Identification of Two Novel Alternative Splicing-related Genes in the Cervical Cancer Immune Microenvironment

Dan Sun
Third Xiangya Hospital of Central South University
Zhifu Zhi
The First Affiliated Hospital of Guangxi Medical University
Aiqian Zhang
Third Xiangya Hospital of Central South University
Bingsi Gao
Third Xiangya Hospital of Central South University
Lingxiao Zou
Third Xiangya Hospital of Central South University
Huan Huang
Third Xiangya Hospital of Central South University
Xingping Zhao
Third Xiangya Hospital of Central South University
Dabao Xu (dabaoxu@yahoo.com)
Third Xiangya Hospital of Central South University

Research Article

Keywords: alternative splicing, immune microenvironment, cervical cancer, SHF, FOXRED2

Posted Date: November 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1030113/v1

Abstract

Background: Cervical cancer (CC) is one of the most common malignant tumors of the female reproductive system. The tumor immunotherapy showed the remarkable effect. Associated alternative splicing (AS)-event signatures provide potential therapeutic targets and improved strategies for new drug development in CC management.

Methods: Clinical information and messenger RNA (mRNA) expression profiles were downloaded from The Cancer Genome Atlas (TCGA) database. Hub genes were extracted from the 7 AS-related genes for correlation analysis with clinical parameters and tumor-immune microenvironment. The relationship between the risk score and the 6 most important checkpoint genes was investigated. Finally, we estimated the Stroma and Immune cells using the Expression data (ESTIMATE) algorithm.

Results: It was revealed that T cells CD8, T cells regulatory (Tregs), T cells CD4 memory activated, neutrophils, mast cells resting, mast cells activated, and macrophages M0 had a significant difference between the low- and high risk-score groups. The genes *SHF* and *FOXRED2* were extracted as hub genes. High expression of *SHF* ($P < 0.002$) and low expression of *FOXRED2* ($P < 0.001$) were associated with poor prognosis..Immune checkpoint genes *IDO1*, *PDCD1*, and *HAVCR2,* were negatively correlated with risk-score.

Conclusions: We used bioinformatics to assess the heterogeneity of tumor-infiltrating immune cells in CESC and discovered out 2 hub genes, *SHF* and *FOXRED2*, from the AS prognostic model. The immune checkpoint genes *IDO1*, *PDCD1*, and *HAVCR2,* showed negative correlations with risk-score. The outcomes were significant for studying tumor progression's immune-related mechanisms and exploring novel prognostic predictors and precise therapy methods.

Introduction

Cervical cancer (CC) is one of the most common malignant tumors of the female reproductive system. Despite advanced screening and prophylactic vaccines, more than 50% of the cases are still detected at an advanced stage (Liontos et al., 2019). The Clinical Practice Guidelines for Cervical Cancer in 2019 (1st Edition) published by the National Comprehensive Cancer Network (NCCN) (Koh et al., 2019) recommended that Pembrolizumab may be used in the second-line treatment of recurrent or metastatic CC with programmed death protein 1 ligand (PD-L1) positive or deficient mismatch repair, (DMMR)/microsatellite instability-high (MSI-H) (2A). The PD-1 inhibitor, Pembrolizumab, was approved to treat advanced and relapsed CC by the Food and Drug Authority (FDA). The immunocheckpoint inhibitors showed the effectiveness in the treatment of CC, but the overall effective rate was less than 30% (Naumann et al., 2019). Therefore, it is important to search for markers to predict the effectiveness of immunotherapy.

Messenger RNA (mRNA) processing is an important maturation process involving mRNA splicing, polyadenylation and capping. Alternative splicing (AS) is an important step in the process of maturation

in the transcription process and regulates eukaryotic gene expression. Defects in this procedure can generate cancer (Tazi et al., 2009). As events in different pre-mRNAs are changed during carcinogenesis, in some cases the relationship between certain splicing events and the development of some cancer features has been established, such as proliferation, angiogenesis, and increased invasion (Hagen et al., 2012; Oltean et al., 2014), leading to the consideration of AS as a new hallmark of cancer (Ladomery, 2013).

Aberrant AS of the *IL1RAP* gene promotes immune evasion and promotes CC (Liu et al., 2018). Ouyang et al. (Ouyang et al., 2020) constructed a clinical prognostic model using prognostic-AS events, and *SNRPA* and *CCDC12* were identified as hub genes for prognosis-associated splicing factors (SFs). Some researchers have recommended that cell SFs contribute to cervical cancer by 2 different but convergent mechanisms: (1) the differential maturation of human papillomavirus (HPV) contributes to the viral replication and viral oncoprotein expression; and (2) by promoting the production of cell mRNA variants and proteins in carcinogenesis functions that may play a role in the development of cervical cancer (Mole et al., 2009; Johansson et al., 2013; Oltean et al., 2014).

Considering the importance of immunotherapy in CC, the characterization of immune infiltration features is essential for further understanding. In this study, the prognostic-related genes which were involved in the AS prognostic model were analyzed using whole-genome analysis. Most importantly, we further analyzed the characteristics of 2 AS-associated genes in the tumor-immune microenvironment. The correlation between 6 most important immune checkpoint genes and the risk score was also investigated. Our research has provided insights into screening of CC related candidate biomarkers that may have a clinical significance in diagnosis, prognosis and treatment.

Methods

Data collection and processing

The mRNA expression profiles and corresponding clinical data of the cervical squamous cell carcinoma (CESC) cohort were downloaded from The Cancer Genome Atlas (TCGA) database (July 2021; https://portal.gdc.cancer.gov/). The AS event data for CESC were obtained online (https://bioinformatics.mdanderson.org/TCGASpliceSeq/) (Ryan et al., 2016). All methods were performed in accordance with the declaration of Helsinki (Declaration of Helsinki, 1968). Usefulness of clinical information was sufficiently evaluated. Several patients were excluded because of the lack of complete clinical features (e.g., age, grade, Federation International of Gynecology and Obstetrics [FIGO] stage, and survival data). A total of 246 patients with complete AS event data and clinical data were included in our analysis. The pathological types contained squamous cell neoplasms, adenomas and adenocarcinomas, cystic neoplasms, mucinous and serous neoplasms, and complex epithelial neoplasms.

The percent spliced in (PSI) value was used to quantify each AS event, which is the ratio of normalized reads indicating the presence of a transcript element versus the total normalized reads for that event. with a rating from 0 to 1. The PSI was calculated as: splice in/splice in+splice out (Wang et,. 2008). The AS data for PSI value > 0.75 was screened. The gene expression and clinical profiles were merged using Perl (v5.30.0; https://www.perl.org/get.html).

Screening for prognostic AS events in CESC

The TCGA SpliceSeq is a database based on TCGA RNA-seq data (Ryan et al., 2016). There are 7 types of selective splicing events including Alternate Acceptor site (AA), Alternate Donor site (AD), Alternate Promoter (AP), Alternate Terminator (AT), Exon Skip (ES), Mutually Exclusive Exons (ME), and Retained Intron (RI). The UpSet.R package (https://cran.r-project.org/web/packages/UpSetR/index.html) was used to analyzed the distributions of all encoded genes in each of the 7 types of AS events and survival-related AS events in the CESC cohort.

Construction of prognostic models and survival analysis

A univariate Cox regression analysis with R package "survival" was used to further evaluate the prognostic value of AS events in CC patients and to determine survival-related different expressed alternative splicing (DEAS) events, including overall survival (OS)-related DEAS events. Next, We then identified the final elimination of potential predictors with non-zero coefficients using the least absolute shrinkage and selection operator (LASSO) regression with the R package "glmnet,". Predictive models were also constructed by multivariate Cox regression analysis depending on the outcomes of LASSO Cox regression. The risk scores of each sample were calculated and the corresponding coefficients were obtained based on PSI values and multivariate Cox analysis, respectively. Risk score was calculated by the formula:

$$\text{risk-score}=\sum_{i=0}^{n} \text{PSI}\times\beta_i$$

where β is the regression coefficient of the AS events. A total of 246 CC patients were divided into high- and low-risk groups bounded by the median of risk-score. Kaplan–Meier survival analysis was then performed to determine the prognosis in each group. Furthermore, receiver operating characteristic (ROC) curves of 1, 3, and 5 years were obtained, and the identification of predictive signatures was shown using the survival ROC package in R (Heagerty et al., 2000) .

Establishment and validation of a predictive nomogram

Risk-score, age, FIGO stage, and grade, were incorporated to construct a nomogram in order to evaluate the probability of 1-, 3-, and 5-year OS of CC. Nomogram was evaluated by the calibration plot using the "rms" package. The nomogram predicted probabilities for the actual rates were evaluated by the calibration curve.

