

# A Nine-Immune-Related Gene Prognostic Signature for Predicting Outcomes in High-Grade Serous Ovarian Cancer

**Yahui Jiang**

Shanghai Jiaotong University

**Tianjiao Lyu**

Shanghai Jiaotong University

**Tianyu Zhou**

Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital

**Yiwen Shi**

Shanghai Jiaotong University

**Weiwei Feng** (✉ [fww12066@rjh.com.cn](mailto:fww12066@rjh.com.cn))

Shanghai Jiaotong University

---

## Research

**Keywords:** high-grade serous ovarian cancer, The Cancer Genome Atlas, immune-related genes, prognostic signature

**Posted Date:** August 30th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-801023/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Recently, immune system has been shown to be indispensable for ovarian cancer progression. The key immune-related genes (IRGs) related to the overall survival of ovarian cancer patients should be taken seriously. Here, we screened 9 survival-related IRGs in high-grade serous ovarian cancer (HGSOC) and build a prognostic signature to predict the outcome of HGSOC patients.

**Methods:** We downloaded RNA-sequence profiles from The Cancer Genome Atlas (TCGA) and Genome Tissue Expression (GTEx) databases to identify differentially expressed genes between normal fallopian tube and HGSOC. Among these genes, IRGs were filtered based on the Immunology Database and Analysis Portal (ImmPort). Using univariate Cox regression, Lasso regression and multivariate Cox regression, we selected 9 survival-related IRGs and established a prognostic signature to compute the risk score. Patients were divided into a low-risk group and a high-risk group, and the immunological feature differences between them were analysed with the ESTIMATE R package, TIMER and GSEA software. Moreover, the prognostic signature was validated by data from Gene Expression Omnibus (GEO) datasets.

**Results:** We obtained 1544 differentially expressed genes in HGSOC compared with normal fallopian tube, among which 99 genes were related to immunology. After univariate Cox regression, Lasso regression and multivariate Cox regression, nine IRGs (HLA-F, PSMC1, PI3, CXCL10, CXCL9, CXCL11, LRP1, STAT1 and OGN) were identified as optimal survival-related IRGs and used to establish a prognostic signature for calculating the risk scores of HGSOC patients. The prognostic signature showed its efficiency in predicting the overall survival of HGSOC patients in TCGA training cohort ( $p=1.018e^{-8}$ ) and GEO test cohort ( $p=2.632e^{-2}$ ). Age and risk scores were independent risk factors for overall survival. As the risk scores increased, the proportions of neutrophil, dendritic cells, CD8<sup>+</sup> T cells, CD4<sup>+</sup> T cells and B cells decreased ( $p$  values were 0.026,  $1.909e^{-4}$ ,  $9.165e^{-10}$ , 0.003 and  $2.658e^{-4}$ , respectively). In addition, 21 out of 24 HLA-related genes were highly expressed in the low-risk group than in the high-risk group. The above might prompt a stronger immune response in the low-risk group.

**Conclusions:** Our study constructed a nine-IRG-based prognostic signature that could effectively predict the overall survival of HGSOC patients and become a promising therapeutic target for HGSOC treatments.

## Background

Ovarian cancer is the most lethal gynaecological malignancy. It accounts for 2.5% of all malignancies among females but 5% of all cancer deaths. Among all the subtypes of ovarian cancer, 90% are epithelial, the most common being serous (52%). Indeed, the incidence of serous ovarian cancer varies by race. The incidence of serous carcinoma of the non-Hispanic white ranks first in all races (5.2 per 100,000), while the one of the Asian/Pacific Islander ranks last (alongside Non-Hispanic black; 3.4 per 100,000). Regrettably, the 5-year survival rate for serous ovarian cancer is 43% in all races (47% in Non-Hispanic

white, 36% in Non-Hispanic black and 47% in Asian/Pacific Islander), which casts light on treatments of serous ovarian cancer.(1)

Epithelial malignancies can be further divided into type I and type II. Type I ovarian cancers are considered low grade ones, which account for only a small fraction of ovarian cancer death. However, type II ovarian cancers are high grade and characterized by aggressive behaviour and poor outcomes. They are thought to originate as fallopian tube carcinomas that spread to the ovaries or peritoneum. Type II cancers are primarily high-grade serous ovarian cancer (HGSOC) that is the most common epithelial subtype. (1, 2) Hence, identifying potential survival-related biomarkers of HGSOC could be beneficial to ovarian cancer treatment and prognosis.

