Submitted: 19 November, 2022 Accepted: 03 April, 2023 Published: 15 October, 2024 DOI:10.22514/ejgo.2024.093

ORIGINAL RESEARCH

European Journal of Gynaecological Oncology

Exploration of prognostic genes in cervical cancer immune microenvironment

Chunli Fang[1,*,†], Ya Zhu[2,†], Feifei Hu[1], Hailin Chen[1], Huajing Xiao[1], Jie Ding[1], Boqun Xu[1]

[1]Department of Gynecology, The Second Affiliated Hospital of Nanjing Medical University, 210009 Nanjing, Jiangsu, China
[2]Department of Oncology, The Second Affiliated Hospital of Nanjing Medical University, 210009 Nanjing, Jiangsu, China

***Correspondence**
fangchunli@njmu.edu.cn
(Chunli Fang)

† These authors contributed equally.

Abstract

Advanced cervical cancer (CESC) is a common gynecological malignancy that threatens females' lives and existing treatments remain ineffective. This study focused on discovering potential biomarkers for the prognosis of CESC patients and exploring potential curative mechanisms and possible therapeutic directions. Based on the gene expression profile and CESC patient survival data in The Cancer Genome Atlas (TCGA) database, prognostic signatures were obtained and using Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses to understand related gene functions, functional and pathway enrichment of potential biomarkers. Gene set enrichment analysis and gene set variation analysis were performed to understand further the biological characteristics and regulatory networks of potential biomarker expression levels on the prognosis of CESC. Immune infiltration analysis to comprehend the functional correlation between potential biomarkers and immune cells and the receiver operating characteristic (ROC) curves to assess the degree of CESC differentiation identification by prognostic signature and hazard score in predicting patient prognosis. Brain-Specific Angiogenesis Inhibitor 1-Associated Protein 2-Like Protein 1 (*BAIAP2L1*) and secreted phosphoprotein 1 (*SPP1*) (hazard ratio (HR) >1) were highly expressed in tumor tissues as CESC risk factors, while *DES* (Desmin) and EF-Hand Calcium Binding Domain 1 (*EFCAB1*) (HR <1) were expressed at low levels in tumor tissues as prognostic protective factors. *SPP1* had the highest positive correlation with Macrophages M0 and *IL3RA* had the highest positive correlation with T cells CD8. Patients with high expression of the risk factor gene *SPP1* had a significant reduction in overall survival (OS) time, whereas those with high expression of the protective factor gene interleukin-3 receptor alpha chain (*IL3RA*) had a long OS time. Our results showed that the prognostic signature, either risk or protective factors, have a great effect on the prognosis of CESC, and understanding the role of these genes in the development of CESC may provide new directions.

Keywords

Cervical cancer; Prognosis; TCGA; Differentially expressed genes (DEGs); *BAIAP2L1*; *SPP1*; *IL3RA*

1. Introduction

Cervical cancer (CESC) remains one of the most serious cancers among females, with nearly 5,700,000 females suffering from the disease, causing more than 310,000 deaths in 2018 (Bray F, Ferlay J, *et al.* [1], 2018). Despite preventive measures, missed opportunities for vaccination and screening are directly correlated with CESC development (Buskwofie A, David-West G, *et al.* [2], 2020). Surgery is the main treatment for early-stage cancer, and radio- and/or chemotherapy has been used to treat patients with advanced or recurrent CESC with limited success (Kagabu M, Nagasawa T, *et al.* [3], 2020).

The 5-year survival rate of advanced CESC, which accounts for nearly one-third of all stages, is approximately 40–50%, and some patients have a poor prognosis when they first present with metastasis (Marquina G, Manzano A, *et al.* [4], 2018). Therefore, increasing attention has been paid to research effective prognostic biomarkers, new molecular mechanisms, and therapeutic directions in CESC, especially in the late stages.

The tumor microenvironment (TME) plays a crucial role in tumor development, and tumor-infiltrating immune cells (TICs) are important components of the TME (Peng L, Hayatullah G, *et al.* [5], 2021). The interaction between tumor cells and TICs ultimately leads to a process that may accelerate tumor growth and invasion, such as immune escape, resistance to apoptosis, and abnormal proliferation (Hinshaw DC, Shevde LA [6], 2019). Hence, analysis to identify dynamic changes in the TME at the genetic level may reveal the underlying mech-

anisms of CESC tumorigenesis and progression and identify new therapeutic targets (Xu F, Shen J, *et al.* [7], 2021).

In the current study, as is shown in Fig. 1, we designed a prognostic signature utilizing the prognosis-related gene expression profile and survival data of CESC from The Cancer Genome Atlas (TCGA). Utilizing the prognostic signature, we obtained differentially expressed genes (DEGs); functional and pathway enrichment analyses were performed to explore the function of these DEGs. Next, DEGs were investigated concerning the characteristics of different immune cells. Furthermore, risk factors were scored and evaluated using the survival and clinical staging data of CESC from the TGCA. Finally, we validated the prognostic signature using GSE7803 and GSE63514, evaluated hazard scores to differentiate the prognosis of patients with CESC, and analyzed the expression of potential biomarkers in patient tissues to explore the mechanisms underlying the induction and promotion of CESC. These results may help improve patient management and enable personalized treatment.

2. Materials and methods

2.1 Data and DEGs acquisition

Gene expression profile data of CESC were obtained from the TCGA database (https://portal.gdc.cancer.gov/) using the R package TCGAbiolinks (https://www.bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html, Colaprico A, *et al.* [8], 2016), and after data cleaning, they were divided into 296 tumor tissue and 3 normal tissue gene expression groups (Table 1). To explore the prognosis-related genes of CESC in more depth, we obtained the survival data of CESC patients from the TCGA database, identified the prognosis-related genes using one-way Cox regression analysis, constructed a prognostic signature using Least absolute shrinkage and selection operator (LASSO) and multifactor Cox regression analyses, and validated the degree of correlation between the signature and the prognosis of CESC.

Using the R package GEOquery (https://www.bioconductor.org/packages/release/bioc/html/GEOquery.html, Davis S, P.S. Meltzer [9], 2007), we downloaded the gene expression profile data of GSE7803 (Zhai Y, Kuick R, *et al.* [10], 2007) from the GEO database (Barrett T, *et al.* [11], 2007). The data were divided into 21 tumors and 10 normal tissues. By choosing the human species and the GPL570 platform, we obtained another data group, GSE63514 (den Boon JA, Pyeon D, *et al.* [12], 2015), and divided them into 28 tumors and 24 normal tissues (Table 1). Differential group analyses were conducted using the R package limma (Ritchie ME, *et al.* [13], 2015). We set the setoff as upregulated with logFC (foldchange) >1 and adj (adjusted) p-value < 0.05 and downregulated with logFC <-1 and adj$p$-value < 0.05. The DEGs obtained were analyzed for differences in expression.

2.2 Function of DEGs and pathway enrichment analysis

Based on this prognostic signature, we classified the tumor dataset into 112 A-types and 184 B-types using the R package limma (https://www.bioconductor.org/packages/release/bioc/html/limma.html, Ritchie ME, *et al.* [13], 2015) for differential analysis between groups. We set genes as upregulated with logFC >1 and adj$p$-value < 0.05, and downregulated with logFC <-1 and adj$p$-value < 0.05.

We performed functional and pathway enrichment analyses to determine the functions of these potential biomarkers. Gene Ontology (GO) analysis (Gene Ontology Consortium [14], 2015) is a common method for large-scale functional enrichment studies of genes, including biological processes, molecular function, and cellular components. The Kyoto Encyclopedia of Genes and Genomes (KEGG) is widely used to store information about genomes, biological pathways, diseases, and drugs (Ogata H, Goto S, *et al.* [15], 1999, Kanehisa M, Goto S. [16], 2000, Wixon J, Kell D. [17], 2000). GO and KEGG analyses of DEGs related to CESC were performed using the R package clusterProfiler (https://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html, Yu G, *et al.* [18], 2012), and p-value < 0.05 was considered statistically significant in this study.

2.3 Gene set enrichment analysis

In a predefined gene set, Gene Set Enrichment Analysis (GSEA) can be used to evaluate the genes' distribution trends in a table of genes ordered by their phenotypic correlation and thus ascertain their contribution to the phenotype (Subramanian A, *et al.* [19], 2005). We obtained the "c2.kegg.v7.4. and h.all.v7.0. symbols" gene sets from the MSigDB database (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp, Liberzon A, *et al.* [20], 2015) to perform GSEA on both datasets separately. GSEA was performed using the "clusterProfiler" R package (Yu G, *et al.* [18], 2012), and a p-value < 0.05 was considered statistically significant.

2.4 Immuno-infiltration analysis of the TCGA-CESC dataset

The degree of infiltration of cells of the tumor microenvironment, as well as immune and stromal cells in the tumor, has a significant impact on prognosis. To better understand the prognostic impact of genes involved in immune and stromal cells, the ESTIMATE package (Yoshihara K, *et al.* [21], 2013) exploits the unique nature of the transcriptional profile of cancer samples to infer the content of tumor cells as well as differentially infiltrating normal cells. We use the ESTIMATE package to calculate immune and mechanistic scores on expression data from the TCGA-CESC dataset and thus assess tumor purity, based on the ESTIMATE algorithm by calculating the characteristics of the tumour expression matrix at Immune Score (Immune Score) and Stromal Score (Stromal Score), ESTIMA Score (ESTIMATE Score) and Tumor Purity to quantify the immune and stromal components of the

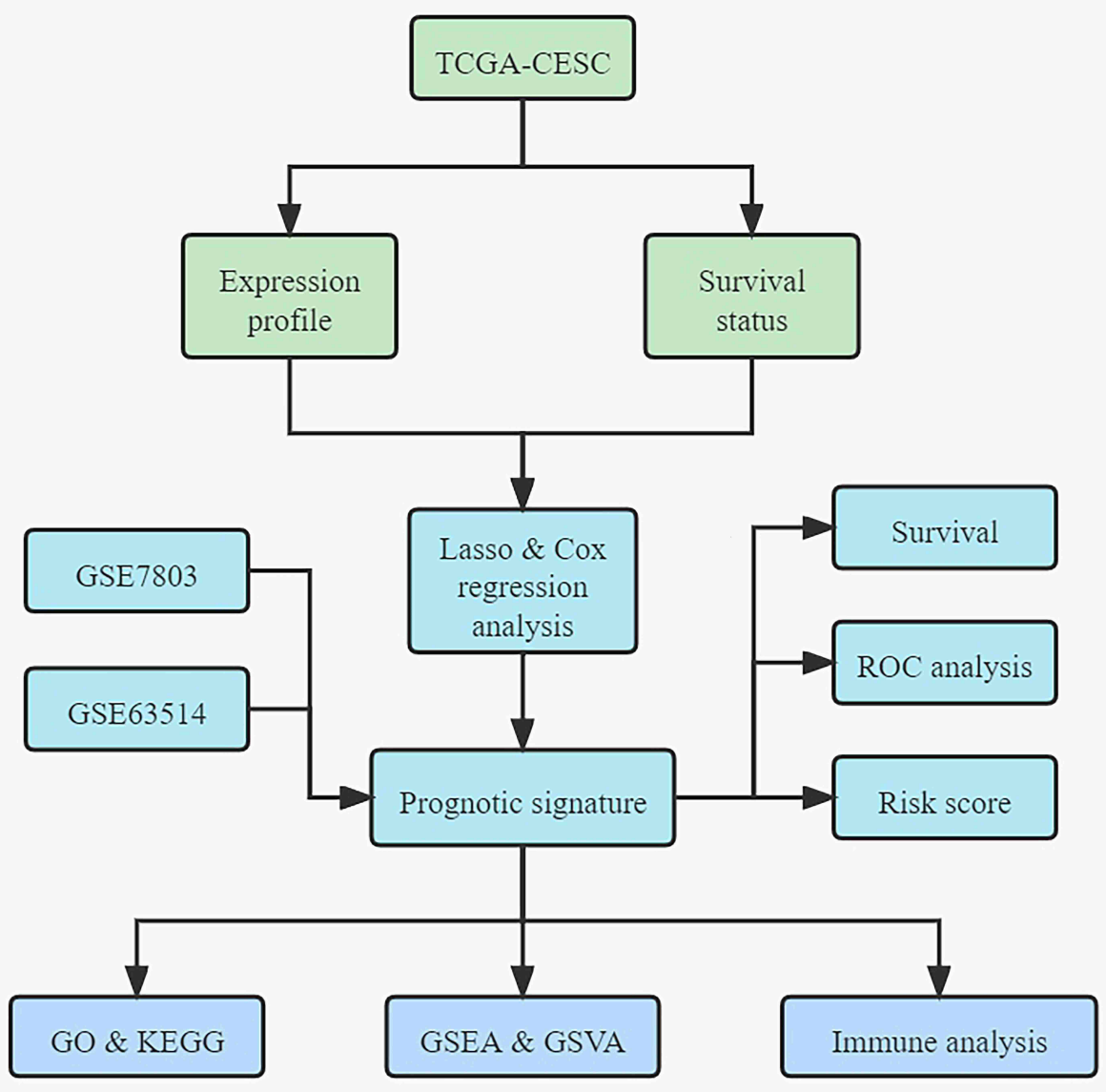


FIGURE 1. Flow chart. TCGA-CESC, The Cancer Genome Atlas-Cervical cancer; GSE, Gene Expression Omnibus Series; GO&KEGG, Gene Ontology & Kyoto Encyclopedia of Genes and Genomes; GSEA & GSVA, Gene Set Enrichment Analysis & Gene Set Variation Analysis.