Immunescore estimate, immune cell infiltrating proportion inference

Normalized RNA expression data were used to infer the immunescore by the estimate package (Yoshihara et al., 2013) and quantify the infiltrating rate of 22 types of immune cells by the "CIBERSORT" package (https://cibersort.stanford.edu/) (Newman et al., 2015). The infiltrating rate of 22 types of immune cells was 100%. The immunity stage was quantified and classified by the Single-sample gene set enrichment analysis (ssGSEA) based on immune-related gene (IRG) sets (He et al., 2018). Next, 47 immune checkpoint genes were analyzed, and 13 of them, which differed from tumor and normal samples, were screened. We then analyzed the differences of 13 hub immune checkpoints among the high- and low-risk groups. The correlations between the 6 most important immune checkpoint genes (*CD274*, *PDCD1*, *PDCD1LG2*, *CTLA4*, *HAVCR2*, *IDO1*) and risk-score were further conducted.

Extraction of AS-related DEGs in CC samples

Seven AS-related DEGs involved in the model construction were analyzed (Table 1). The expression of these genes had differences between CC and normal samples using the R package "limma" with the threshold of $|log2FC|>1$ and $P < 0.05$.

Integrate AS-related DEGs with clinical features and prognosis

The gene expression was used to sepetate the gub genes into high- and low-expression groups. Then they were used to analyze clinical indicators, such as age, grade, and FIGO stage. Finally, the prognosis of the gub genes in the 2 groups was analyzed using the "survival" and "survminer" packages.

Analysis of the relationship between stromal/immune scores and AS-related DEGs of CC immune microenvironment

The Expression data (ESTIMATE) algorithm was applied to analyze the stromal score, immune score, ESTIMATE score, and tumor purity by the estimation of Stromal and Immune cells in Malignant Tumors to testify the effect of ssGSEA grouping based on transcriptome profiles of cervical cancer. In addition,

the stromal score, immune score, ESTIMATE score, and tumor purity in the high and low expression groups of gub genes were compared by Limma.R and ggpubr.R packages. Next, we analyzed the correlation between hub genes and immune cells.

Construction of potential SF-AS regulatory network

The data of 404 SFs were downloaded from the SpliceAid2 database (http://www.introni.it/splicing.html). The correlation between the expression level of SFs and PSI values of OS-related AS events was analyzed by R packages (BiocManager, limma). The absolute value of correlation coefficient > 0.6 and P < 0.001 were considered to be significant. Finally, the potential SF-AS regulatory network was visualized by Cytoscape software (v3.7.2; https://cytoscape.org/).

Statistical analysis

All statistical analyses were applied by R version 4.1.0 (R packages: survival, survminer, UpSetR, glmnet, estimate, ggpubr, e1071, rms, preprocessCore, vioplot, ggExtra, GSVA, GSEABase, reshape2, pheatmap, corrplot, ggplot, ggplot2, BiocManager; https://cran.r-project.org/bin/windows/base/). A P value < 0.05 was regarded as statistical significance.

Results

Overview of AS events in TCGA CESC cohort

A total of 246 patients with CESC were identified. The mRNA splicing data enrolled in this study includes 17,069 AS events in 9,608 genes. As shown in Fig. 1A, a single gene could have up to 6 different splicing modes, and most genes had more than 1 AS event. Exon skipping (ES) was the most frequent splice type in the 7 AS types, followed by an alternate terminator (AT) and alternate promoter (AP). The top 20 significantly survival-related AS events for 7 AS types were presented in Fig. 1D-K.

Prognostic index models featured by AS events for CC

To explore the prognostic utility of each AS signature, AS events associated with OS were identified by fitting univariate Cox proportional hazard regression models after merging the clinical data in the CC cohort. As a result, 2,269 AS events were determined with p < 0.05, containing 1,153 high-risk survival-associated AS events (hazard ratio [HR] >1) and 1,116 low-risk survival-associated AS events (HR < 1). The intersecting sets between different genes and AS events were visualized by the UpSet plot (Fig. 1B), indicating that 1 gene might have more than 1 survival-associated AS event. The most frequent splice type in the survival-associated AS events was AP, followed by ES and AT.

After conducting univariate regression analysis, LASSO regression was performed. The prediction models were constructed by the optimal survival-related AS events to avoid the overfitting of OS based models (Fig. 1C). In the meantime, a 7-AS event signature was identified as a survival predictor of CC through the Cox proportional hazards regression model (Table 1). A distribution diagram of survival risk-score, survival status of CC patients, and clustering heatmap of the PSI levels of survival-associated AS events are shown from top to bottom in Fig. 1L. The horizontal axis represents the patients in risk-score order from low to high (Fig. 1L).

Kaplan-Meier curves and log-rank tests were gererated to analyze the relationship between risk-score and survival status. The survival rates of low-risk patients were higher than that of high-risk patients, exactly as illustrated in Fig. 2F ($P < 0.001$). The ROC analysis was then applied to compare the predictive power of prognostic models, and the ROC curve (area under the curve [AUC]) in 1, 2, and 3 years were all greater than 0.850. Moreover, the AUC of the risk score model predicting a 1-year survival rate was larger than that of the age, grade, and FIGO stage (Fig. 2A, 2B).

Table 1

The 7 AS events associated with the OS of patients with CC

ID	Coefficient	HR	HR.95L	HR.95H	P-value
FCF1|28425|AD	3.8372	46.3961	1.7893	1203.0625	0.0209
FOXRED2|62052|AP	2.4601	11.7062	0.8097	169.2506	0.0711
HNRNPA1|301521|ES	3.4369	31.0901	0.2586	3738.2195	0.1596
NDUFA3|51782|ES	5.1283	168.7360	12.9262	2202.6492	0.0001
SHF|30409|AP	1.7134	5.5477	1.4991	20.5309	0.0103
NHLRC3|25701|ES	-4.3616	0.0128	0.0005	0.3207	0.0080
MCC|73005|AP	1.0190	2.7703	0.7634	10.0533	0.1213

HR, hazard ratio

Construction and evaluation of the nomogram

Univariate and multivariate Cox regression methods were used and combined patient clinical characteristics (age, grade, and FIGO stage) to analyze whether the 7-AS event signatures could be an independent predictor of CC patients survival As shown in Fig. 2, it is shown that the risk-score could still be used as a reliable and stable independent risk predictor in the CC cohort ($P < 0.001$; Fig. 2C). We constructed a predictive nomogram based on the multivariate analysis (Fig. 2D) that included risk scores and clinical characteristics. As a result, the risk-score showed satisfactory diagnostic ability together with clinical characteristics.

Risk-score and AS events are associated with the infiltration of immune cells in the CESC microenvironment

First, immunescore in 29 types of infiltrating immune cells and immune function was assessed by the ssGSEA method (He et al., 2018). The immune score difference of each immune cell in the low risk-score and high risk-score groups is shown in Fig. 3B and Fig. 3C. We further explored the impact of risk-score on the infiltration of 22 types of immune cells in the tumor microenvironment (TME) by the CIBERSORT algorithm. The landscape of 22 types of immune cells infiltrating in the low risk-score group and the high risk-score group is shown in Fig. 3A. Differential analysis results showed that 7 types of immune cells (T cells CD8, T cells regulatory [Tregs], T cells CD4 memory activated, neutrophils, mast cells resting, mast cells activated, and macrophages M0) had a significant difference between the 2 groups ($P < 0.05$). The correlations with risk-score and the number of those 7 types of immune cells were shown in Fig.3D-J. Positive correlation could be found between risk-score and neutrophils (R=0.24, P=0.001, Fig.3G), mast cells activated (R=0.33, P=4.3e-0.6, Fig.3I) as well as macrophages M0 (R=0.3, P=3.2e-0.5, Fig.3J).

Finally, the violin plot was used to assess the difference in tumor purity, ESTIMATE score, immune score, and stromal score calculated using the ESTIMATE algorithm between the 2 groups (Fig. 8A). Tumor purity, ESTIMATE score, stromal score, and the immune score had no differences in the low- and high risk-score groups ($P \geq 0.05$).

The risk-score is associated with the key immune checkpoint genes in the immune microenvironment of CESC

The difference in the expression level of 47 immune checkpoint genes in the low risk-score and high risk-score groups was evaluated, and 13 genes showed the significant differences (Fig. 4A). Next, R packages (limma, corrplot, ggpubr, ggExtra) were used to screen the risk-score associated with the 6 most important checkpoint genes (*CD274*, *PDCD1*, *PDCD1LG2*, *CTLA4*, *HAVCR2*, and *IDO1*). With the absolute threshold value of P less than 0.001, 3 immune checkpoint genes, *IDO1*, *PDCD1*, and *HAVCR2,* were identified (Fig. 4B). Scatter plots displaying the correlation of those 3 genes and risk-score were plotted separately. Though 3 of the correlation coefficients did not reach 0.3, the scatter plot showed a negative correlation (Fig. 4C-4E).

Extraction of IRGs depending on AS events and their correlation with clinical parameters

The expression of 7 genes (Table 1) was identified for difference analysis between CC and normal cohort by "limma" package with the threshold of |log2FC|≥1 and P≤0.05, and the 2 genes, *SHF* (logFC= -1.5396, P<0.05) and *FOXRED2* (logFC= 1.3212, P < 0.05) were extracted for further analysis. Next, we divided

tumor patients into high expression and low expression groups according to the optimal cut-off in *SHF* and *FOXRED2* for clinical prognostic analysis.

