y, Zhang X. TKTL1 participated in malignant progression of cervical cancer cells *via* regulating AKT signal mediated PFKFB3 and thus regulating glycolysis. Cancer Cell International. 2021; 21: 678.

[22] Wen LJ, Hu XL, Li CY, Liu J, Li ZY, Li YZ, *et al.* Myosin 1b promotes migration, invasion and glycolysis in cervical cancer *via* ERK/HIF-1$\alpha$ pathway. American Journal of Translational Research. 2021; 13: 12536–12548.

[23] Xiong G, Stewart RL, Chen J, Gao T, Scott TL, Samayoa LM, *et al.* Collagen prolyl 4-hydroxylase 1 is essential for HIF-1$\alpha$ stabilization and TNBC chemoresistance. Nature Communications. 2018; 9: 4456.

[24] Zhou H, He Y, Li L, Wu C, Hu G. Overexpression of P4HA1 is correlated with poor survival and immune infiltrates in lung adenocarcinoma. BioMed Research International. 2020; 2020: 8024138.

[25] Shang X, Yuan B, Li J, Xi F, Mao J, Zhang C, *et al*. TGFBI is involved in the formation of polyploid cancer cells and the response to paclitaxel. Annals of Translational Medicine. 2021; 9: 693–693.

[26] Corona A, Blobe GC. The role of the extracellular matrix protein TGFBI in cancer. Cellular Signalling. 2021; 84: 110028.

[27] Fico F, Santamaria-Martínez A. TGFBI modulates tumour hypoxia and promotes breast cancer metastasis. Molecular Oncology. 2020; 14: 3198–3210.

[28] Yin R, Zhai X, Han H, Tong X, Li Y, Deng K. Characterizing the landscape of cervical squamous cell carcinoma immune microenvironment by integrating the single-cell transcriptomics and RNA-Seq. Immunity, Inflammation and Disease. 2022; 10: e608.

[29] Peng M, Yang D, Hou Y, Liu S, Zhao M, Qin Y, *et al.* Intracellular citrate accumulation by oxidized ATM-mediated metabolism reprogramming *via* PFKP and CS enhances hypoxic breast cancer cell invasion and metastasis. Cell Death & Disease. 2019; 10: 228.

[30] Shen J, Jin Z, Lv H, Jin K, Jonas K, Zhu C, *et al*. PFKP is highly expressed in lung cancer and regulates glucose metabolism. Cellular Oncology. 2020; 43: 617–629.

[31] Sha X, Wang K, Wang F, Zhang C, Yang L, Zhu X. Silencing PFKP

restrains the stemness of hepatocellular carcinoma cells. Experimental Cell Research. 2021; 407: 112789.

[32] Sun G, Ni K. The role of cavin3 in the progression of lung cancer and its mechanism. BioMed Research International. 2020; 2020: 6364801.

[33] Chen L, Wu Q, Xu X, Yang C, You J, Chen F, *et al.* Cancer/testis antigen LDHC promotes proliferation and metastasis by activating the PI3K/Akt/GSK-3$\beta$-signaling pathway and the in lung adenocarcinoma. Experimental Cell Research. 2021; 398: 112414.

[34] Kuo YH, Chan TC, Lai HY, Chen TJ, Wu LC, Hsing CH, *et al.* Overexpression of pyruvate dehydrogenase kinase-3 predicts poor prognosis in urothelial carcinoma. Frontiers in Oncology. 2021; 11: 749142.

[35] Kurt B, Buendgens L, Wirtz TH, Loosen SH, Schulze-Hagen M, Truhn D, *et al.* Serum perilipin 2 (PLIN2) predicts multiple organ dysfunction in critically ill patients. Biomedicines. 2021; 9: 1210.

[36] Li K, Li J, Ye M, Jin X. The role of Siah2 in tumorigenesis and cancer therapy. Gene. 2022; 809: 146028.

[37] Nie C, Qin H, Zhang L. Identification and validation of a prognostic signature related to hypoxic tumor microenvironment in cervical cancer. PLOS ONE. 2022; 17: e0269462.

[38] Yang Y, Li Y, Qi R, Zhang L. Constructe a novel 5 hypoxia genes signature for cervical cancer. Cancer Cell International. 2021; 21: 345.

[39] Xie F, Dong D, Du N, Guo L, Ni W, Yuan H, *et al*. An 8-gene signature predicts the prognosis of cervical cancer following radiotherapy. Molecular Medicine Reports. 2019; 20: 2990–3002.

[40] Mei J, Xing Y, Lv J, Gu D, Pan J, Zhang Y, *et al*. Construction of an immune-related gene signature for prediction of prognosis in patients with cervical cancer. International Immunopharmacology. 2020; 88: 106882.

[41] Nguyen NNY, Choi TG, Kim J, Jung MH, Ko SH, Shin Y, *et al*. A 70-gene signature for predicting treatment outcome in advanced-stage cervical cancer. Molecular Therapy—Oncolytics. 2020; 19: 47–56.

**How to cite this article:** Tingting He, Xiaoyu Tang, Siru Chen, Xin Chen, Fuye Lin, Minmin Yu, *et al*. Comprehensive analysis of hypoxia-related gene signature in cervical cancer. European Journal of Gynaecological Oncology. 2023; 44(6): 105-121. doi: 10.22514/ejgo.2023.105.