TABLE 1. Data information sheet.

Data	Normal	Cancer
TCGA-CESC	3	296
GSE7803	10	21
GSE63514	24	28

TCGA: The Cancer Genome Atlas; CESC: cervical cancer.

tumor. The differences in immune scores, stromal scores and ESTIMA scores were investigated for model risk groups (high risk group: high, low risk group: low).

Single sample gene set enrichment analysis (ssGSEA) (Barbie DA, *et al.* [22], 2009) The algorithm, which allows estimation of the number of specific immune infiltrating cells and the activity of specific immune responses, was derived from the research content of published articles on tumor immune infiltration to obtain 28 gene sets used to mark different tumor infiltrating immune cell types, including various human immune cell subtypes: CD8+ T cells, dendritic cells, macrophages, regulatory T cells, *etc*. regulatory T cells, *etc*. Enrichment scores calculated by analysis of the ssGSEA algorithm in the R package GSVA (Gene Set Variation Analysis) package were used to represent the level of infiltration of each immune cell type in each sample, and correlations between immune infiltrating cells and between immune cells and DEGs were determined by Spearman correlation analysis.

2.5 Immuno-infiltration analysis (CIBERSORT)

CIBERSORT (https://cibersort.stanford.edu/) (Newman AM, *et al.* [23], 2019) is an R/web-based tool for deconvolution of expression matrices of human immune cell subtypes based on the principle of linear support vector regression (LSVR). It can assess the infiltration status of immune cells in sequenced samples based on the set of gene expression features of 22 known immune cell subtypes. In this study, the immune cell infiltration status of the dataset was evaluated by the CIBERSORT algorithm and the interrelationships between the various immune cells were calculated by Spearman correlation. Subsequently, we determined the correlation between DEGs and immune infiltrating cells by Spearman correlation analysis and visualized them using the R package "ggplot2" to create heat maps and lollipop plots.

2.6 Diagnostic and prognostic analysis of tumors by prognostic signature

Gene expression levels and clinical characteristics were correlated with patient prognosis. To further evaluate the impact of the prognostic signature on tumor diagnosis and patient prognosis, we first analyzed the ROC curves of *IL3RA*, secreted phosphoprotein 1 (*SPP1*), and *EFCAB1* in the model. We then analyzed the subgroup Kaplan-Meier curves in the clinical stages III and IV groups. The log-rank test was used for statistical analysis, and a p-value < 0.05 was considered as statistically significant.

2.7 Risk factor analysis in CESC

We used survival data and clinical stages of CESC patients in the TCGA database to score risk factors and drew Kaplan-Meier curves and nomogram diagrams based on risk scores to evaluate risk score efficacy.

2.8 Immunohistochemical analysis of DEGs in CESC

The Human Protein Atlas (HPA) database (www.proteinatlas.org/) is based on proteomic, transcriptomic, and systems biology data and can be used to map tissues, cells, organs, *etc*. It covers tumor and normal tissues for protein expression. This database was used to evaluate the expression of potential biomarkers in the tissues of CESC patients (Sjöstedt E, Zhong W, *et al.* [24], 2020).

2.9 Statistical analysis

All data calculations and statistical analyses were performed using R software (https://www.r-project.org/, version 4.0.2). Independent Student's t-tests were used to calculate the differences between normally distributed variables, and Mann-Whitney U-tests were used to calculate the differences between non-normally distributed variables to compare two groups of continuous variables. ROC curves were plotted using R's pROC package, and the area under the curve was calculated to estimate the accuracy of the risk score for prognosis. All bilateral p-values were statistically significant at $p < 0.05$.

3. Results

3.1 Design of CESC prognostic model

We performed LASSO regression on the cleaned TCGA dataset (Fig. 2A–C) and screened for CESC-related prognostic markers using Cox multi-factor analysis (Table 2). The results yielded 22 key genes, and the results were plotted as forest plots (Fig. 2D), which showed high prognostic efficacy for most genes. There were 18 genes screened for $p < 0.05$, including 10 genes with a risk ratio (HR >1) and eight genes with a risk ratio (HR <1).

Next, we selected genes with HR >1 and HR <1 to plot survival curves (Fig. 3). The results showed that these genes' expression levels greatly impacted the CESC patients' prognosis, and the trends were consistent with the initially obtained results. The differential expression results of potential biomarkers are shown as box plots (Fig. 4A–D). As risk factors for the prognosis of CESC patients, *BAIAP2L1* (BAI1-associated protein 2-like 1, IRTKS) and *SPP1* (HR >1) were highly expressed in tumor tissues compared with para-carcinoma. As protective prognostic factors of CESC patients, *DES* and *EFCAB1* (HR <1) were lower in tumor tissues than in para-carcinoma.

3.2 Design of CESC prognostic subtypes

To better demonstrate the expression of genes associated with CESC prognosis, we used the "ConsensusClusterPlus" (Wilkerson MD, Hayes DN [25], 2010) R package to perform consistent clustering on the TCGA CESC dataset using the above prognostic model with a number of subtype parameters of two, which can classify the samples well. We then used the obtained model genes to group the genes using CESC prognostic subtypes and plotted the heat map (Fig. 4E). The results showed a trend toward differences in gene expression. We obtained DEGs from the two groups using limma package differential analysis. We set the genes as upregulated with logFC >1 and adj$p$-value < 0.05 and downregulated with logFC <-1 and adj$p$-value < 0.05. In total, 2381 DEGs were identified.

3.3 Functional enrichment analysis of DEGs

To analyze the relationship between DEGs and their biological processes, molecular functions, cellular components, biological pathways, and diseases, we performed a functional enrichment analysis of DEGs associated with CESC prognosis (Fig. 5A, Tables 3 and 4). The DEGs were mainly enriched in biological processes such as extracellular matrix (ECM) organization, extracellular structure organization, ossification, collagen fibril organization, collagen metabolic process, collagen-containing ECM, endoplasmic reticulum lumen, collagen trimer, basement membrane, fibrillar collagen trimer, ECM structural constituent, glycosaminoglycan binding, collagen binding, heparin binding, and other molecular functions. Pathway enrichment analysis of DEGs showed that they were enriched in focal adhesion, the

TABLE 2. Multi-factor cox analysis.

gene	HR	HR.95L	HR.95H	*p* value
PRND	9.533687429	3.770321238	24.10701642	1.90×10^{-6}
BAIAP2L1	2.968129900	1.796483673	4.903910474	2.17×10^{-5}
DES	0.794385928	0.64925279	0.971961942	0.025335807
PPP1R14A	1.431543187	1.100511884	1.862147901	0.007500269
PTPRB	1.659247056	1.137518183	2.420269701	0.008565775
SERPINF1	0.641406826	0.481112007	0.85510798	0.002471436
ROR1	1.896527765	1.356314551	2.651905164	0.000182764
ERG	1.364394707	1.016130555	1.832021395	0.038792207
EFCAB1	0.448711782	0.2618933	0.768795015	0.003533628
IL3RA	0.613499277	0.429047693	0.877248307	0.00741192
PRKCDBP	1.340518197	1.044954683	1.719681308	0.021111773
REEP1	0.718225557	0.546411952	0.944064178	0.017663603
HIST1H3G	0.701053775	0.531012215	0.925546309	0.01221603
SDS	0.605572144	0.445293321	0.823541706	0.001385605
LHX2	0.680254019	0.530885958	0.871647711	0.002319355
SPP1	1.528004078	1.238546563	1.885109961	7.61×10^{-5}
CPE	1.275021287	1.074468943	1.513007233	0.005393059
REG1A	1.227243246	1.058431854	1.422978703	0.006685745

HR: hazard ratio; PRND: Prion-like protein doppel; BAIAP2L1: Brain-Specific Angiogenesis Inhibitor 1-Associated Protein 2-Like Protein 1; DES: Desmin; PPP1R14A: protein phosphatase 1, regulatory (inhibitor) subunit 14A; PTPRB: protein tyrosine phosphatase, receptor type B; SERPINF1: serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1; ROR1: receptor tyrosine kinase like orphan receptor 1; ERG: erythroblast transformation-specific (ETS) transcriptions factors; EFCAB1: EF-hand calcium binding domain 1; IL3RA: interleukin-3 receptor alpha chain; PRKCDBP: protein kinase C, delta binding protein; REEP1: Homo sapiens receptor accessory protein 1; HIST1H3G: histone cluster 1, H3g; SDS: shwachman-bodian-diamond ribosome maturation factor; LHX2: Homo sapiens LIM homeobox 2; SPP1: secreted phosphoprotein 1; CPE: Clostridium perfringens enterotoxin; REG1A: regenerating islet-derived 1 alpha.

PI3K-Akt (phosphatidylinositol-3-kinase-protein kinase B) signaling pathway, human papillomavirus infection, protein digestion and absorption, ECM-receptor interaction, and other biological pathways (Fig. 5B), while the most remarkably enriched pathway was hsa04510: focal adhesion (Fig. 5C).

3.4 GSEA

To validate the prognosis of CESC affected by different gene expression levels, we analyzed the association between gene expression and the biological processes, cellular components affected, and molecular functions performed in the two datasets (Table 5). The results showed that prognosis-related differential genes mainly affected graft versus host disease, pathways in cancer, vascular smooth muscle contraction, focal adhesion, hypertrophic cardiomyopathy, dilated cardiomyopathy, arrhythmogenic right ventricular cardiomyopathy, ECM-receptor interaction, and other pathways (Fig. 6A), as well as typical results of ECM-receptor interaction (Fig. 6B), focal adhesion (Fig. 6C), and dilated cardiomyopathy (Fig. 6D). There were also related functions with apoptosis, KRAS (Kirsten ratsarcoma viral oncogene homolog) signaling up, apical junction, UV (ultraviolet) response DN (Genes down), myogenesis, coagulation, and angiogenesis, epithelial mesenchymal transition *et al.* (Fig. 6E) and apical junction (Fig. 6F), coagulation (Fig. 6G), and epithelial mesenchymal transition (Fig. 6H).

3.5 Gene set variation analysis

To examine the gene set enrichment results, Gene Set Variation Analysis was used to transform the gene expression matrices between samples into gene set expression matrices to assess whether different metabolic pathways are enriched between samples (Table 6). The results were visualized using the pheatmap package (Fig. 7A,B). We found that sample grouping could distinguish the GSEA results.