The immune system has a dual effect on ovarian cancer outcomes, which leads to tumour inhibition or progression depending on the function of prevalent immune cell subsets. For instance, ovarian tumours with high T cell contents had better 5-year overall survival rate than ovarian tumours with low T cell contents, at 38% vs 4.5%, respectively.(3, 4) This makes immunotherapy a promising treatment for ovarian cancer. Currently, immunotherapies for ovarian cancer mainly include three parts: (1) monoclonal antibodies as receptor mediators, including immune checkpoint inhibitors (targeting cytotoxic lymphocyte-associated antigen 4 and programmed death receptor); (2) cancer vaccines (one notable example is Vigil®); (3) adoptive immunotherapies alone or in combination with other approaches (including natural killer cells, cytokine-induced killer cells and chimaeric antigen receptor T cells, etc.). (5) However, the applications and effects of existing immunotherapies in ovarian cancer are still limited compared with other tumours.

Here, we aimed to screen for the differentially expressed immune-related genes (IRGs) between fallopian tube (FT) and HGSOC and conduct a prognostic IRG model to predict the survival outcomes of HGSOC patients. In addition, we further investigated the immunological features of tumours in the low-risk and high-risk groups identified according to the model. Our study revealed a prognostic IRG model of serous ovarian cancer consisting of 9 IRGs, which provided new biomarkers and new immune therapeutic targets in HGSOC.

## Methods

### TCGA, GTEx and GEO data acquisition

To obtain the RNA-Seq gene expression profile of HGSOC tissues and FT tissues, we downloaded the TCGA and GTEx gene expression profiles re-computed from raw RNA-Seq profiles by the USCS Xena project. The clinical data (including age, stage, grade, overall survival and survival state) for corresponding ovarian cancer patients were downloaded at the same time. The imbalance between the TCGA and GTEx data, which might cause heterogeneity, was normalized using the affy Bioconductor library `NormalizeBetweenArrays` by R software. According to the NCCN Guideline Version 2020 of

Epithelial Ovarian Cancer, grade 2 serous is considered high-grade, therefore we include the serous ovarian cancer patients with grade 2, grade 3 and grade 4 into our TCGA training cohort.

We then screened for RNA-Seq gene expression datasets of HGSOC that had patient prognosis information in the GEO database for the sake of validating the IRG-based prognostic signature trained by TCGA data. GSE49997, GSE140082, GSE26712 and GSE32062 were selected as the test groups. The batch effect between the TCGA and GEO data was removed via the “sva” package in R software.

### **Identification of differentially expressed genes (DEGs) and IRGs**

There were 5 FT tissues and 341 HGSOC tissues in total. To identify DEGs between FT tissues and HGSOC tissues, we used the Wilcoxon test after replicate probes in the expression profile were averaged by the “limma” package in R software. The filtering criteria were set to false discovery rate (FDR) <0.05 and | Log2-fold change (FC) |>2.

The immunologically relevant gene list was downloaded from the ImmPort database (<https://www.immport.org>).<sup>(28)</sup> Then, the differentially expressed IRGs between FT tissues and HGSOC tissues were obtained by taking the intersection of DEGs and immunologically relevant genes in the list. The results were visualized by heatmaps and volcano plots via the “pheatmap” package.