The correlations between 2 hub genes expression and clinicopathological parameters were evaluated using R packages (limma, survival, survminer). The expression of SHF was higher in normal group ($P < 0.01$, Fig. 5B) while the expression of FOXRED2 was lower in normal group ($P < 0.05$, Fig. 5G). However, no notable association between 2 genes and age, grade, or stage were found ($P \geq 0.05$) (Fig. 5C-E, H-J). Survival analysis revealed that high expression of *SHF* ($P < 0.002$) and low expression of *FOXRED2* ($P < 0.001$) were associated with poor prognosis (Fig. 5A, 5F).

Relationship between *SHF* expression and immune cell infiltration

The landscape of 22 types of immune cells infiltrating in the low expression group and high expression group of *SHF* is shown in Fig. 6A. The B cell naive, plasma cells, T cells CD8+, T cells CD4 memory activated, T cells CD4 memory resting, natural killer (NK) cells resting, macrophages M1, had differences in 2 groups. The immune score difference of each immune cell in the 2 groups is shown in Fig. 6B. The expression of *SHF* had positive correlations with T cells CD4 memory resting, plasma cells, mast cells resting, and B cells naive, while being negatively correlated with T cells CD4 memory activated, T cells CD8, NK cells resting, and macrophages M1 ($P < 0.05$) (Fig. 6D-6K). However, we found no strong correlation between immune cells infiltration and *SHF* expression. Given that the risk score was related to tumor immunity, we finally appraised the correlation between the gene signature and the expression of immune checkpoints. The immune checkpoints with differential expression in the low expression and high expression groups of SHF are shown in Fig. 6C.

Correlations between FOXRED2 expression and immune cell infiltration

Fig. 7A showed the landscape of 22 types of immune cells infiltrating in the low expression and high expression groups of *FOXRED2*. The number of B cell naive was different in 2 groups. The immune score difference of each immune cell in the 2 groups is shown in Fig. 7B. In additon, the association between *FOXRED2* expression and the number of tumor-infiltrating immune cells was investigated. The results exhibited that *FOXRED2* expression had positive correlations with T cells CD4 memory resting, mast cells resting, B cells naive, while being negatively correlated with NK cells resting (P□0.05) (Fig. 7D-7G). However, we also found no strong correlation between immune cells infiltration and *FOXRED2* expression. The immune checkpoints with differential expression in the low and high expression group of *FOXRED2* are shown in Fig. 7.

The tumor purity, ESTIMATE score, immune score, and stromal score between the low and high expression groups of *SHF* and *FOXRED2*

In order to verify the effectiveness of the grouping strategy between the low and high expression groups of *SHF* and *FOXRED2*, the ESTIMATE method was applied to evaluate tumor purity, ESTIMATE score, immune score, and stromal score. Compared with the high expression group in *SHF*, the low expression group had a higher immune score ($P < 0.001$) (Fig. 8B). Other parameters had no differences between 2 groups in *SHF* ($P \geq 0.05$) (Figure 8B). Compared with the low expression group in *FOXRED2*, the high expression group had a lower ESTIMATE score, immune score, and stromal score ($P < 0.001$), but had higher tumor purity ($P < 0.001$) (Fig. 8C).

Potential Regulatory Network Between SFs and AS Events

A total of 31 SFs (blue) were significantly associated with 227 survival-associated AS events consisting of 135 low-risk AS events (green) and 92 high-risk AS events (red). Most of low-risk AS events showed negatively correlation with the SF expression (green lines), and most of the high-risk AS events were positively correlated with SF expression (red lines) (Fig. 9).

Discussion

The process of AS enables the generation of distinct mRNA and protein isoforms from a single gene; splicing perturbations are common in cancer and are associated with mutations (Dong et al., 2019). 84% of RNA-binding proteins and more than 70% of splicing factors are dysregulated at the mRNA expression level in cancers (Dvinge et al., 2016; Sebestyén et al., 2016; Sveen et al., 2016). In addition to HPV infections, dysregulated pathways are a basic feature of CC development and progression. Therefore, the researches which elucidate the modifications of the cancer-related pathways and AS are rapidly increasing. Some studies showed abnormal AS and the dysregulation of gene expression in CC (Wu et al., 2018; Sharma et al., 2014; Campos-Parra et al., 2016; Lin et al., 2019). Related molecular signatures provide potential therapeutic targets and improved strategies for new drug development in CC management. This was the meaning and purpose of our research.

IRGs were extracted depending on AS events and their correlation with clinical parameters. Finally, 2 hub genes, *SHF*|30409|AP and *FOXRED2*|62052|AP, were extracted from the 7 genes involved in the AS prognostic model. Interestingly, the AS events in the 2 genes were both APs. Demircioğlu et al. inferred active promoters using RNA-seq data from 18,468 cancer and normal samples, demonstrating that APs are a major contributor to context-specific regulation of transcription (Demircioğlu et al., 2019). Promoters are deregulated in tissues, cancer types, and patients that affect known oncogenes and new candidates. Promoter activity provides a more accurate predictor of patient survival than gene expression because of genes with independently regulated promoters (Demircioğlu et al., 2019). Our research showed that the

expression of *SHF* was higher while the expression of *FOXRED2* was lower in the tumor group leading to a poor prognosis. However, the research of AP in *SHF* and *FOXRED2* has not been found. How the AS events influence the gene expression, and the tumor prognosis needs to be further analyzed.

Recently, antibody immunotherapy represented by immune checkpoint inhibitors has become a research hotspot. Immune checkpoints are regulatory molecules in the immune system, and it plays an important role in immune tolerance and tumor immune escape. The CTLA-4 inhibitor, Ipilimumab, was first approved by FDA for the treatment of malignant melanoma (Lipson et al., 2011). The FDA then approved pembrolizumab to treat advanced or recurrent CC in patients with disease progression during or after chemotherapy combined with tumor tissue positive for PD-L1 (Chung et al., 2019).

We found 3 immune checkpoint genes *IDO1*, *PDCD1*, and *HAVCR2*, that were negatively correlated with risk-score. Indoleamine 2, 3-dioxygenase 1 (*IDO1*) is a rate-limiting metabolic enzyme, and numerous studies have demonstrated that *IDO1* is highly expressed in multiple types of human cancer (Zhai et al., 2018). Research has shown that *IDO1* inhibition combined with PDCD1 blockade will soon become the "new backbone" for cervical immunotherapy. Because *IDO1* inhibitors act to block the destruction of important amino acid, tryptophan, in the TME. This helps tip the immune balance from regulatory T cells toward cytotoxic ones (2018). The *HAVCR2* gene, known as Tim-3 (T cell immunoglobulin and mucin-domain containing-3) is a co-inhibitory receptor that is expressed on interferon (IFN)-g-producing T cells, FoxP3+ Treg cells, and innate immune cells that are shown to suppress their responses to the interaction with their ligand/s (Monney et al., 2002; Meyers et al., 2005; McIntire et al., 2001). It has become prominent as a potential candidate for cancer immunotherapy, and it can enhance anti-tumor immunity and suppresse tumor growth in several preclinical tumor models (Das et al., 2017). How the AS event regulates immune checkpoint and thus affects the efficacy of immune checkpoint inhibitors remains to be further explored.

There are several limitations to current research. Though it was based on bioinformatics analysis, there were no recruitment cohorts for the prognosis verification.

Conclusion

Our research assessed the heterogeneity of tumor-infiltrating immune cells in CESC and discovered 2 hub genes, *SHF* and *FOXRED2*, from the 7 genes involved in the AS prognostic model. There were 3 immune checkpoint genes, *IDO1*, *PDCD1*, and *HAVCR2*, which showed negative correlations with risk-score. The outcomes were significant for studying tumor progression's immune-related mechanisms and exploring novel prognostic predictors and precise therapy methods.

Declarations

Author contributions

DB Xu conceived and designed the study with XP Zhao. D Sun drafted the manuscript and analyzed the data. ZF Zhi handled the pictures, AQ Zhang and Bingsi Gao handled the article format. LX Zou and H Huang reviewed the data. All authors contributed to the article and approved the submitted version.

Acknowledgments

Funding:

This study was supported by the Hunan Science and Technology Department (2020 SK4017), the National Key Research and Development Program of China (2018YFC1004800), and the Hunan Provincial Clinical Medical Technology Innovation Guiding Project (2020SK53605 and 2020SK53606).