3.6 Immunoscore and immune-infiltration analysis of the TCGA-CESC dataset

We used the R package ESTIMATE to calculate various immune and matrix scores for the TCGA-CESC dataset using the collated TCGA-CESC dataset expression spectrum data, and after processing the results we obtained the Stromal Score, Immune Score, ESTIMATE Score. Combining the information from the TCGA-CESC dataset with the high and low risk groupings (low group, high group), we present the results of the Stromal Score, Immune Score, and ESTIMATE Score scores in group comparison plots (Fig. 8A–C), which show that the Immune Score of the TCGA-CESC dataset Score (Fig. 8B) was statistically significant ($p < 0.05$) between the different subgroups of the TCGA-CESC dataset sample.

We collated the expression profiles from the TCGA-CESC dataset and used the ssGSEA algorithm to calculate the correlation between the expression profiles of 28 immune cells

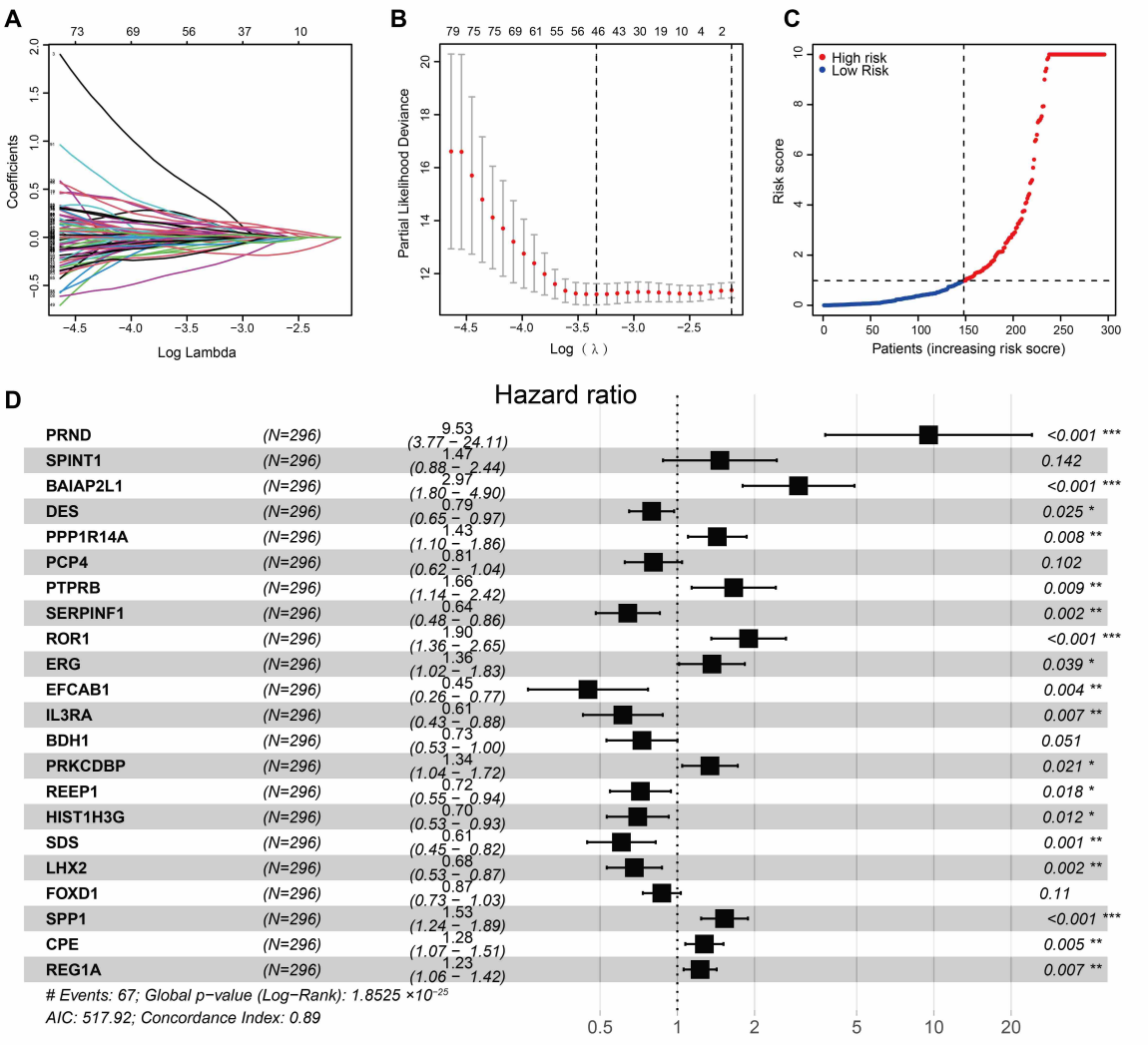


FIGURE 2. Construction of CESC prognostic model. (A–C) lasso regression + Cox multifactor regression analysis on TCGA dataset, lasso regression (A), optimal λ value (B), and risk profile (C). (D) forest plot showing the obtained prognostic model genes. CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma; Lasso: Least absolute shrinkage and selection operator; TCGA: The Cancer Genome Atlas; "***": $p < 0.001$; "**": $p < 0.01$; "*": $p < 0.05$.

and the high and low risk subgroups (low group, high group) from the TCGA-CESC dataset. We plotted group comparisons (Fig. 8D) based on differences in the degree of infiltration of immune cells in the low group, the high group, and for those immune cells showing significant ($p < 0.05$) differences, we analyzed the correlations between immune cells and DEGs (*SPP1*, *IL3RA*) and selected combinations with significant ($p < 0.05$) correlations for heat map display. The results show that a total of 11 immune cells (Activated B cell, Activated CD8 T cell, CD56dim natural killer cell, Central memory CD8 T cell, Effector memory CD8 T cell, Eosinophil Immature B cell, Immature dendritic cell, Memory B cell, Monocyte, Neutrophil) were significantly ($p < 0.05$) infiltrated differently in the low and high groups.

3.7 Immuno-infiltration analysis (CIBERSORT)

We collated the TCGA-CESC expression profiles and uploaded them to the CIBERSORT online website, using the CIBERSORT algorithm to calculate the correlation between the expression profiles of 22 immune cells and the high and low groups of the dataset. We first plotted a histogram of the infiltration abundance of immune cells in the sample (Fig. 9A). We then analyzed the differences between immune cells in the high and low groups according to the infiltration abundance of immune cells (Fig. 9B), where immune cells T cells CD8, T cells CD4 naive, T cells CD4 memory activated, Monocytes, Macrophages M0, Dendritic cells resting, Dendritic cells activated, Mast cells resting, Mast cells activated, Neutrophils were significantly ($p < 0.05$) infiltrated differently in the High and Low groups. We plotted the correlation heat map of these 10 immune cells (Fig. 9C) and the correlation heat map of DEGs (*SPP1*, *IL3RA*) with the 10 immune cells (Fig. 9D). The results showed that *SPP1* had the highest positive correlation with Macrophages M0 and *IL3RA* had the highest positive correlation with T cells CD8.

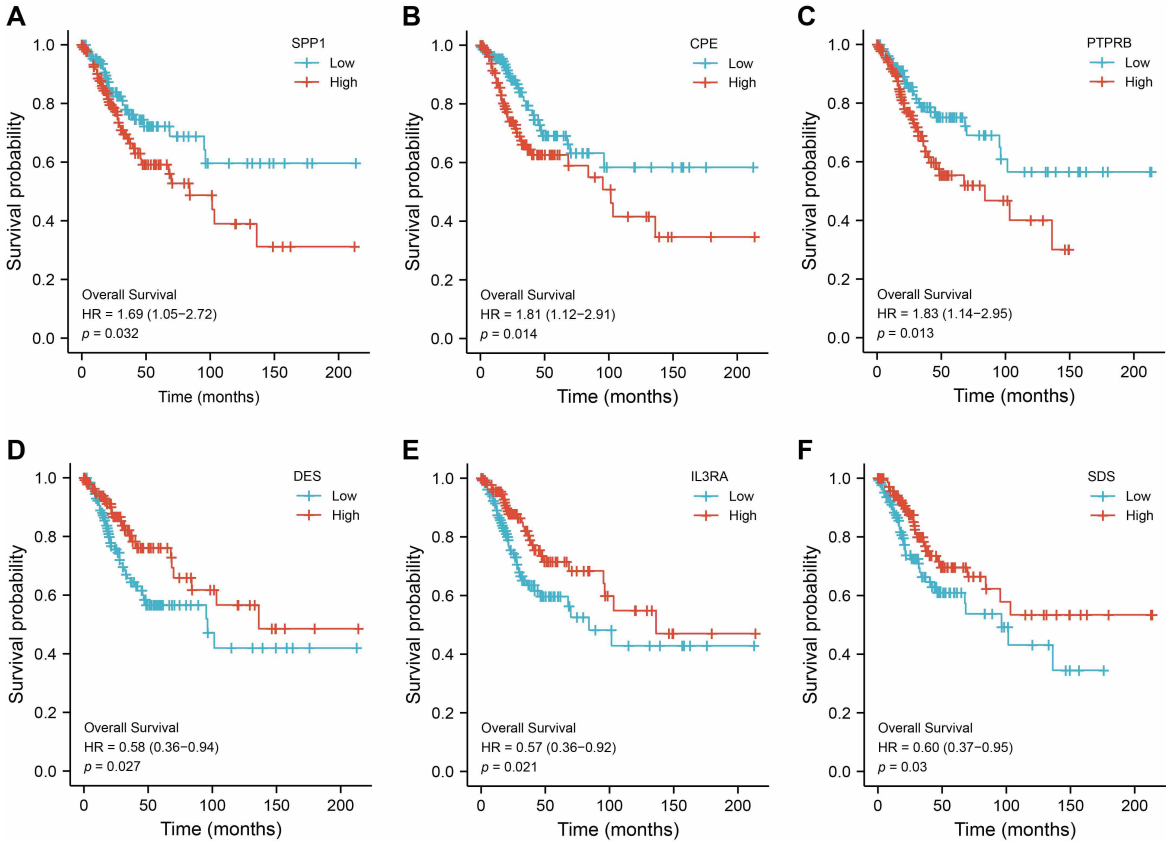


FIGURE 3. Survival analysis of prognosis-related genes. (A) KM survival curve of *SPP1*. (B) KM survival curve of *CPE*. (C) KM survival curve of *PTPRB*. (D) KM survival curve of *DES*. (E) KM survival curve of *IL3RA*. (F) KM survival curve of SDS. KM: Kaplan-Meier.