### **Screening for survival-related IRGs**

After merging the clinical data with the corresponding IRG expression levels of certain HGSOC patients, we used the “survival” package in R software to screen for survival-related IRGs by univariate Cox proportional hazard regression analysis when the p value <0.05 in the TCGA training cohort. The p value and hazard ratio of survival-related IRGs were visualized by forest plots.

### **Construction of IRG prognostic signature**

We performed least absolute shrinkage and selection operator (LASSO) regression analysis to search for IRGs with good prognostic value by using the “glmnet” package in R software. Then, the “survival” package was used to construct the IRG regression model for predicting the survival outcome of HGSOC patients by managing multivariate Cox regression. The risk scores of patients in the TCGA training cohort and GEO test cohort were calculated at the same time by utilizing the expression levels of IRGs involved in the regression model and its regression coefficients. The risk scoring model is shown as follows:

$$\text{Risk score} = \sum_{i=1}^n \text{Exp}_i * \text{Coe}_i$$

Here, “ $\text{Exp}_i$ ” represents the expression level of prognostic IRGs involved in the regression model, and “ $\text{Coe}_i$ ” represents the regression coefficients of prognostic IRGs.

The accuracies of this prognostic IRG model in the TCGA training cohort and GEO test cohort were evaluated using a time-dependent receiver operating characteristic (ROC) curve obtained via the “survivalROC” package in R software. The area under the ROC curve was calculated to determine the effectiveness of the IRG prognostic signature.

### **Survival analysis of IRG prognostic signature**

Patients in the TCGA training cohort and GEO test cohort were divided into a low-risk group and a high-risk group, respectively, according to the average risk scores of each cohort. In each cohort, Kaplan-Meier curves of the low-risk group and high-risk group were plotted, and survival analysis was performed using the log-rank test with the “survival” and “survminer” packages in R software.

To show the relationship between risk scores and HGSOC patient survival status more intuitively, the “pheatmap” package was used to draw risk plots that illustrated the risk score distribution and its relationship with survival status. The expression level of prognostic IRGs was also visualized in the form of a heatmap sorted according to the risk scores.

### **Construction of prognostic predictive nomogram**

A prognostic predictive nomogram was obtained via the “rms” package based on patients’ clinical characteristics (including age, pathological grade and stage) and the risk scores derived from the above methods. Then, the one-, three- and five-year survival rate of HGSOC patients could be predicted using this nomogram by calculating the total score.

### **Computational methods of multiple immune-related indexes and immunocyte infiltration status**

The stromal scores, immune scores and ESTIMATE scores of HGSOC tissues were calculated by the “estimate” package of R software. The stromal score and immune score represent stromal content and immune infiltration in tumour tissue, respectively. Tumour purity was reflected by the ESTIMATE score. The immune cell infiltration status of tumour tissues in the TCGA training cohort was downloaded from the Tumour Immune Estimation Resource (TIMER) database (<https://cistrome.shinyapps.io/timer/>).<sup>(29)</sup> TIMER is a web resource for systematic evaluation of the clinical impact of different immune cells in diverse cancer types. The abundance of six immune cell types, B cells, CD4 T cells, CD8 T cells, neutrophils, macrophages and dendritic cells (DCs), in the tumour microenvironment was estimated using a novel statistical method. The correlation test between immune cell infiltration status and risk score was performed by R software.

### **Functional enrichment analysis**

The functional enrichment analysis of gene expression profiles between low-risk and high-risk patients was performed by Gene Set Enrichment Analysis (GSEA) software (version 4.0.1).<sup>(30)</sup> We chose

the 5 significant immune-related pathways in the low-risk group and high-risk group. The results are shown in one plot using the “plyr”, “ggplot2”, “grid” and “gridExtra” packages in R software.

### **Identification of the potential transcription factors that regulate the expression of the nine prognostic IRGs**

The transcription factors which were involved in tumorigenesis were downloaded from <http://www.cistrome.org/>. First, we screened out the differentially expressed transcription factors in HGSOC tissues compared with FT. Then, we analysed the expression correlation between differentially expressed transcription factors and the nine prognostic IRGs by R software. The correlation filter was 0.3 and the p-value filter was 0.001. The results were visualized as a network using the Cytoscape software.