Conflicts of Interest:

The authors have no conflicts of interest to declare.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

References

1. Liontos, M., Kyriazoglou, A., Dimitriadis, I., Dimopoulos, M. A. & Bamias, A. Systemic therapy in cervical cancer: 30 years in review. *Crit Rev Oncol Hematol. May*, **137**, 9–17 : 10.1016/j.critrevonc.2019.02.009 (2019).
2. Koh, W. J. *et al.* Cervical Cancer, Version 3.2019, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw. Jan*, **17** (1), 64–84 : 10.6004/jnccn.2019.0001 (2019).
3. Naumann, R. W. *et al.* (2019). Safety and Efficacy of Nivolumab Monotherapy in Recurrent or Metastatic Cervical, Vaginal, or Vulvar Carcinoma: Results From the Phase I/II CheckMate 358 Trial. J Clin Oncol. Nov 1;37(31):2825-2834. doi: 10.1200/JCO.19.00739
4. Tazi, J., Bakkour, N. & Stamm, S. Alternative splicing and disease. *Biochim Biophys Acta. Jan*, **1792** (1), 14–26 (2009).
5. Hagen, R. M. & Ladomery, M. R. Role of splice variants in the metastatic progression of prostate cancer. *Biochem Soc Trans. Aug*, **40** (4), 870–4 : 10.1042/BST20120026 (2012).
6. Oltean, S. & Bates, D. O. (2014). Hallmarks of alternative splicing in cancer. Oncogene. Nov 13;33(46):5311-8. doi: 10.1038/onc.2013.533

7. Ladomery, M. (2013). Aberrant alternative splicing is another hallmark of cancer. Int J Cell Biol. 2013:463786. doi: 10.1155/2013/463786
8. Liu, F. *et al.* SRSF10-mediated IL1RAP alternative splicing regulates cervical cancer oncogenesis via mIL1RAP-NF-κB-CD47 axis. *Oncogene. May*, **37** (18), 2394–2409 (2018).
9. Ouyang, D., Yang, P., Cai, J., Sun, S. & Wang, Z. Comprehensive analysis of prognostic alternative splicing signature in cervical cancer. *Cancer Cell Int. Jun*, **8**, 20:20221 (2020).
10. Mole, S. *et al.* RNA splicing factors regulated by HPV16 during cervical tumour progression. *J Pathol. Nov*, **219** (3), 383–91 : 10.1002/path.2608 (2009).
11. Johansson, C. & Schwartz, S. Regulation of human papillomavirus gene expression by splicing and polyadenylation. *Nat. Rev. Microbiol*, **11**, 239–251 (2013).
12. Oltean, S. & Bates, D. O. (2014). Hallmarks of alternative splicing in cancer. Oncogene. Nov 13;33(46):5311-8. doi: 10.1038/onc.2013.533
13. Ryan, M. *et al.* (2016). TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. Nucleic Acids Res. Jan 4;44(D1):D1018-22. http://doi: 10.1093/nar/gkv1288
14. (1968). Declaration of Helsinki.Phys Ther. Dec;48(12):1418–9. doi: 10.1093/ptj/48.12.1418
15. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes., **456**, 470–476 : 10.1038/nature07509 (2008).
16. Heagerty, P. J., Lumley, T. & Pepe, M. S. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics. Jun*, **56** (2), 337–44 (2000).
17. Yoshihara, K. *et al.* (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612. doi: 10.1038/ncomms3612
18. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*, **May;12** (5), 453–7 (2015).
19. He, Y., Jiang, Z., Chen, C. & Wang, X. Classification of triple-negative breast cancers based on Immunogenomic profiling. *J Exp Clin Cancer Res. Dec*, **29** ;37((1), 327 (2018).
20. Dong, M., Dong, Z., Zhu, X., Zhang, Y. & Song, L. Long non-coding RNA MIR205HG regulates KRT17 and tumor processes in cervical cancer via interaction with SRSF1. *Exp Mol Pathol. Dec*, **111**, 104322 : 10.1016/j.yexmp.2019.104322 (2019).
21. Dvinge, H., Kim, E., Abdel-Wahab, O. & Bradley, R. K. RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer. Jul*, **16** (7), 413–30 : 10.1038/nrc.2016.51 (2016).
22. Sebestyén, E. *et al.* Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res. Jun*, **26** (6), 732–44 (2016).
23. Sveen, A., Kilpinen, S., Ruusulehto, A., Lothe, R. A. & Skotheim, R. I. (2016). Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. Oncogene. May 12;35(19):2413-27. doi: 10.1038/onc.2015.318
24. Wu, K. *et al.* Identification of key pathways and genes in the progression of cervical cancer using bioinformatics analysis. *Oncol Lett*, **Jul;16** (1), 1003–1009 : 10.3892/ol.2018.8768 (2018).

25. Sharma, G., Dua, P. & Agarwal, S. M. A Comprehensive Review of Dysregulated miRNAs Involved in Cervical Cancer. *Curr Genomics. Aug*, **15** (4), 310–23 : 10.2174/1389202915666140528003249 (2014).
26. Campos-Parra, A. D. *et al.* Comprehensive transcriptome analysis identifies pathways with therapeutic potential in locally advanced cervical cancer. *Gynecol Oncol*, **Nov;143** (2), 406–413 : 10.1016/j.ygyno.2016.08.327 (2016).
27. Lin, M., Ye, M., Zhou, J., Wang, Z. P. & Zhu, X. Recent Advances on the Molecular Mechanism of Cervical Carcinogenesis Based on Systems Biology Technologies. *Comput Struct Biotechnol J. Feb*, **7**, 17:17241–250 (2019).
28. Demircioğlu, D. *et al.* A Pan-cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative Promoters. *Cell. Sep*, **5** ;178((6), 1465–147717 : 10.1016/j.cell.2019.08.018 (2019).
29. Lipson, E. J. & Drake, C. G. Ipilimumab: an anti-CTLA-4 antibody for metastatic melanoma. *Clin Cancer Res. Nov*, **15** ;17((22), 6958–62 (2011).
30. Chung, H. C. *et al.* Efficacy and Safety of Pembrolizumab in Previously Treated Advanced Cervical Cancer: Results From the Phase II KEYNOTE-158 Study. *J Clin Oncol. Jun*, **10** ;37((17), 1470–1478 : 10.1200/JCO.18.01265 (2019).
31. Zhai, L. *et al.* IDO1 in cancer: a Gemini of immune checkpoints. *Cell Mol Immunol. May*, **15** (5), 447–457 : 10.1038/cmi.2017.143 (2018).
32. (2018).Blocking IDO1 Helps Shrink Bladder, Cervical Tumors.Cancer Discov.Jan;8(1):OF3. doi: 10.1158/2159-8290.CD-NB2017-167
33. Monney, L. *et al.* Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease. *Nature.Jan*, **31** ;415((6871), 536–41 (2002).
34. Meyers, J. H., Sabatos, C. A., Chakravarti, S. & Kuchroo, V. K. The TIM gene family regulates autoimmune and allergic diseases. *Trends Mol Med. Aug*, **11** (8), 362–9 (2005).
35. McIntire, J. J. *et al.* Identification of Tapr (an airway hyperreactivity regulatory locus) and the linked Tim gene family. *Nat Immunol*, **Dec;2** (12), 1109–16 : 10.1038/ni739 (2001).
36. Das, M., Zhu, C. & Kuchroo, V. K. Tim-3 and its role in regulating anti-tumor immunity. *Immunol Rev. Mar*, **276** (1), 97–111 : 10.1111/imr.12520 (2017).