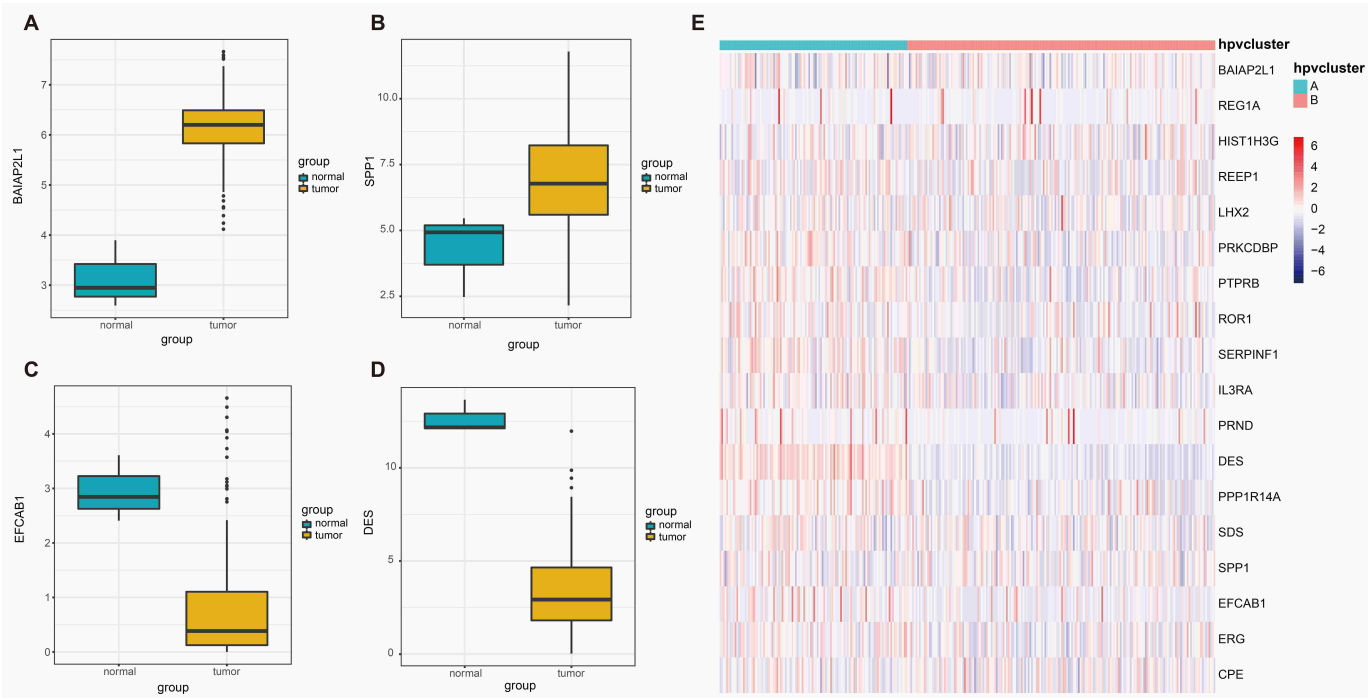


FIGURE 4. CESC subtype construction. (A–D) box plots of differential expression of *BAIAP2L1* (A), *SPP1* (B), *DES* (C) and *EFCAB1* (D) in the tumor group versus the control group. (E) heat map according to cervical cancer subtype grouping. Red in the heat map indicates activation and blue indicates inhibition. *BAIAP2L1*: Brain-Specific Angiogenesis Inhibitor 1-Associated Protein 2-Like Protein 1; *SPP1*: secreted phosphoprotein 1; *DES*: Desmin; *EFCAB1*: EF-Hand Calcium Binding Domain 1.

TABLE 3. GO enrichment analysis of differentially expressed genes.

Ontology	Description	*p*.adjust
BP	extracellular matrix organization	8.48×10^{-50}
BP	extracellular structure organization	8.48×10^{-50}
BP	collagen fibril organization	1.32×10^{-22}
BP	collagen metabolic process	3.06×10^{-12}
BP	Ossification	1.96×10^{-11}
BP	transmembrane receptor protein serine/threonine kinase signaling pathway	8.27×10^{-9}
BP	collagen catabolic process	2.00×10^{-8}
BP	connective tissue development	3.30×10^{-8}
BP	cell-substrate adhesion	4.70×10^{-8}
BP	cellular response to transforming growth factor beta stimulus	1.50×10^{-7}
CC	collagen-containing extracellular matrix	7.05×10^{-59}
CC	collagen trimer	2.99×10^{-16}
CC	endoplasmic reticulum lumen	5.97×10^{-15}
CC	basement membrane	7.01×10^{-14}
CC	fibrillar collagen trimer	3.21×10^{-11}
CC	banded collagen fibril	3.21×10^{-11}
CC	complex of collagen trimers	4.72×10^{-11}
CC	lysosomal lumen	1.15×10^{-6}
CC	Golgi lumen	1.90×10^{-6}
CC	Microfibril	1.41×10^{-5}
MF	extracellular matrix structural constituent	3.04×10^{-52}
MF	collagen binding	6.01×10^{-21}
MF	glycosaminoglycan binding	6.74×10^{-15}
MF	extracellular matrix structural constituent conferring tensile strength	1.15×10^{-14}
MF	heparin binding	1.21×10^{-11}
MF	growth factor binding	1.09×10^{-10}
MF	extracellular matrix structural constituent conferring compression resistance	1.14×10^{-10}
MF	integrin binding	1.80×10^{-10}
MF	extracellular matrix binding	4.80×10^{-10}
MF	sulfur compound binding	4.80×10^{-10}

BP: Biological Process; CC: Cell Component; MF: Molecular Function.

TABLE 4. KEGG enrichment analysis of differentially expressed gene.

ID	Description	*p*.adjust
hsa04974	Protein digestion and absorption	0.000000002
hsa04510	Focal adhesion	0.000000002
hsa04512	ECM-receptor interaction	0.000000003
hsa04151	PI3K-Akt signaling pathway	0.000019600
hsa05165	Human papillomavirus infection	0.000045100
hsa04350	TGF-beta signaling pathway	0.000186447
hsa04270	Vascular smooth muscle contraction	0.000186447
hsa05205	Proteoglycans in cancer	0.000530173
hsa04926	Relaxin signaling pathway	0.000979279
hsa04933	AGE-RAGE signaling pathway in diabetic complications	0.001702110

KEGG: Kyoto Encyclopedia of Genes and Genomes; ECM: extra cellular matrix; TGF-beta: transforming growth factor-beta; AGE-RAGE: advanced glycosylation end products-receptor of AGEs.

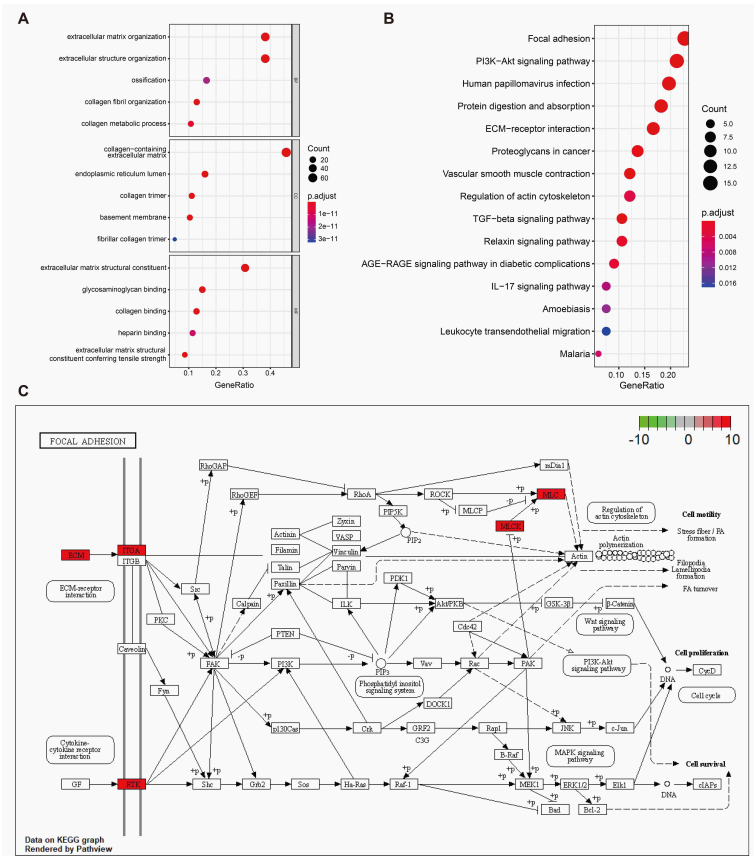


FIGURE 5. GO and KEGG enrichment analysis. (A) GO functional enrichment analysis, horizontal coordinate is gene ratio, vertical coordinate is GO terms, node size indicates the number of genes enriched in the pathway, node color indicates −log10 (*p* value). (B) KEGG pathway enrichment analysis, horizontal coordinate is gene ratio, vertical coordinate is Pathway name, node size indicates the number of *genes* enriched in the pathway, node color indicates −log10 (*p* value). (C) The most significantly enriched pathway plot hsa04510: Focal adhesion. GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.

3.8 Clinical relevance of the prognostic signature

We further evaluated the extent to which the prognostic signature discriminated the identification of CESC and plotted the ROC curve (Fig. 10A–C). The results showed that risk and protective factors could better distinguish between tumors and non-tumors. Next, we analyzed and plotted subgroup survival curves, selecting clinical stage III or IV for subgroup analysis (Fig. 10D–F) and found that the protective factor *IL3RA* remained a protective factor for cancer in the subgroup analysis, and patients with high *IL3RA* expression had a significantly longer overall survival (OS) time than those with low expression. In contrast, high expression of the risk factor *SPP1* was associated with an obviously shorter OS time in patients with CESC.

3.9 Validation of the prognostic signature

To assess better the signature's reliability, we validated it using GSE7803 and GSE63514 datasets. We calibrated the two datasets separately and presented the calibrated effects in box plots (Fig. 11A,B). Differential analysis was performed on the datasets, setting genes as upregulated with logFC >1 and adj*p*-value < 0.05 and downregulated with logFC <−1 and adj*p*-value < 0.05.

Among them, 753 DEGs were obtained from the GSE7803 dataset and plotted in volcanoes for display (Fig. 11C). We verified the expression of *SPP1* as a risk factor and *IL3RA* as a protective factor in GSE7803 (Fig. 11D,E). The results showed high expression of *SPP1* and low expression of *IL3RA* in the tumor group compared to normal controls.

We then analyzed the GSE63514 dataset similarly and obtained 793 DEGs with the same screening criteria, and the results were presented in a volcano plot (Fig. 11F). Similarly, we analyzed the expression of *SPP1* and *IL3RA* in the GSE63514 dataset (Fig. 11G,H), which also resulted in high expression of *SPP1* as a risk factor in the tumor group and low expression of *IL3RA* as a protective factor in the tumor group compared to the normal control.

3.10 Prognostic correlation analysis of the prognostic hazard score

Next, clinical factors such as OS time, age, and clinical stage were included for hazard score analysis, and the hazard score was calculated using LASSO + Cox regression analyses. The Kaplan-Meier survival curve was plotted according to the hazard score (Fig. 12A), which revealed that this hazard score could well distinguish the prognosis of CESC patients. We also plotted a nomogram (Fig. 12B) for this hazard score and found that it could make a good judgment of the prognosis of

TABLE 5. GSEA analysis.

Description	enrichment Score	*p*.adjust
KEGG_ECM_receptor_interaction	−0.821278412	0.000000009
KEGG_focal_adhesion	−0.690243170	0.000000009
KEGG_dilated_cardiomyophathy	−0.723481435	0.000000009
KEGG_hypertrophic_cardiomyopathy	−0.720182166	0.000000037
KEGG_arrhythmogenic_right_ventricular_cardiomyopathy	−0.734157688	0.000000082
KEGG_graft_versus_host_disease	0.709494059	0.000001350
KEGG_pathways_in_cancer	−0.548865956	0.000002060
KEGG_vascular_smooth_muscle_contraction	−0.631938464	0.000031200
KEGG_TGF_β_signaling_pathway	−0.653315180	0.000078200
KEGG_WNT_signaling_pathway	−0.583594965	0.000092400
HALLMARK_apical_junction	−0.641040630	0.000000001
HALLMARK_coagulation	−0.706726018	0.000000001
HALLMARK_epithelial_mesenchymal_transition	−0.840581779	0.000000001
HALLMARK_myogenesis	−0.685495760	0.000000001
HALLMARK_UV_response_down	−0.650962589	0.000000012
HALLMARK_KRAS_signaling_up	−0.611778568	0.000000083
HALLMARK_angiogenesis	−0.773264148	0.000009180
HALLMARK_apoptosis	−0.540973675	0.001742462
HALLMARK_Notch_signaling	−0.690197821	0.004296571
HALLMARK_TGF_β_signaling	−0.632152990	0.005483642

ECM: ExtraCellular Matrix; TGF: Transforming Growth Factor; WNT: Wnt signaling pathway; UV: Ultravioletradiation; KRAS: Kirsten ratsarcoma viral oncogene homolog.