### **Statistical analysis**

All analyses were performed in R software (version 4.0.2) and perl (version 5.30.0). Statistical significance was defined as  $p < 0.05$ .

## **Results**

### **Identification of differentially expressed IRGs in HGSOC patients**

We downloaded and normalized the gene expression profiles of 5 FT tissues and 341 HGSOC from the GTEx and TCGA databases, respectively. Using the Wilcoxon test, we finally screened 1544 genes that were differentially expressed between FT and HGSOC (Figure 1A). Of these, 99 genes served as differentially expressed IRGs after intersection with the immunologically relevant gene list downloaded from the ImmPort database. The results were visualized by heatmaps and volcano plots and are shown in Figure 1B-C.

### **Evaluation of survival-related IRGs**

The inclusion criteria of the patients involved in our study were as follows: (1) the pathological type was HGSOC, (2) clinical parameters were completed, and (3) minimum follow-up of 90 days. Finally, 341 patients from the TCGA database were included in the training cohort to obtain the IRG prognostic signature (Additional table 1). We also aimed to use the gene expression profiles and clinical information of ovarian cancer patients from the GEO database to validate our IRG prognostic signature. Therefore, we selected 757 patients from GSE49997, GSE140082, GSE26712 and GSE32062 that met the inclusion criteria for the GEO test cohort. After normalizing the gene expression profiles in both the TCGA and GEO databases, nine survival-related IRGs were identified by performing univariate Cox proportional hazard regression analysis in the TCGA training cohort (Figure 2A).

### **Construction of the prognostic signature using the TCGA training cohort**

Through further Lasso regression analysis and multivariate Cox regression analysis, we obtained 9 optimal survival-related IRGs and combined them into a prognostic signature of HGSOc patients. The prognostic model included HLA-F, PSMC1, PI3, CXCL10, CXCL9, CXCL11, LRP1, STAT1 and OGN. The survival prognosis of a certain patient was predicted by calculating the risk score, which was the sum of each survival-related IRG above times its coefficient, as shown in Figure 2B. The comprehensive risk score was imputed as follows:  $(-0.08217 \times \text{expression level of HLA-F}) + (-0.12360 \times \text{expression level of PSMC1}) + (0.12159 \times \text{expression level of PI3}) + (0.28363 \times \text{expression level of CXCL10}) + (-0.07392 \times \text{expression level of CXCL9}) + (-0.39712 \times \text{expression level of CXCL11}) + (0.07471 \times \text{expression level of LRP1}) + (-0.03230 \times \text{expression level of STAT1}) + (0.11022 \times \text{expression level of OGN})$ . Patients in the TCGA training cohort were divided into low-risk and high-risk groups according to the median risk score, which was -0.15764. As shown in Figure 2C-D, the AUC (area under the curve) values for the prognostic signature at 1, 2, and 3 years were 0.622, 0.709 and 0.670, respectively, which proved the accuracy of this signature, and the survival prognosis of the low-risk group was significantly better than that of the high-risk group ( $p=1.018e^{-8}$ ). The 3-year survival rates of the low-risk and high-risk groups were 76.7% (95% CI: 69.8–84.2%) and 55.4% (95% CI: 47.8–64.2%), respectively. The 5-year survival rates of the low-risk and high-risk groups were 45.5% (95% CI: 37.2–55.5%) and 21.2% (95% CI: 15.1–29.6%), respectively. To present the relationship between the risk score and survival status more intuitively, we ranked the HGSOc patients according to risk score. As the risk score increased, the survival time decreased gradually, and the survival status was more likely to be death. The different expression patterns of those 9 IRGs in the low-risk group and high-risk group were visualized in the heatmap (Figure 2E-G). The expression of OGN, LRP1 and PI3 in the high-risk group was relatively higher than that in the low-risk group, and the expression of CXCL9, CXCL10, CXCL11, PSMC1, HLA-F and STAT1 was lower than that in the low-risk group.