Figures

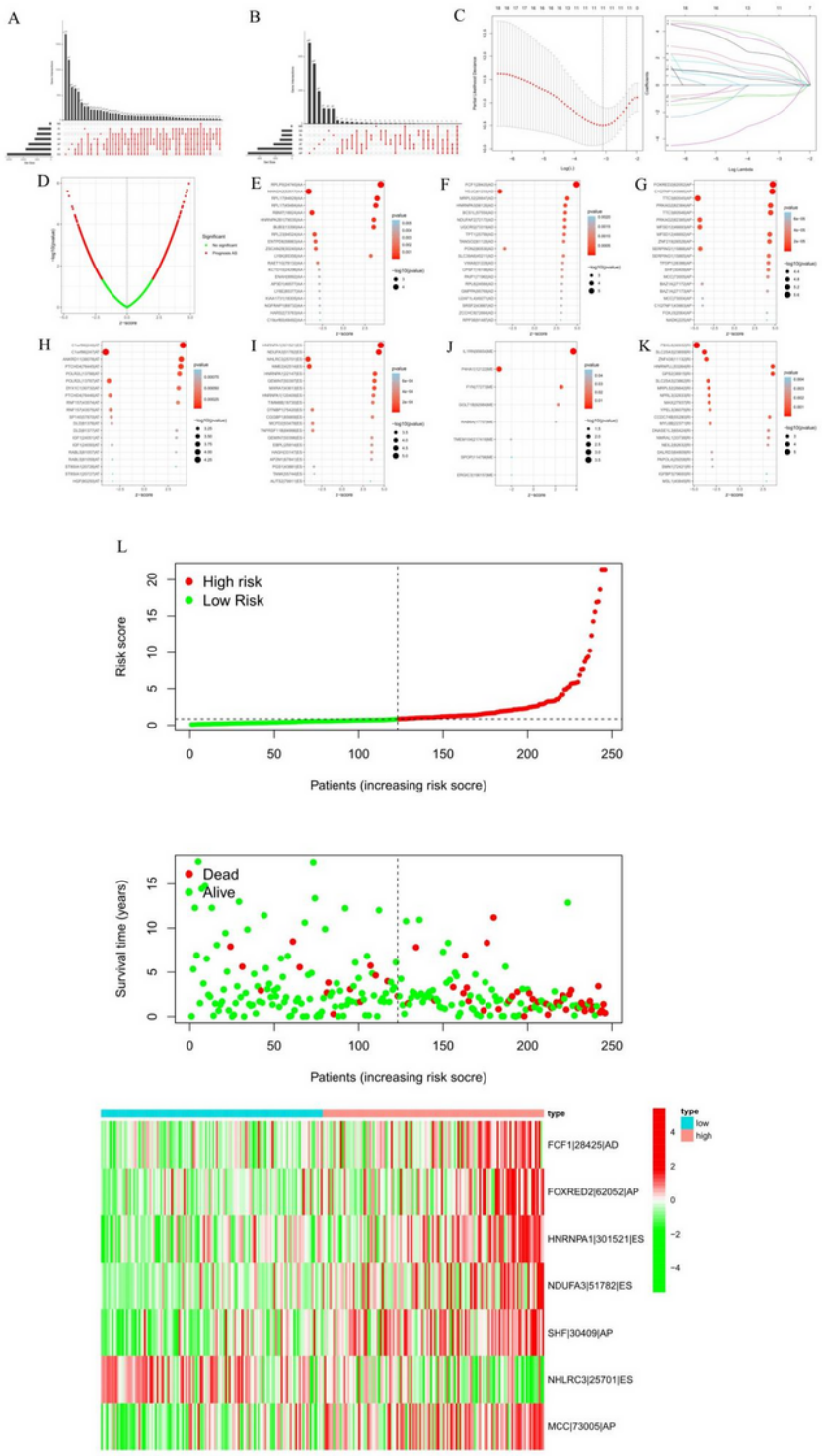

Figure 1

Identification and prognostic value of the AS markers in cervical cancer. The upSet plot of intersections and aggregates among diverse types of AS (A) and survival-associated AS events (B) in CESC. (C) LASSO coefficient profiles of survival-associated AS events and 10 x cross-validation for tuning parameter selection in the LASSO model. (D) The Volcano plot depicts the P-values from the univariate Cox analysis of 17,069 AS events. (E–K) Forest plots of z-score of the top 20 significantly survival-related AS events

for 7 splicing types (ME has only 8 events). (L) Risk score distribution of survival-associated AS events signature in the TCGA cohort including risk cores, survival status, and heatmap of the PSI profiles from top to bottom. AS, alternate splicing; CESC, the cervical squamous cell carcinoma; LASSO, least absolute shrinkage, and selection operator; AA, alternate acceptor; AD, alternate donor; AP, alternate promoter; AT, alternate terminator; ES, exon skip; ME, mutually exclusive exons; RI, retained intron; TCGA, The Cancer Genome Atlas; PSI, percent spliced in

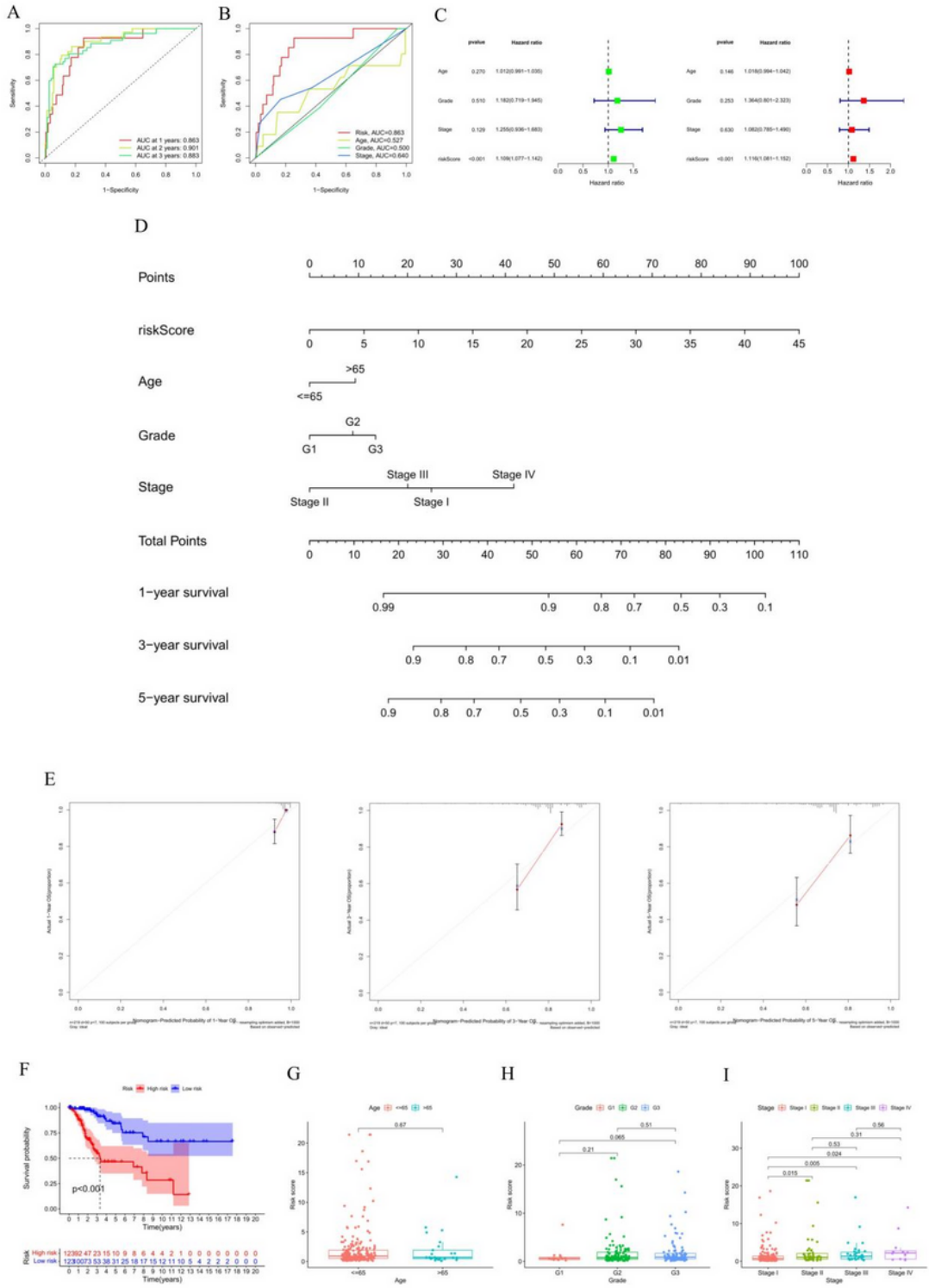

Figure 2

The risk-score based on AS events are associated with survival and clinical parameters of CC patients. (A, B) ROC curve in the predicted groups (high and low-risk groups) of the survival-associated AS events in the CESC cohort. (C) Univariate analysis and multivariate analysis of risk-scores and clinical characteristics that were simultaneously associated with OS. (D) The nomogram consisted of age, gender, FIGO stage, risk-score was used to predict the 1-, 3 - and 5-year survival probability of CC patients. (E) Calibration plots of the AS-clinic nomograms are in agreement between nomogram-predicted and observed 1, 3, and 5-year outcomes of the CESC cohort. The nomogram-predicted survival probability is plotted on the x-axis, and the actual survival is plotted on the y-axis. The 45º dashed line represents the ideal performance. The red lines represent the actual performances of the model, and the figures from left to right depict the 1-, 3-, and 5-year results. (F) Kaplan-Meier analysis for OS of CESC patients. Differences of risk-score in age (G), grade (H), and FIGO stage (I) groups. The bottom and top of the boxes show the 25th and 75th percentiles (interquartile range). AS, alternate splicing; CESC, the cervical squamous cell carcinoma; OS, overall survival; ROC, receiver operating characteristic; CC, cervical cancer; FIGO, Federation International of Gynecology and Obstetrics