TABLE 6. GSVA analysis.

Description	logFC	adj.*p*.Val
KEGG_vascular_smooth_muscle_conteraction	−0.434373082	0.007970771
KEGG_base_excision_repair	0.561972639	0.016442738
KEGG_DNA_replication	0.738938018	0.025738759
KEGG_cell_cycle	0.482209259	0.030952893
KEGG_calcium_signaling_pathway	−0.285262661	0.047859151
HALLMARK_mtorc1_signaling	0.462952875	0.002365503
HALLMARK_E2F_targets	0.687053829	0.006645147
HALLMARK_myogenesis	−0.407280453	0.006906586
HALLMARK_G2M_checkpoint	0.601508744	0.015351567
HALLMARK_glycolysis	0.299040739	0.024698092

patients, and its efficacy far exceeded that of clinical staging. Calibration plots (Fig. 12C–E) showed that the predictions at 1, 3 and 5 years were very close to the true values.

3.11 Immunohistochemical results of prognosis-related genes

Using the HPA database, we further analyzed the immunohistochemical results of *SPP1* and *IL3RA* in normal and CESC tissues (Fig. 13). *SPP1* was more highly expressed in CESC tissues than in normal tissues, while *IL3RA* was expressed to a higher extent in both CESC and normal tissues, which may have undergone molecular regulatory function during the development of the tumor, and the related molecular regulatory mechanism needs to be further clarified by subsequent studies.

4. Discussion

Based on the expression profile and survival status of CESC patients, we established a predictive model and validated the correlation between the model and CESC prognosis. Based on this model, 22 prognosis-related potential biomarkers of CESC were identified, and most genes had high prognostic efficacy. Among these potential biomarkers, those with a greater impact on CESC prognosis were selected to plot survival curves. We found that as prognosis-related risk factors for CESC patients, *BAIAP2L1* and *SPP1* were highly expressed in tumor tissues compared to paraneoplastic tissues. The prognostic relevance of *BAIAP2L1* in CESC has not yet been reported. *BAIAP2L1* is a member of the inverse BAR (Bin/amphiphysin/Rvs) domain proteins and can mediate cellular morphology and motility

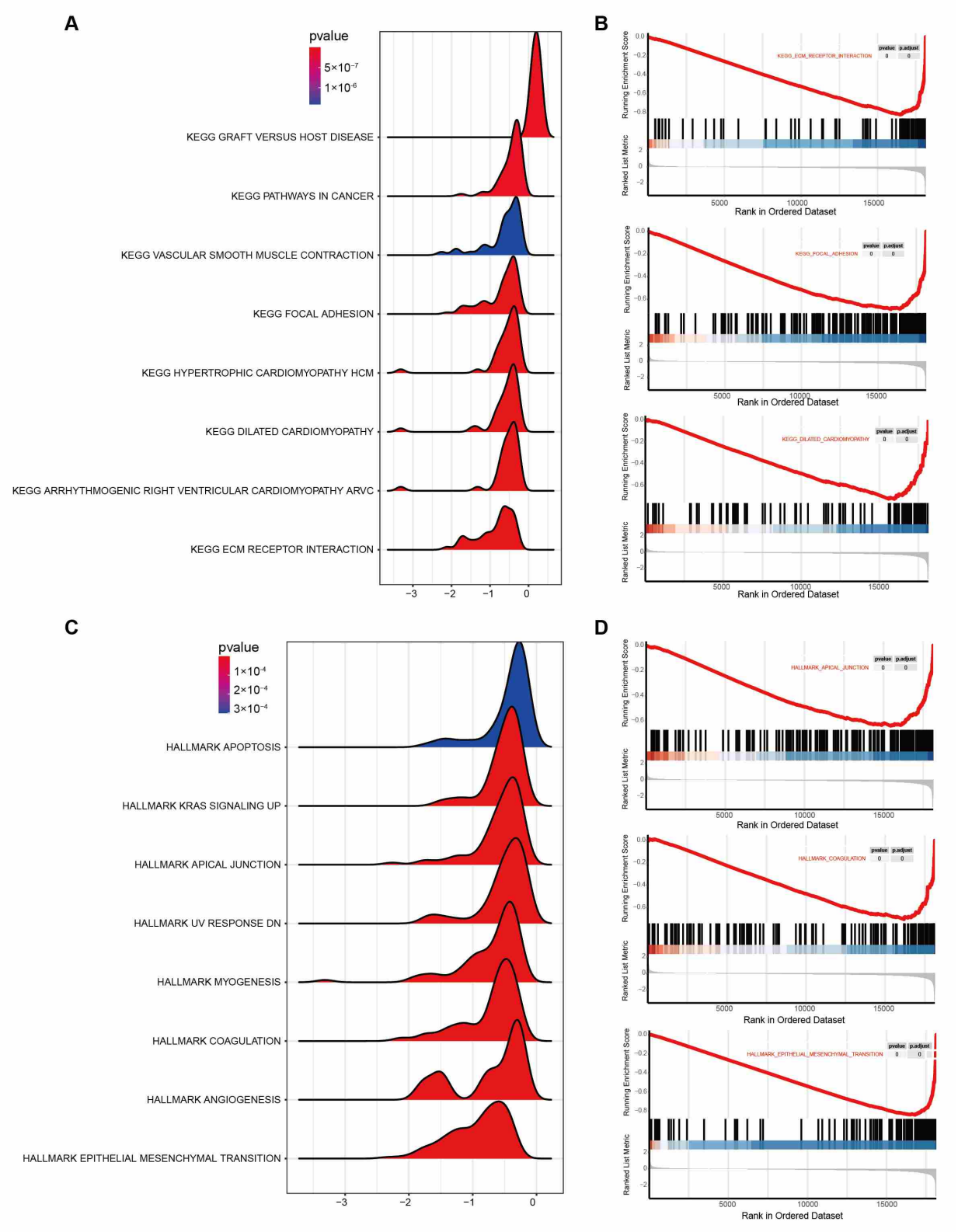


FIGURE 6. GSEA analysis. (A) GSEA analysis performed on cervical cancer subtypes, results visualized as a mountain range plot, horizontal coordinates are gene ratio, vertical coordinates are pathways, colors indicate *p*-value. (B–E) GSEA-KEGG typical results of ECM receptor interaction (B), focal adhesion (C), dilated cardiomyopathy (D) show. (E) GSEA analysis of cervical cancer subtypes, results visualized as a mountain range plot, horizontal coordinate is gene ratio, vertical coordinate is HALLMARK, color indicates *p*-value. (F–H) GSEA-HALLMARK typical results apical junction (F), coagulation (G), and epithelial mesenchymal transition (H) show. GSEA: Gene Set Enrichment Analysis; KEGG: Kyoto Encyclopedia of Genes and Genomes; ECM: ExtraCellular Matrix.

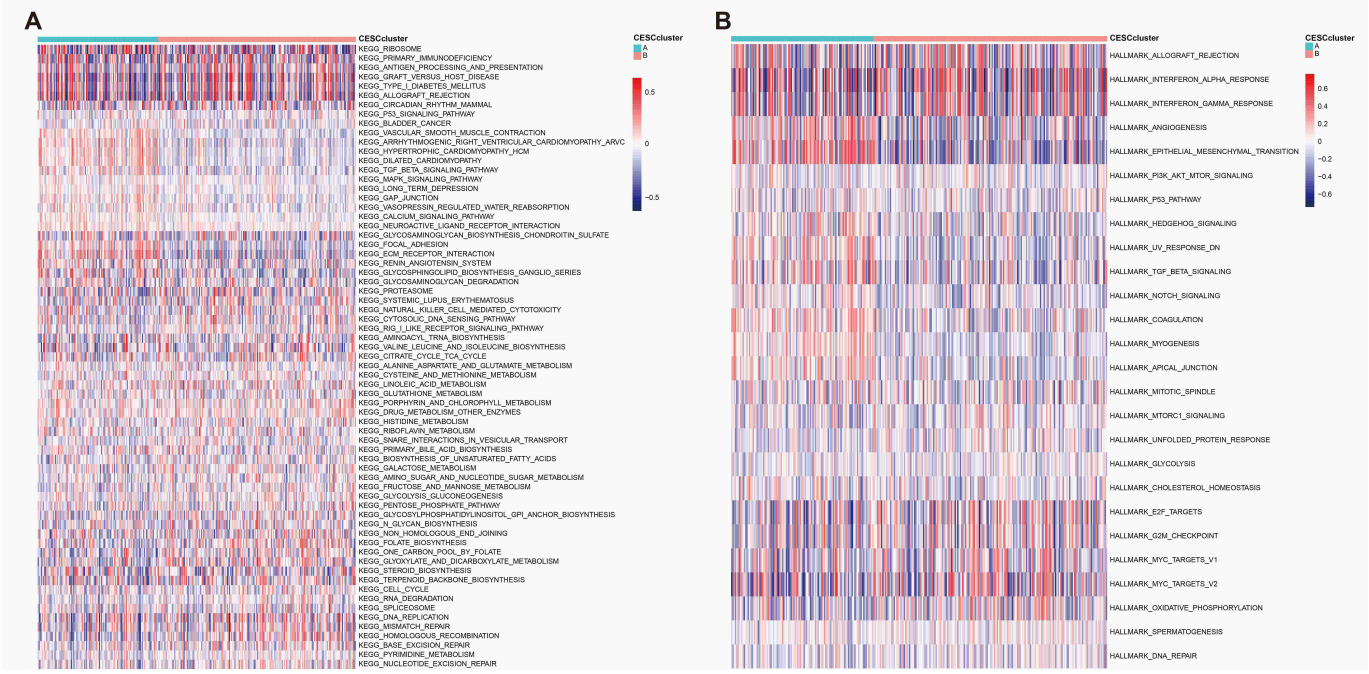
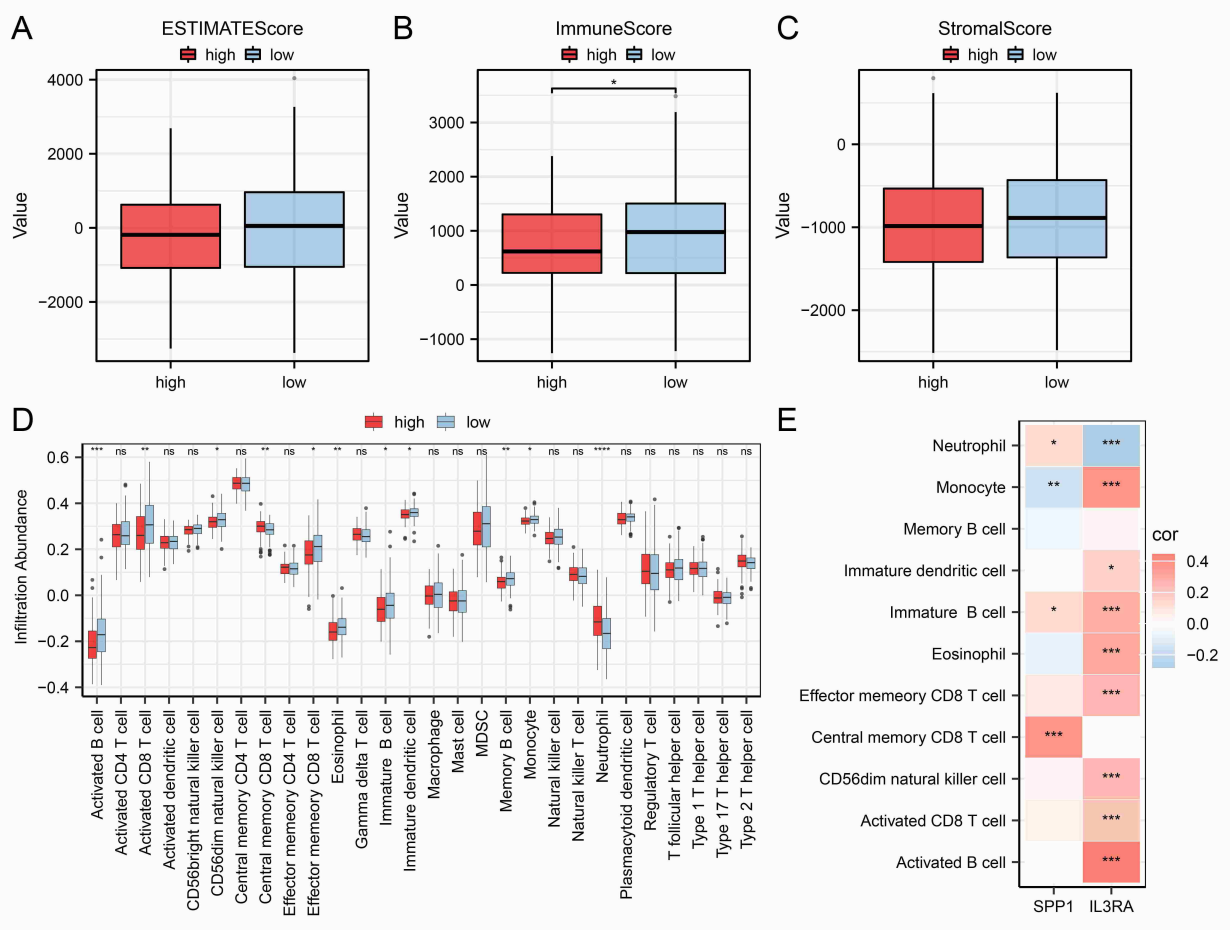