### **Verification prognostic signature using GEO test cohort**

Next, we calculated the risk scores of 757 patients in the GEO test cohort (Additional table 2) according to the above formula. Using -0.15764 as the cut-off value, we divided 387 patients into the low-risk group and 370 patients into the high-risk group. In the GEO test cohort, the low-risk group also showed a better survival prognosis than the high-risk group ( $p=2.632 e^{-2}$ ). The 3-year survival rates of the low-risk and high-risk groups were 66.7% (95% CI: 61.6–72.3%) and 59.3% (95% CI: 54.0–65.2%), respectively. The 5-year survival rates of the low-risk and high-risk groups were 44.1% (95% CI: 37.6–51.7%) and 39.6% (95% CI: 33.1–47.4%), respectively (Figure 3A). The AUC values for the prognostic signature at 1, 2, and 3 years were 0.545, 0.581 and 0.572, respectively (Figure 3B). When patients were ranked in ascending order of the risk score, the survival prognosis (including status and time) worsened. The different expression patterns of these 9 IRGs in the low-risk and high-risk groups had the same tendency as those in the TCGA training cohort (Figure 3C-E).

### **Age and risk score were independent prognostic factors of HGSOc**

To evaluate the independent prognostic significance of the risk score in the TCGA training cohort, we conducted univariate Cox regression analyses and multivariate Cox regression analyses, in which other clinical characteristics (including age, grade, stage) were also involved. All the factors were treated as category variables. As a result, age and risk score were considered independent prognostic factors of HGSOE (Figure 4A-B). Elderly patients might have had a worse survival outcome (hazard ratio of death = 1.016 [95% CI, 1.002-1.029;  $p=0.020$ ], and patients who had a higher risk score were more likely to suffer a poor prognosis (hazard ratio of death = 2.876 [95% CI, 2.084-3.969;  $p<0.001$ ]). In addition, patients aged  $\geq 60$  years had a higher risk score than patients aged  $<60$  years ( $p=0.018$ ). However, there was no significant relationship between the risk score and pathological stage (Figure 4C-D). Finally, for better prediction of the 1-, 3-, and 5-year survival rates of HGSOE patients, we constructed a prognostic nomogram combining risk score, age and stage (Figure 4E). The total points were used to evaluate patient survival outcomes.

### **Immunological feature differences between the low-risk group and the high-risk group**

First, we investigated the contribution of immune cells and stromal cells in tumour tissues of the TCGA training cohort. Based on the ESTIMATE algorithm, the stromal score, immune score and ESTIMATE score were computed, which represented stromal content, immune infiltration and tumour purity in tumour tissue, respectively. Tumour tissues in the high-risk group have fewer immune elements ( $p=0.028$ ) and tend to have more stromal components, although the differences were not significant ( $p=0.126$ ). There was also no difference in tumour purity between these two groups ( $p=0.379$ ) (Figure 5A-C). Second, regarding the relationship between immune cell infiltration conditions and risk scores, we found that as the risk scores increased, the proportions of neutrophil, DCs, CD8<sup>+</sup> T cells, CD4<sup>+</sup> T cells and B cells decreased (the correlation indices were -0.122, -0.202, -0.326, -0.160 and -0.198; the  $p$  values were 0.026,  $1.909e^{-4}$ ,  $9.165e^{-10}$ , 0.003 and  $2.658e^{-4}$ , respectively). The risk scores had no significant correlation with macrophages ( $p=0.889$ ) (Figure 5D). Then, regarding the difference in HLA-related gene expression between the low-risk group and the high-risk group, we found that among 24 HLA-related genes, 21 genes were expressed at higher levels in the low-risk group than in the high-risk group (including HLA-F, HLA-E, HLA-DRB1, HLA-DRA, HLA-DQA2, HLA-DPB2, HLA-DPB1, HLA-DOB, HLA-DMB, HLA-DMA, HLA-C, HLA-B, HLA-A) (Figure 5E). This result indicated that patients in the low-risk group might have greater antigen processing and presentation capability. Finally, we also used gene set enrichment analysis (GSEA) to analyse the pathways enriched in patients in the low-risk group and high-risk group. There were 14 gene sets significantly enriched at  $p<0.05$  in the high-risk group and 24 gene sets significantly enriched at  $p<0.05$  in the low-risk group, which are listed in additional table 3. We selected 5 immune-related or meaningful pathways enriched in the low-risk group and high-risk group and showed them in Figure 5F. As a result, ECM receptor interaction, hedgehog signalling pathway, focal adhesion, adherens junction and tight junction were key pathways enriched in patients in the high-risk group. Intestinal immune network for IgA production, primary immunodeficiency, antigen processing and presentation, autoimmune thyroid disease and allograft rejection were key pathways enriched in patients in the low-risk group.