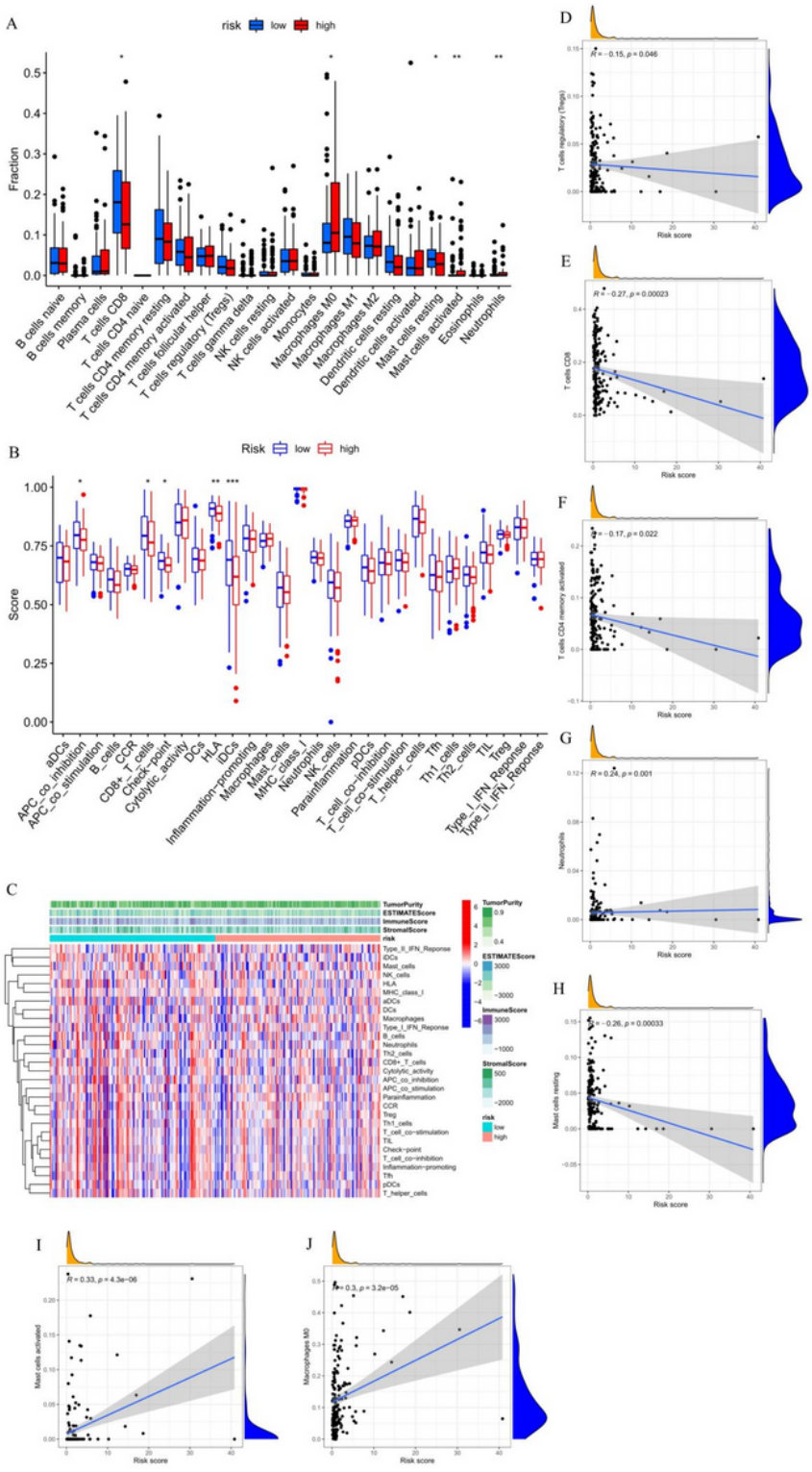

Figure 3

The relationship of the risk-score and immune cells infiltrating in CESC tumor-immune microenvironment. (A) The landscape of 22 types of infiltrating immune cells in the low- and high risk-score group. (B) The landscape of 29 types of infiltrating immune cells and immune function in 2 groups. (C) The heatmap showed a difference in infiltrating immune cells in the 2 groups in a CESC immune microenvironment. (D-J) The correlation with risk-score and the number of immune cells. The bottom and top of the boxes are

the 25th and 75th percentiles (interquartile range). Blue: low risk, red: high risk. ***p□0.001, **p□0.01, *p□0.05. CESC, the cervical squamous cell carcinoma

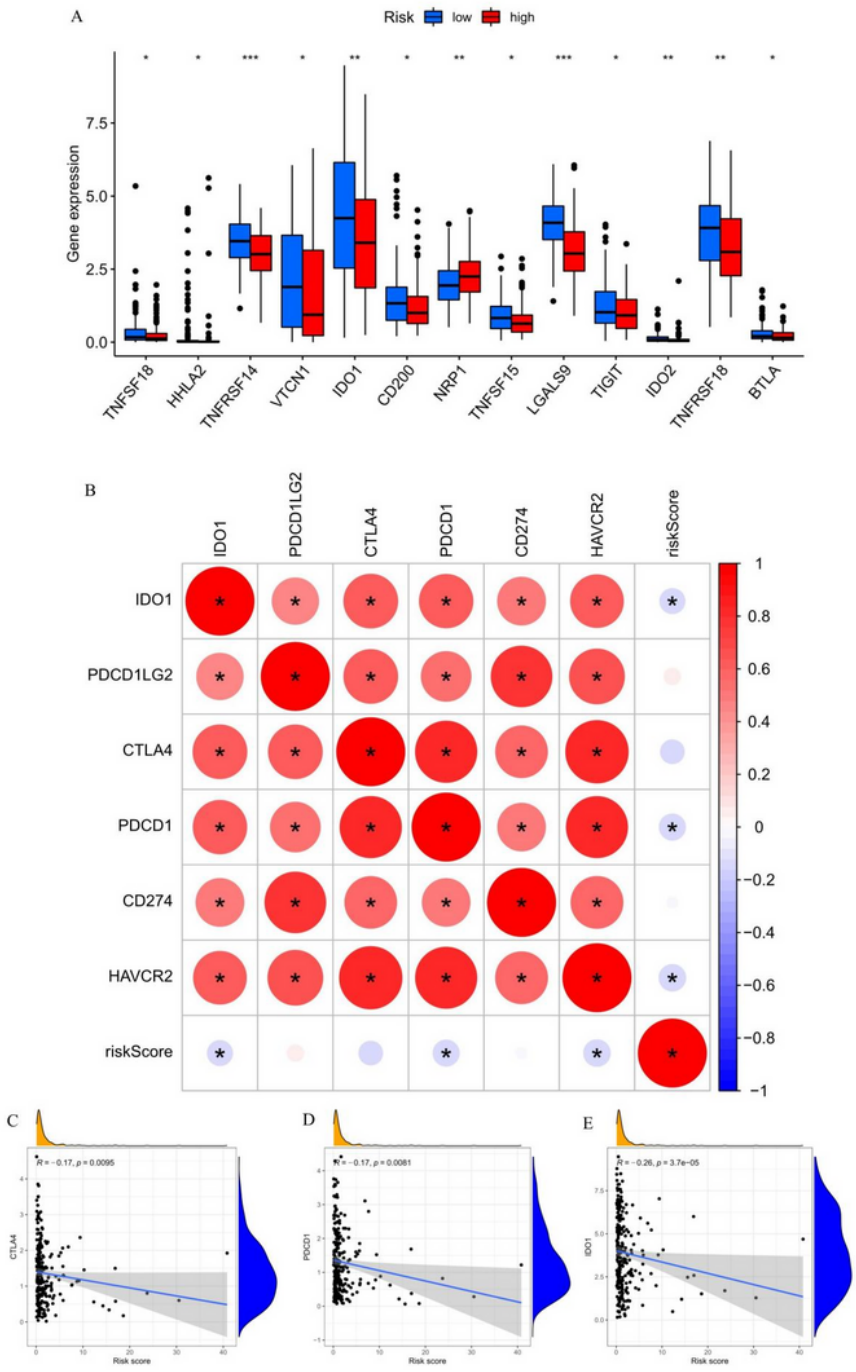

Figure 4

The key immune checkpoint genes are related to risk-score in CESC immune microenvironment. (A) The landscape of 26 types of immune checkpoint genes in low and high risk-score groups.***p□0.001□**p□0.01□*p□0.05. (B-E) The correlation of the risk-score and the 6 most important checkpoint genes (CD274,

PDCD1, PDCD1LG2, CTLA4, HAVCR2, IDO1). The bottom and top of the boxes are the 25th and 75th percentiles (interquartile range). *:statistically significant; red: positive correlation, blue: negative correlation. CESC, the cervical squamous cell carcinoma

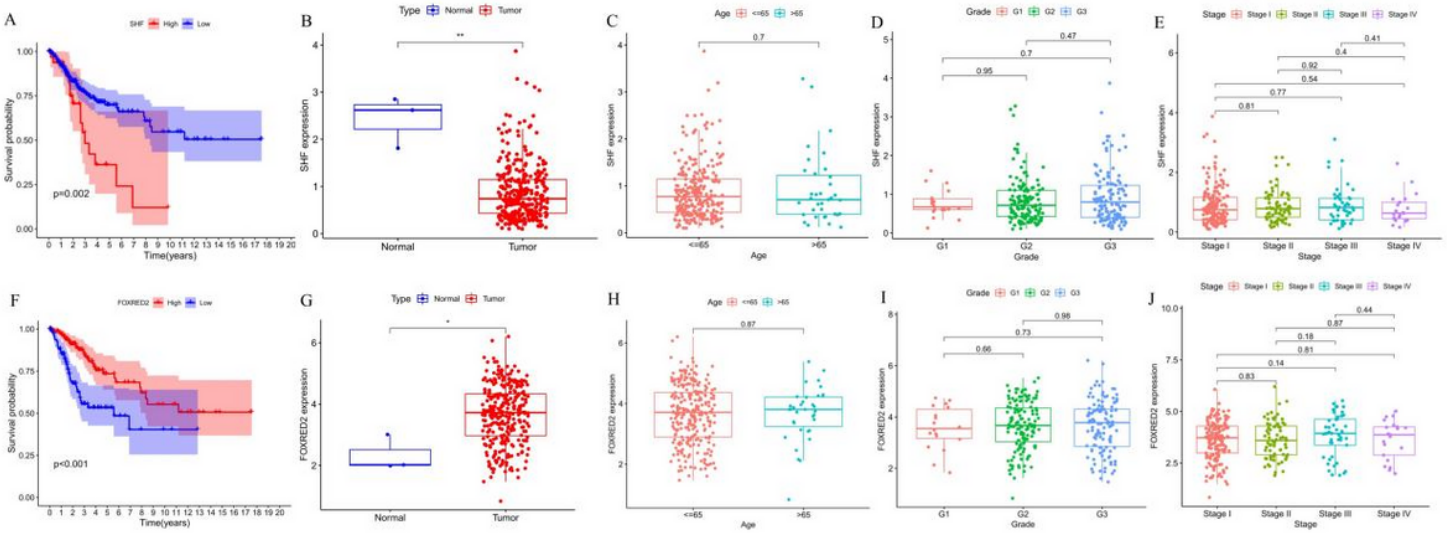

Figure 5

The prognostic signature of SHF and FOXRED2 expression SHF (B-E) and FOXRED2 (G-J) expression in type, age, grade, and stage groups. Kaplan-Meier survival curve of SHF (A) and FOXRED2 (F) in high expression and low expression groups. The bottom and top of the boxes respectively show the 25th and 75th percentiles (interquartile range). **p□0.01□*p□0.05.