FIGURE 7. GSVA analysis. (A) GSVA analysis and heat map visualization of GSEA-KEGG enrichment results. (B) GSVA analysis and heat map visualization of GSEA-HALLMARK enrichment results. Red indicates activation, blue indicates inhibition. GSVA: Gene Set Variation Analysis; GSEA: Gene Set Enrichment Analysis; KEGG: Kyoto Encyclopedia of Genes and Genomes.

by regulating actin polymerization. Several recent studies have shown that high expression levels are associated with advanced stages and metastasis of cancer. It can be used as a biomarker for human ovarian cancer (Kao C, Hsu A, *et al.* [26], 2015), correlated with progression and OS time of gastric cancer (Huang LY, Wang X, *et al.* [27], 2018), and can serve as a biomarker in diagnosis and prognosis for lung cancer patients (Xu L, Du H, *et al.* [28], 2019, Liu XQ, Tufman A, *et al.* [29], 2020). The association of this gene with CESC needs to be further validated in future studies. *SPP1*, also known as osteopontin or early T-lymphocyte activation 1 protein, is present in a wide range of tissues throughout the body under physiological and disease conditions, and its biological functions are diverse and specific. It has been repeatedly reported in the development of several types of cancers and is considered to be associated with poor prognosis (Lamort AS, Giopanou I, *et al.* [30], 2019). Recent studies have shown a close relationship between *SPP1* and CESC; it has been reported that there is a significantly high expression of *SPP1* mRNA and protein in CESC, which may lead to cancer cell resistance to cisplatin (Chen X, Xiong D, *et al.* [31], 2019, Xu X, Jiang X, *et al.* [32], 2019). It has also been suggested to be a promising biomarker for predicting OS and immune checkpoint inhibitor treatment for CESC (Qu X, Shi Z, *et al.* [33], 2021, Li X, Zhang Q, *et al.* [34], 2021, Deepti P, Pasha A, *et al.* [35], 2022, Zhao K, Ma Z, *et al.* [36], 2022).

A total of 2381 DEGs were identified by analyzing the prognostic subtypes of CESC. GO, KEGG and GSEA enrichment analyses showed that the biological functions of the DEGs were mainly associated with alterations in the extracellular microenvironment and cell proliferation, motility, and survival. The most significant enrichment pathway is focal adhesion, a predominantly integrin-mediated anchoring junction located on the basal surface of epithelial cells that serves primarily to integrate the surrounding ECM and the actin cytoskeleton. This indicates a primary role in the development of tumors, such as in regulating cell survival and mediating tumor cell proliferation and migration (Mao D, Xu R, *et al.* [37], 2021). Several studies have shown the important role of this pathway in promoting the migratory, invasive, and adhesive capacities of CESC cells (Xu F, Zhang J, *et al.* [38], 2017, Du Q, Wang W, *et al.* [39], 2020, Chen H, Suo K, *et al.* [40], 2013). Inhibiting the function of key enzymes in this pathway can improve the metastasis and invasion of CESC cells (Hao Z, Yang J, *et al.* [41], 2015, Yamamoto N, Kinoshita T, *et al.* [42], 2013).

According to the infiltration assessment, 22 types of immune cells in CESC tissues play an important role in the immune microenvironment of CESC, especially killer immune cells such as various types of macrophages and CD8+ T cells. The TME components provide an immunosuppressive environment, thus promoting tumor development. Among these components, the role of tumor-associated macrophages has received increasing attention (Wang Q, Steger A, *et al.* [43], 2019). The phenotype of macrophages constantly changes during the different stages of CESC, which affects the ability of cancerous tissues to proliferate, invade, and metastasize (Liu Y, Li L, *et al.* [44], 2020), and the M2-like subtype is considered to be functional in cervical tumor progression, whereas the M1-like subtype has the opposite effect (Wang Q, Steger A, *et al.* [43], 2019). Regulation of macrophage polarization has been reported to control tumor progression through the TME [41]. Macrophages are heterogeneous tissue-resistant cells that exhibit relevant functions in the microenvironment. During CESC development, the quantity and type of tumor-infiltrating T cells may also influence the type of macrophage. Patients



F I G U R E 8. Immunoscore and immune infiltration analysis of the TCGA-CESC dataset. (A–C) TCGA-CESC dataset low group and high group, ESTIMATE Score (A), Immune Score (B), Stromal Score (C). (D) Comparison of grouping of immune cells in TCGA-CESC dataset low group and high group. (E) Heat map of the correlation between immune cells and EDGs. Symbol * equals $p < 0.05$, statistically significant; symbol ** equals $p < 0.01$, highly statistically significant; symbol *** equals $p < 0.001$, highly statistically significant. CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma; TCGA: The Cancer Genome Atlas.

with a high CD8+ T cell/Treg ratio or CD8+ T cell/Foxp3+ tumor-infiltrating lymphocytes ratio within the epithelium of CESC have been shown to have further improved survival as an independent prognostic factor, and they could enhance and activate M1-like macrophages which could be an option for immunotherapy (de Vos van Steenwijk PJ, *et al.* [45], 2013, Liang Y, Lü W, *et al.* [46], 2018).

In the survival curve analysis of stage III and IV CESC subgroups, we found that patients with high expression of *IL3RA* had longer OS times than those with low expression levels. In contrast, patients with high SPP1 expression had a significantly shorter OS time. The *IL3RA*, CD123, is coded by chromosomes Xp22.3 and Yp11.3 (Milatovich A, Kitamura T, *et al.* [47], 1993). IL3RA is mainly found in hematologic malignancies and is considered a potential therapeutic target for leukemia (Shi M, Su RJ, *et al.* [48], 2019). To the best of our knowledge, very few studies have reported the role of *IL3RA* in CESC. In a high-HPV-associated abnormality, vulvar intraepithelial neoplasia, a study showed that CD123+ plasmacytoid DCs were distributed throughout the thickness of the epidermis and dermis, and the numbers were significantly increased compared with healthy controls. Another study showed that the CD123+ DCs ratio in the peripheral blood of CESC patients was lower than that of cervical intraepithelial neoplasia and healthy controls. The decrease in circulating DCs may lead to an altered immune function involving cytokines, antigen presentation, and puerile T cells, which could cause disease progression (van Seters M, Beckmann I, *et al.* [49], 2008).

A shortcoming of the current study was that the results were not validated using an independent cohort of patients. Therefore, further *in vitro* or *in vivo* experiments are needed to validate our findings.

5. Conclusions

In the current study, using a prognostic model for CESC based on the TCGA database, we aimed to understand the roles of potential biomarkers in the prognostic development of CESC, explore the possible molecular mechanisms involved in the development of CESC, and identify possible loci for targeted, individualized therapy. The gene *BAIAP2L1* was first identified as a risk factor for the prognosis of CESC, while IL3RA was identified as a protective factor. Understanding

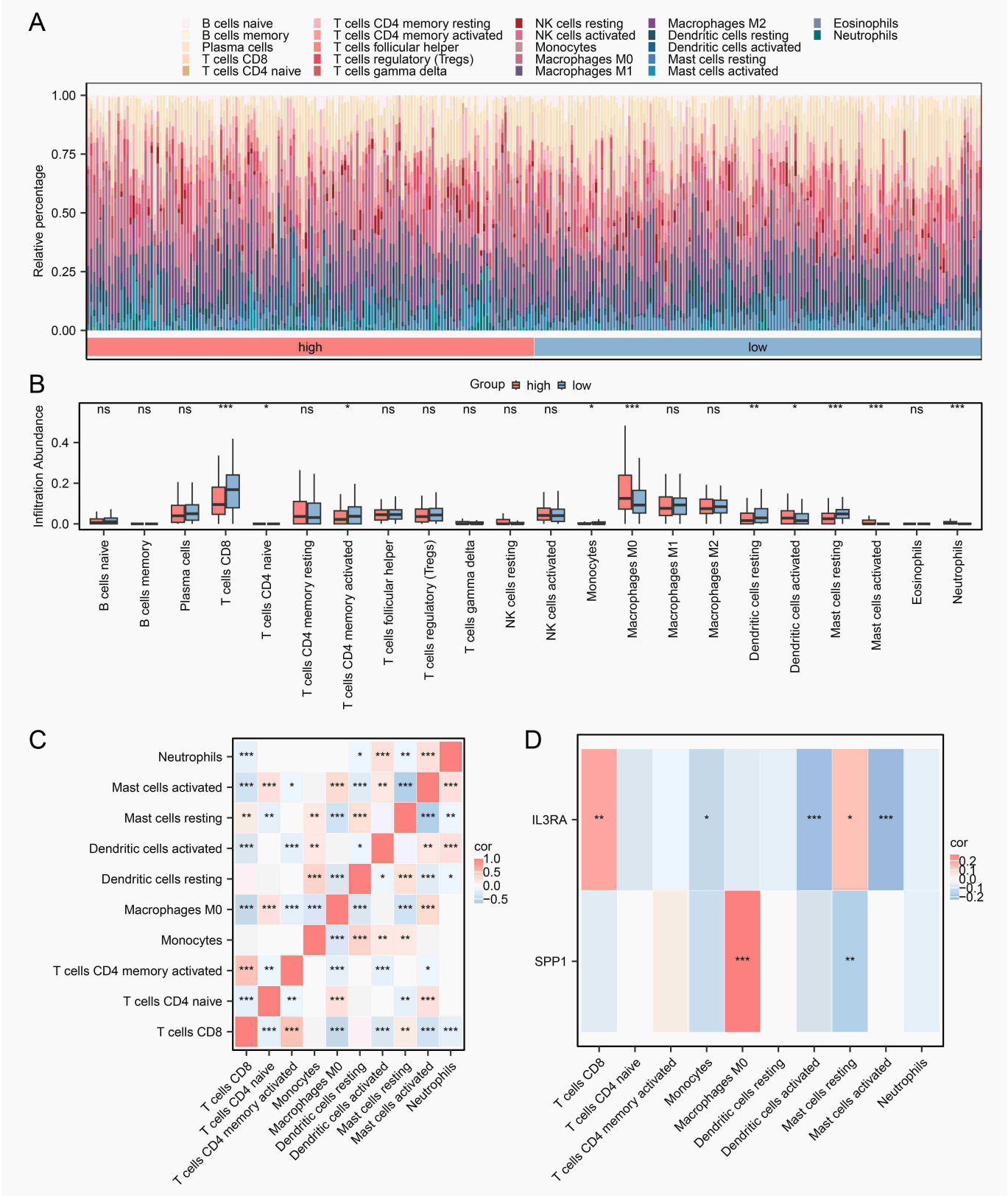


FIGURE 9. Immuno-infiltration analysis (CIBERSORT). (A) Stacked histogram of immune infiltration in TCGA-CESC. (B) Comparative graph of grouping of immune cells in different subgroups (high, low) of TCGA-CESC. (C) Heat map of correlation of immune cells. (D) Lollipop plot of correlation of DEGs with immune cells. Symbol ns equals $p \geq 0.05$, which is not statistically significant; symbol * equals $p < 0.05$, which is statistically significant; symbol ** equals $p < 0.01$, which is highly statistically significant; symbol *** equals $p < 0.001$, which is highly statistically significant. CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma; TCGA: The Cancer Genome Atlas.