## Identification of the potential transcription factors that regulate the expression of the nine prognostic IRGs

There were 32 transcription factors which differentially expressed between FT and HGSOC. After expressive correlation analysis, we found 3 transcription factors could regulate the expression of the nine IRGs in prognostic signature. As shown in Figure 5G, FOS, NR4A1 and NR2F1 could upregulate the expression of LRP1. NR2F1 could not only upregulate the expression of OGN, but also downregulate the expression of CXCL11 and CXCL10.

## Discussion

Due to its nonspecific early clinical symptoms and lack of useful therapeutic strategies, ovarian cancer has been considered the leading cause of death among gynaecologic malignancies. Therefore, seeking effective therapeutic targets is extremely urgent for ovarian cancer treatments. Peritoneal metastasis is one of the dominant mechanisms of ovarian cancer metastasis. During this process, tumour cells dynamically interact with tumour microenvironment (TME) components, which could affect ovarian cancer progression and deeply influence disease prognosis by means of genetic or epigenetic methods.

(6) Both immune cells in the innate and adaptive immune systems, as well as cancer-associated fibroblasts interplaying with tumour cells, are involved in creating a tumour-promoting and immunosuppressive TME.(7) Apart from principle treatments such as surgery and chemotherapy, various immune-related therapies (such as atezolizumab and vigil) are under investigation in clinical trials as an effective ovarian cancer therapies. Some of them gained surprising and good outcomes.

(6) Hence, the immune system is a considerable link in ovarian cancer development.

Numerous studies have improved the prognostic value of IRGs in multiple tumours. For example, Guo et al. demonstrated that a risk model conducted by 9 optimal survival-related IRGs (HSPA6, CACYBP, DKK1, EGF, FGF19, OSM, GAST, ANGPTL3, NR2F2) could effectively predict the outcomes of oesophageal cancer patients.(8) Thus, we speculated that the expression of IRGs could also be associated with ovarian cancer patient outcomes. An immune-based prognostic score for ovarian cancer (IPSOV) was developed by Shen et al., and they reported that IPSOV and IPSOV-clinical integrated signatures could estimate the overall survival of ovarian cancer patients. However, IPSOV consisted of 129 IRGs belonging to 15 immune categories and defined as the combined effect of scores in different categories using the coefficients generated from the multivariable Cox regression model.

(9) Although IPSOV might provide a prognostic model that contains comprehensive IRGs, subsequent clinical application would be difficult because it requires the expression of large amounts of IRGs. Since HGSOC was deemed to derived from the malignant change of FT. Therefore, we first screened for IRGs between FT and HGSOC in the GTEx and TCGA databases. Then, we established a prognostic signature that consisted of 9 optimal IRGs and calculated the immune-related risk score of HGSOC using the data from the TCGA training cohort. According to the results of the survival analysis, patients who had a significantly higher risk score had worse overall survival. The ROC curve and risk plot proved that our 9-IRG-based prognostic signature could significantly distinguish the high-risk group among HGSOC

patients. Finally, we validated our prognostic signature in the GEO test cohort, which included 757 HGSOC patients. Similarly, patient prognosis in the high-risk group was poorer than that in the low-risk group according to the Kaplan-Meier curve and risk plot. Furthermore, after univariable analysis and multivariable Cox regression, it was found that age and risk score were independent risk factors for HGSOC.