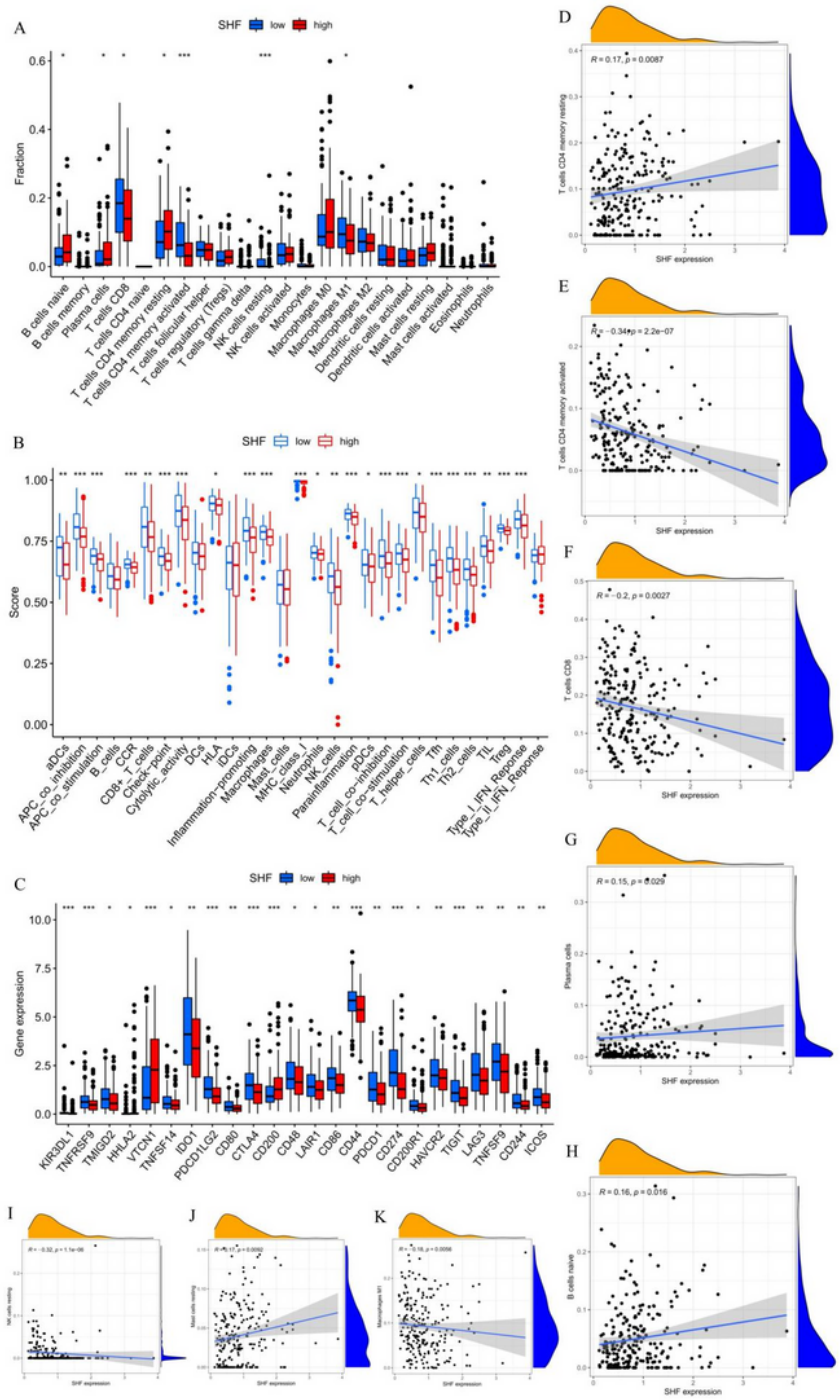

Figure 6

The relationship of the SHF expression and immune cells infiltrating in CESC immune microenvironment. (A) The box plot shows the proportion difference of each immune cell between the low and high expression group of SHF. (B) The landscape of infiltrating immune cells and immune function in 2 groups. (C) The expression of SHF is associated with the key immune checkpoint genes in CESC microenvironment. (D-K) Correlation between the SHF expression and the infiltration of immune cells. The

bottom and top of the boxes are the 25th and 75th percentiles (interquartile range). Blue: low risk, red: high risk. ***p□0.001, **p□0.01, *p□0.05. CESC, the cervical squamous cell carcinoma

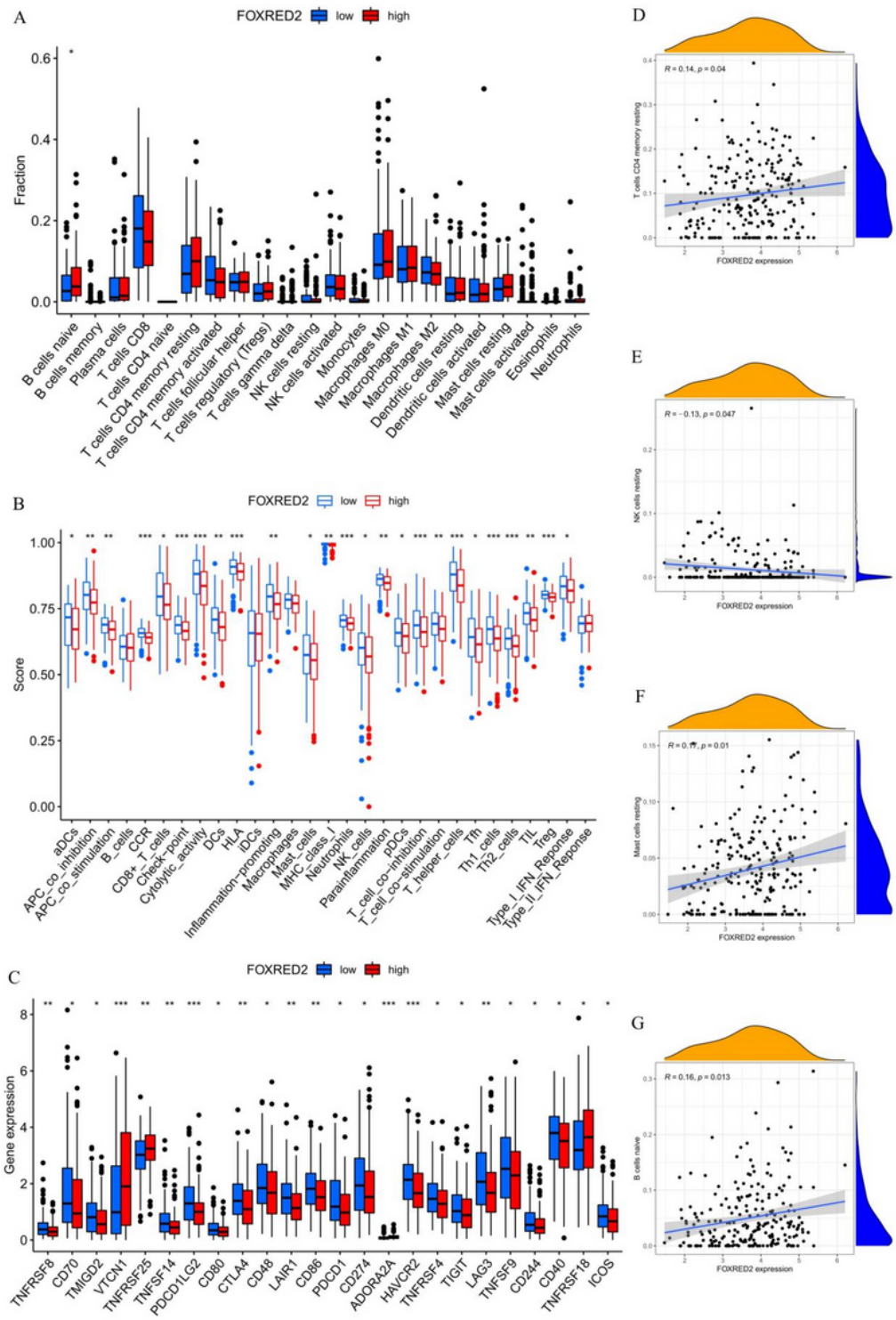

Figure 7

The relationship of FOXRED2 expression and immune cells infiltrating in CESC immune microenvironment. (A) The box plot shows a difference in the proportion of each immune cell between the low- and high-expression group of FOXRED2. (B) The landscape of infiltrating immune cells and

immune function in 2 groups. (C) The expression of FOXRED2 is related to the key immune checkpoint genes in TME. (D-G) Correlation between FOXRED2 expression and the infiltration of immune cells. The bottom and top of the boxes are the 25th and 75th percentiles (interquartile range). Blue: low risk, red: high risk. ***p<0.001, **p<0.01, *p<0.05. CESC, the cervical squamous cell carcinoma; TME, tumor microenvironment