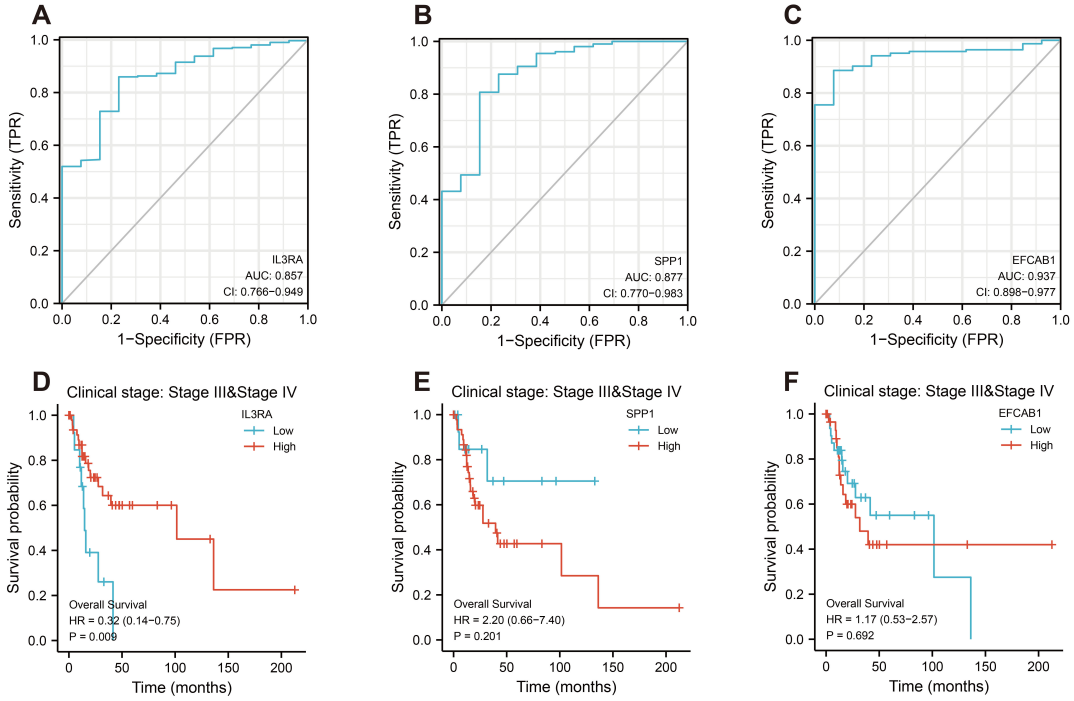


FIGURE 10. Clinical relevance of the prognostic model. (A–C) ROC curves for *IL3RA* (A), *SPP1* (B) and *EFCAB1* (C). (D–F) Overall survival curves for subgroups KM of *IL3RA* (D), *SPP1* (E) and *EFCAB1* (F). ROC: receiver operating characteristic; KM: Kaplan-Meier.

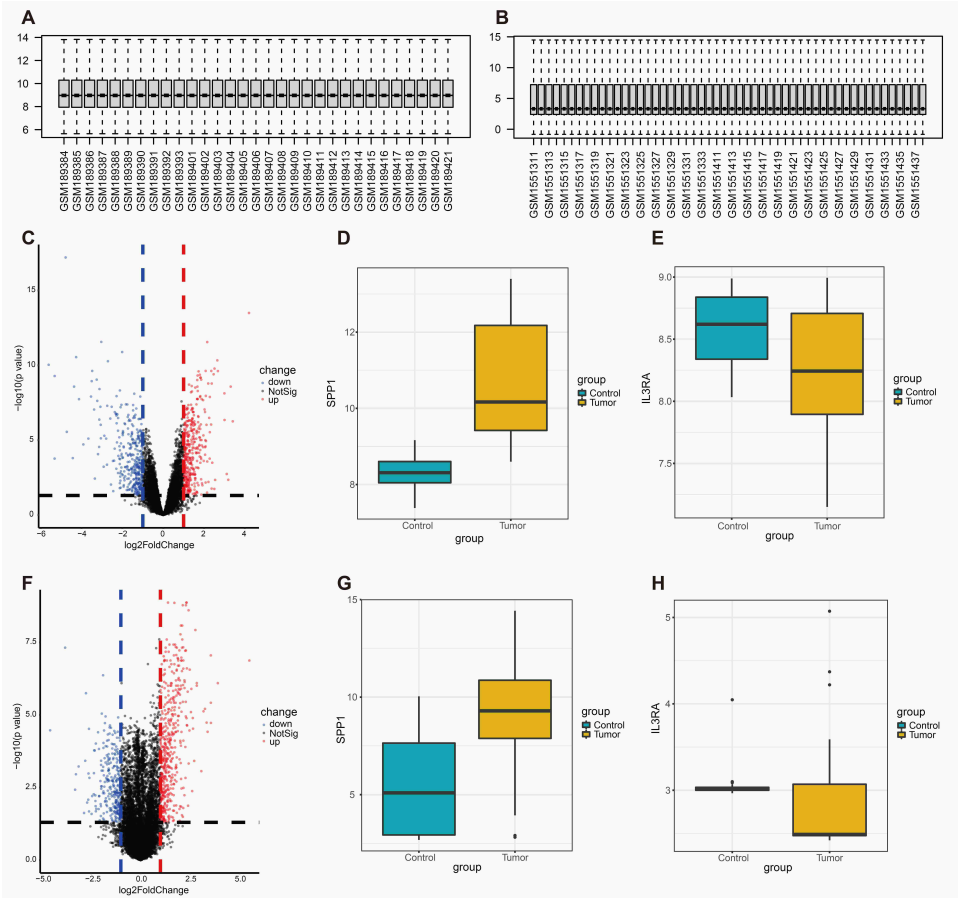


FIGURE 11. Validation of the prognostic model. (A,B) box plots of GSE7803 (A) and GSE63514 (B) datasets after correction. (C) box plots of GSE7803 differential gene volcanoes. (D,E) box plots of *IL3RA* (D) and *SPP1* (E) expression differences in the GSE7803 dataset. (F) box plots of GSE63514 differential gene volcanoes. (G,H) box plots of *IL3RA* (G) and *SPP1* (H) box plots of differential expression in the GSE63514 dataset. GSE: Gene Expression Omnibus Series.

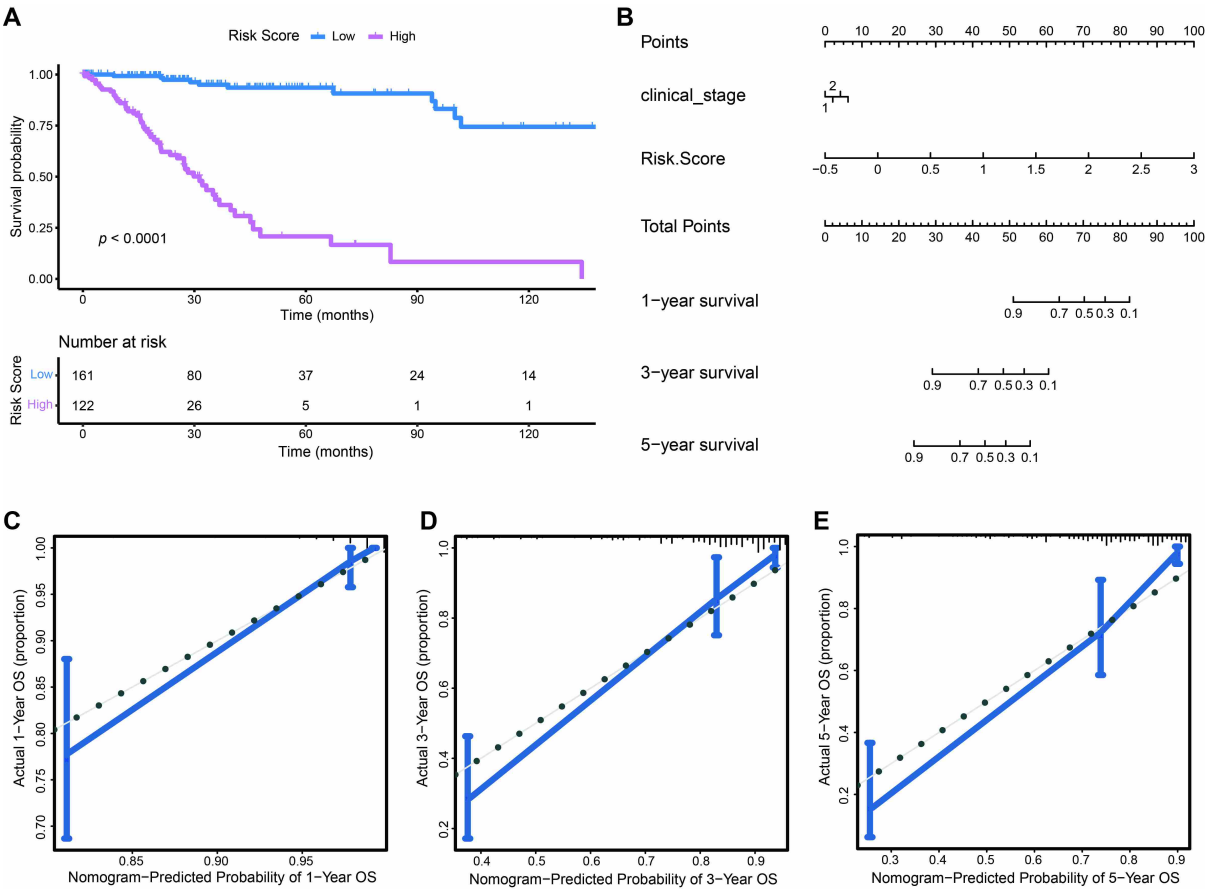


FIGURE 12. Prognostic correlation analysis of prognostic hazard scores. (A) Overall survival curves for hazard scores. (B) Columnar plots of hazard score correlations. (C–E) calibration plots for 1 year (C), 3 years (D), and 5 years (E). OS: Overall Survival.

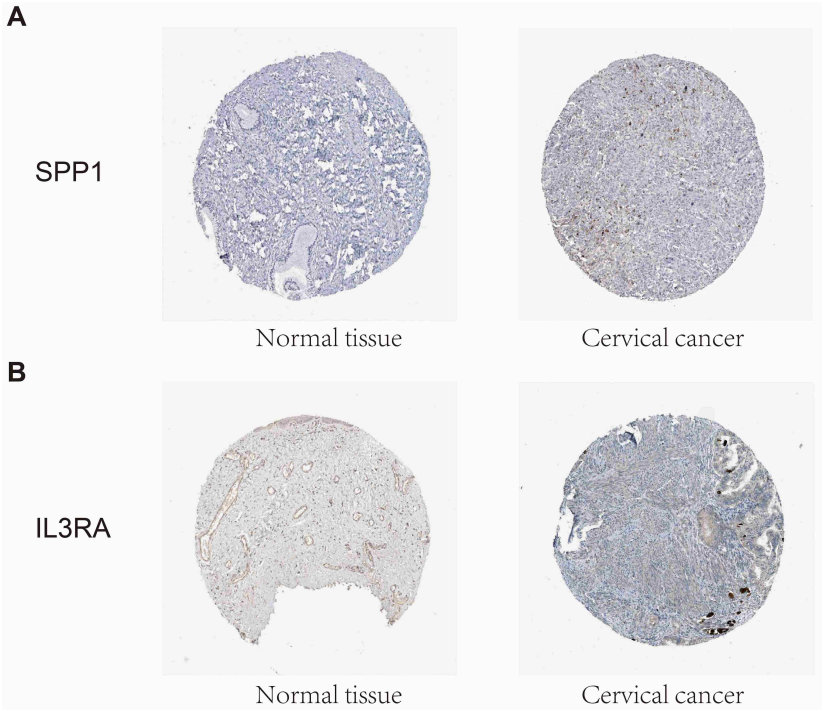


FIGURE 13. Immunohistochemical results of prognosis-related genes. (A) Immunohistochemical results of *SPP1* in normal cervical tissues and cervical cancer tissues. (B) Immunohistochemical results of *IL3RA* in normal cervical tissues and cervical cancer tissues. *SPP1*: secreted phosphoprotein 1; *IL3RA*: OSinterleukin-3 receptor alpha chain.

the potential biomarkers and their biological functions in the TME of CESC will help develop new ideas about the possible mechanisms of CESC development and provide new therapeutic options. The shortcomings of this study are that it was not further validated in combination with wet experiments and lacked corresponding clinical correlation studies. More specific pathogenesis and molecular targets require further validation.

AVAILABILITY OF DATA AND MATERIALS

The authors confirm that the data supporting the findings of this study are available within the article.

AUTHOR CONTRIBUTIONS

CLF and YZ—contributed to the conception of the study, the data analysis and wrote the manuscript. FFH, HLC and HJX—helped the data analysis and manuscript writing. JD and BQX—performed the analysis with constructive discussions. All authors approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This study is supported by the Science and Technology Development Fund of Nanjing Medical University (NMUB2018038).

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

[1] 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: A Cancer Journal for Clinicians. 2018; 68: 394–424.

[2] Buskwofie A, David-West G, Clare CA. A review of cervical cancer: incidence and disparities. Journal of the National Medical Association. 2020; 112: 229–232.

[3] Kagabu M, Nagasawa T, Sato C, Fukagawa Y, Kawamura H, Tomabechi H, *et al*. Immunotherapy for uterine cervical cancer using checkpoint inhibitors: future directions. International Journal of Molecular Sciences. 2020; 21: 2335.

[4] Marquina G, Manzano A, Casado A. Targeted agents in cervical cancer: beyond bevacizumab. Current Oncology Reports. 2018; 20: 40.

[5] Peng L, Hayatullah G, Zhou H, Chang S, Liu L, Qiu H, *et al*. Tumor microenvironment characterization in cervical cancer identifies prognostic relevant gene signatures. PLoS One. 2021; 16: e0249374.

[6] Hinshaw DC, Shevde LA. The tumor microenvironment innately modulates cancer progression. Cancer Research. 2019; 79: 4557–4566.

[7] Xu F, Shen J, Xu S. Multi-omics data analyses construct a six immune-related genes prognostic model for cervical cancer in tumor microenvironment. Frontiers in Genetics. 2021; 12: 663617.

[8] Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, *et al*. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Research. 2016; 44: e71.

[9] Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. Bioinformatics. 2007; 23: 1846–1847.

[10] Zhai Y, Kuick R, Nan B, Ota I, Weiss SJ, Trimble CL, *et al*. Gene expression analysis of preinvasive and invasive cervical squamous cell carcinomas identifies HOXC10 as a key mediator of invasion. Cancer Research. 2007; 67: 10163–72.

[11] Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, *et al*. NCBI GEO: mining tens of millions of expression profiles–database and tools update. Nucleic Acids Research. 2007; 35: D760–D765.

[12] den Boon JA, Pyeon D, Wang SS, Horswill M, Schiffman M, Sherman M, *et al*. Molecular transitions from papillomavirus infection to cervical precancer and cancer: role of stromal estrogen receptor signaling. Proceedings of the National Academy of Sciences. 2015; 112: E3255–64.

[13] 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.

[14] Gene Ontology Consortium. Gene ontology consortium: going forward. Nucleic Acids Research. 2015; 43: D1049–56.

[15] Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research. 1999; 27: 29–34.

[16] Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research. 2000; 28: 27–30.

[17] Wixon J, Kell D. The Kyoto encyclopedia of genes and genomes—KEGG. Yeast. 2000; 17: 48–55.

[18] 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.

[19] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005; 102: 15545–15550.

[20] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular signatures database (MSigDB) hallmark gene set collection. Cell Systems. 2015; 1: 417–425.

[21] Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, *et al*. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature Communications. 2013; 4: 2612.

[22] Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, *et al*. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009; 462: 108–12.

[23] Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, *et al*. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nature Biotechnology. 2019; 37: 773–782.

[24] Sjöstedt E, Zhong W, Fagerberg L, Karlsson M, Mitsios N, Adori C, *et al*. An atlas of the protein-coding genes in the human, pig, and mouse brain. Science. 2020; 367: eaay5947.

[25] Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010; 26: 1572–1573.

[26] Chao A, Tsai CL, Jung SM, Chuang WC, Kao C, Hsu A, *et al*. BAI1-associated protein 2-like 1 (BAIAP2L1) is a potential biomarker in ovarian cancer. PLoS One. 2015; 10: e0133081.

[27] Huang L, Wang X, Cui X, Li H, Zhao J, Wu C, *et al*. IRTKS is correlated with progression and survival time of patients with gastric cancer. Gut. 2018; 67: 1400–1409.

[28] Xu L, Du H, Zhang Q, Wang C, Yan L, Tian G, *et al*. BAI1 associated protein 2 like 2 is a potential biomarker in lung cancer. Oncology Reports. 2019; 41: 1304–1312.

[29] Liu XQ, Tufman A, Kiefl R, Li GF, Ma QL, Huber RM. Identification of lung adenocarcinoma-specific exosome RNAs in peripheral blood by RNA-Seq analysis. European Review for Medical and Pharmacological Sciences. 2020; 24: 1877–1886.

[30] Lamort AS, Giopanou I, Psallidas I, Stathopoulos GT. Osteopontin as a link between inflammation and cancer: the thorax in the spotlight. Cells. 2019; 8: 815.

[31] Chen X, Xiong D, Ye L, Yang H, Mei S, Wu J, *et al*. SPP1 inhibition improves the cisplatin chemo-sensitivity of cervical cancer cell lines. Cancer Chemotherapy and Pharmacology. 2019; 83: 603–613.

[32] Xu X, Jiang X, Chen L, Zhao Y, Huang Z, Zhou H, *et al*. MiR-181a promotes apoptosis and reduces cisplatin resistance by inhibiting osteopontin in cervical cancer cells. Cancer Biotherapy & Radiopharmaceuticals. 2019; 34: 559–565.

[33] Qu X, Shi Z, Guo J, Guo C, Qiu J, Hua K. Identification of a novel six-gene signature with potential prognostic and therapeutic value in cervical cancer. Cancer Medicine. 2021; 10: 6881–6896.

[34] Li X, Zhang Q, Chen G, Luo D. Multi-omics analysis showed the clinical value of gene signatures of C1QC^{+} and SPP1^{+} TAMs in cervical cancer. Frontiers in Immunology. 2021; 12: 694801.

[35] Deepti P, Pasha A, Kumbhakar DV, Doneti R, Heena SK, Bhanoth S, *et al*. Overexpression of secreted phosphoprotein 1 (SPP1) predicts poor survival in HPV positive cervical cancer. Gene. 2022; 824: 146381.

[36] Zhao K, Ma Z, Zhang W. Comprehensive analysis to identify *SPP1* as a prognostic biomarker in cervical cancer. Front Genet. 2022; 12: 732822.

[37] Mao D, Xu R, Chen H, Chen X, Li D, Song S, *et al*. Cross-talk of focal adhesion-related gene defines prognosis and the immune microenvironment in gastric cancer. Frontiers in Cell and Developmental Biology. 2021; 9: 716461.

[38] Xu F, Zhang J, Hu G, Liu L, Liang W. Hypoxia and TGF-β1 induced PLOD2 expression improve the migration and invasion of cervical cancer cells by promoting epithelial-to-mesenchymal transition (EMT) and focal adhesion formation. Cancer Cell International. 2017; 17: 54.

[39] Du Q, Wang W, Liu T, Shang C, Huang J, Liao Y, *et al*. High expression of integrin α3 predicts poor prognosis and promotes tumor metastasis and angiogenesis by activating the c-Src/extracellular signal-regulated protein kinase/focal adhesion kinase signaling pathway in cervical cancer. Frontiers in Oncology. 2020; 10: 36.

[40] Chen H, Suo K, Cheng Y, Zheng B, Xu L. Vascular endothelial growth factor C enhances cervical cancer migration and invasion via activation of focal adhesion kinase. Gynecological Endocrinology. 2013; 29: 20–24.

[41] Hao Z, Yang J, Wang C, Li Y, Zhang Y, Dong X, *et al*. MicroRNA-7 inhibits metastasis and invasion through targeting focal adhesion kinase in cervical cancer. International Journal of Clinical and Experimental Medicine. 2015; 8: 480–7.

[42] YAMAMOTO N, KINOSHITA T, NOHATA N, ITESAKO T, YOSHINO H, ENOKIDA H, *et al*. Tumor suppressive microRNA-218 inhibits cancer cell migration and invasion by targeting focal adhesion pathways in cervical squamous cell carcinoma. International Journal of Oncology. 2013; 42: 1523–1532.

[43] Wang Q, Steger A, Mahner S, Jeschke U, Heidegger H. The formation and therapeutic update of tumor-associated macrophages in cervical cancer. International Journal of Molecular Sciences. 2019; 20: 3310.

[44] Liu Y, Li L, Li Y, Zhao X. Research progress on tumor-associated macrophages and inflammation in cervical cancer. BioMed Research International. 2020; 2020: 1–6.

[45] de Vos van Steenwijk PJ, Ramwadhdoebe TH, Goedemans R, Doorduijn EM, van Ham JJ, Gorter A, *et al*. Tumor-infiltrating CD14-positive myeloid cells and CD8-positive T-cells prolong survival in patients with cervical carcinoma. International Journal of Cancer. 2013; 133: 2884–2894.

[46] Liang Y, Lü W, Zhang X, Lü B. Tumor-infiltrating CD8+ and FOXP3+ lymphocytes before and after neoadjuvant chemotherapy in cervical cancer. Diagnostic Pathology. 2018; 13: 93.

[47] Milatovich A, Kitamura T, Miyajima A, Francke U. Gene for the alpha-subunit of the human interleukin-3 receptor (IL3RA) localized to the X-Y pseudoautosomal region. American Journal of Human Genetics. 1993; 53: 1146–53.

[48] Shi M, Su RJ, Parmar K, Chaudhry R, Sun K, Rao J, *et al*. CD123: a novel biomarker for diagnosis and treatment of leukemia. Cardiovascular & Hematological Disorders-Drug Targets. 2019; 19: 195–204.

[49] van Seters M, Beckmann I, Heijmans-Antonissen C, van Beurden M, Ewing PC, Zijlstra FJ, *et al*. Disturbed patterns of immunocompetent cells in usual-type vulvar intraepithelial neoplasia. Cancer Research. 2008; 68: 6617–22.

How to cite this article: Chunli Fang, Ya Zhu, Feifei Hu, Hailin Chen, Huajing Xiao, Jie Ding, *et al*. Exploration of prognostic genes in cervical cancer immune microenvironment. European Journal of Gynaecological Oncology. 2024; 45(5): 42-59. doi: 10.22514/ejgo.2024.093.