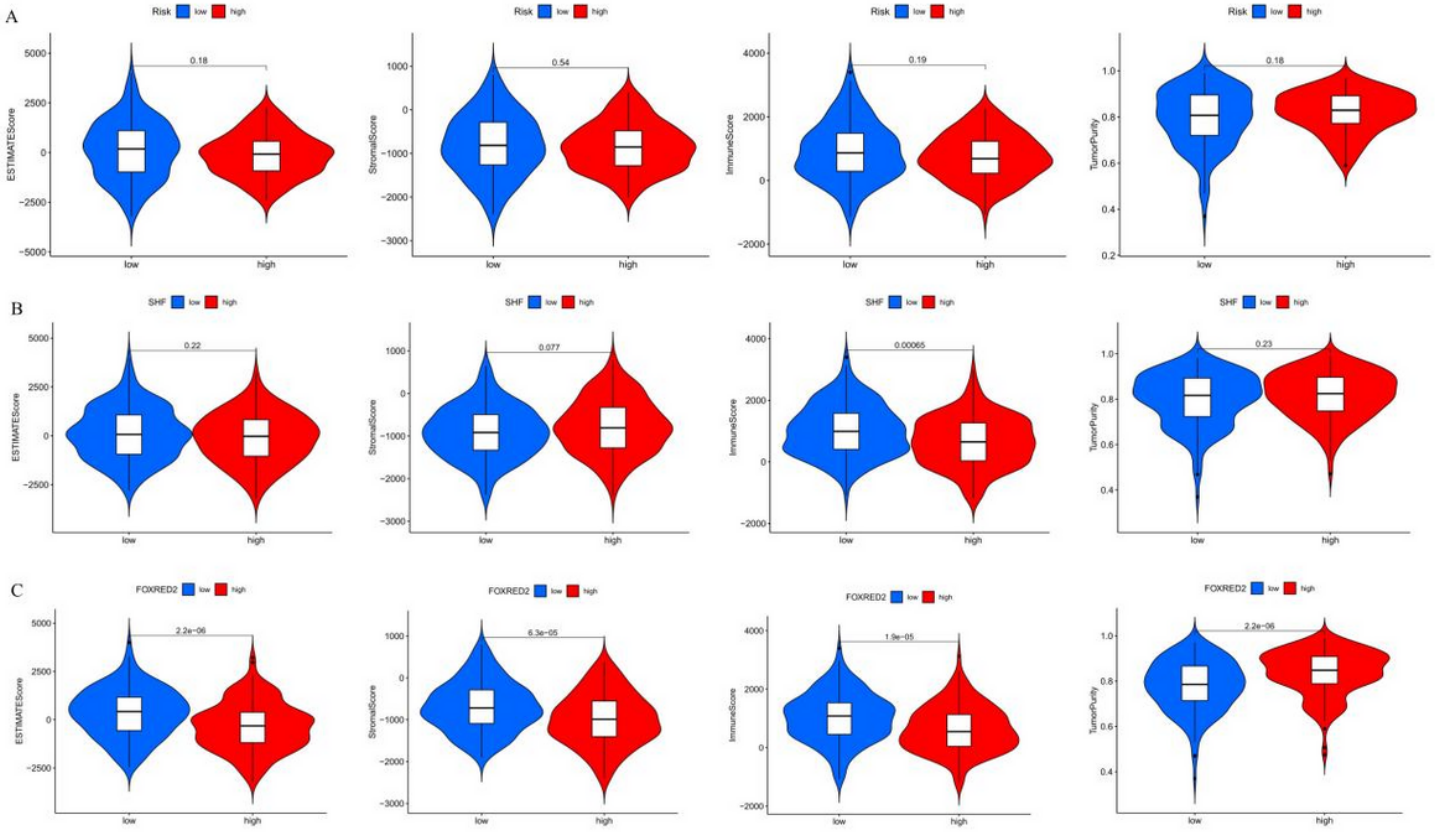

Figure 8

Construction and verification of the immune-related CESC subgroups. (A) The violin plot shows the difference in tumor purity, ESTIMATE score, immune score, and stromal score calculated using the ESTIMATE algorithm between the high and low risk-score groups. The stromal score, immune score, ESTIMATE score, and tumor purity were compared between 2 groups of SHF (B) and FOXRED2 (C). CESC, the cervical squamous cell carcinoma

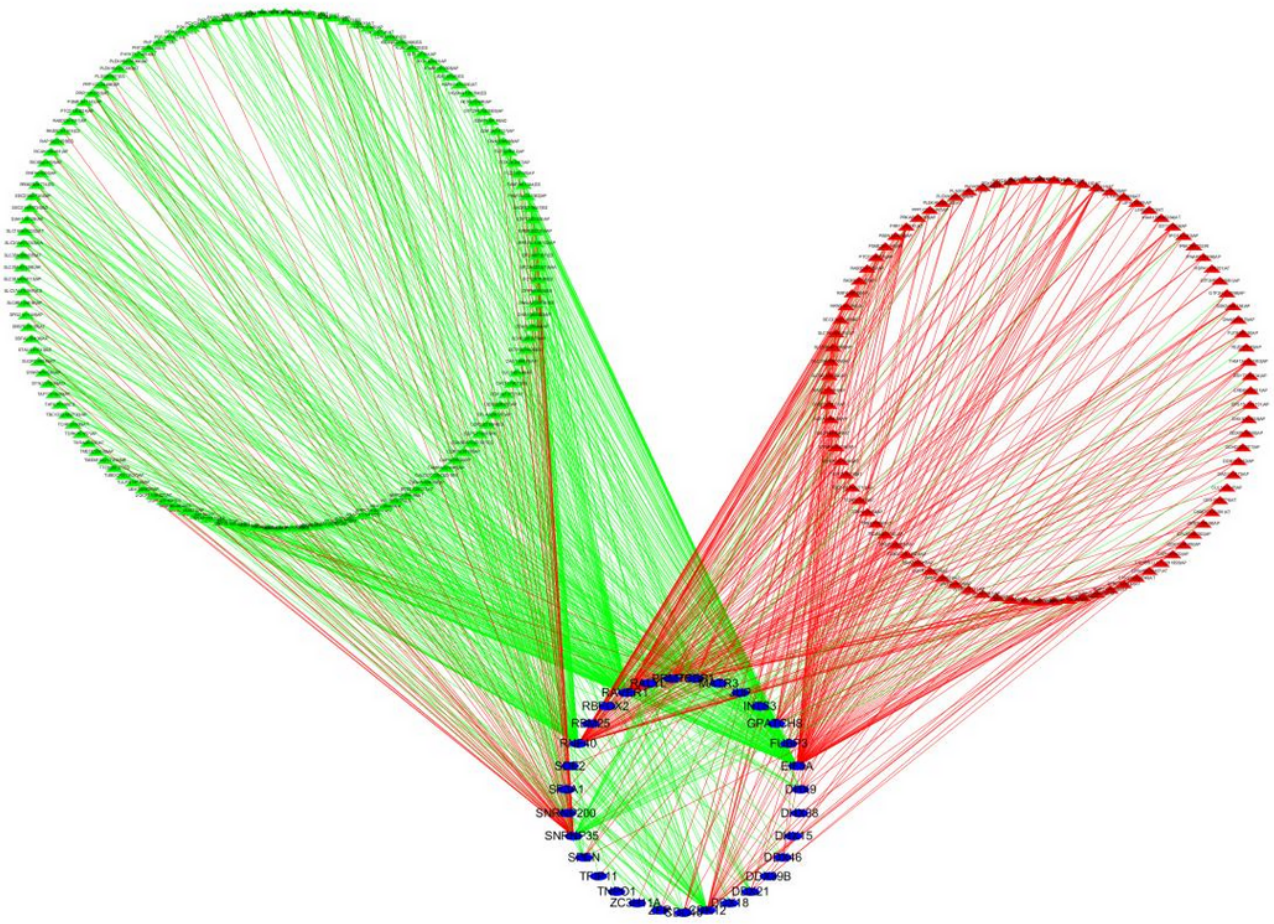

Figure 9

The regulatory network between SFs and AS events. Blue circle: spliced factors, green triangle: low-risk AS events, red triangle: high-risk AS events, green line: negative correlation, red line: positive correlation.. SF, splicing factors; AS, alternative splicing