In addition, there were several prognostic signatures of HGSOC which had already published based on the expression of other molecular subtypes from TCGA dataset. An article published in *Nature* termed the four HGSOC subtypes *Immunoreactive*, *Differentiated*, *Proliferative* and *Mesenchymal* based on gene content in the clusters. Using the integrated expression data set from 215 samples from TCGA training cohorts, a 193 gene transcriptional signature predictive of overall survival was defined. The predictive power was validated on a set of 255 TCGA test samples as well as three independent expression data sets, and the p-values of Log rank test were 0.0200, 0.0002, 0.0010 and 0.0050 respectively. Our nine-IRGs signature could also show statistically significant association with survival in TCGA train cohort and GEO test cohort, whose the p-values were  $1.018e^{-8}$  and 0.0263, respectively.(10)

The immune-related prognostic signature in our study could not only predict survival outcomes but also suggest that several neglected IRGs that might be promising targets for HGSOC treatments. Among the 9 IRGs involved in our prognostic signature, CXCL9, CXCL10 and CXCL11 are well investigated in ovarian cancer. Chemokines, interacting with chemokine receptors, can contribute to cell proliferation, inflammation, metastasis and tumorigenesis. Sonja Lieber reported that TAMs could produce chemokines CXCL9, CXCL10 and CXCL11, which attract CXCR3-expressing CD8<sup>+</sup> effector memory T cells from the periphery. CD8<sup>+</sup> effector memory T cells migrating into the TME contribute to tumour eradication and a better survival outcome.(11) It was also reported that high expression of CXCL9, CXCL11 and CXCR4 hinted at longer survival of HGSOC patients with the TP53 mutation.(12) Josahkian et al. reinforced the association of high STAT1 expression with better response to chemotherapy, which made it a prognostic biomarker of HGSOC.(13) In terms of immune regulation, STAT1 enhances anti-tumorigenic immune responses and proinflammation function via inducing cytokines secretion and increasing amount of anti-tumour immune lymphocytes.(14) However, although some studies have shown that the expression of HLA-F, PSMC1, PI3, LRP1 and OGN plays a role in tumorigenesis, few studies have shed light on the effects of the above genes on ovarian cancer tumorigenesis. According to the article of Feng et al., LRP1 is associated with worsened outcomes in clear-cell renal cell carcinoma and is related to cancer immune modulation.(15) OGN had different effects on tumorigenesis in different tumour types. OGN could promote meningioma development through downregulation of neurofibromatosis type 2 and activation of mTOR signalling.(16) However, in colorectal cancer, OGN expression is positively related to CD8<sup>+</sup> cell infiltration in the tumour niche and associated with better survival.(17) According to our results, OGN seems to act as an oncogene in serous ovarian cancer. Therefore, further studies should be performed to validate the impact of the above genes on ovarian cancer development.

To explore the immunological feature differences between the low-risk group and the high-risk group, we primarily computed the stromal score, immune infiltration score and tumour purity

in the tissues of these two groups. Recent studies have shown that the TME contributes to ovarian cancer progression.(18, 19) As the primary nontumour components of the TME, tumour-infiltrating immune cells and stromal cells might also be involved. We found that there were fewer immune cells in the high-risk group than in the low-risk group. Then, the relationships between the risk score and 6 types of immune cells were further studied. With the rise in risk score, the proportions of neutrophil, DCs, CD8<sup>+</sup> T cells, CD4<sup>+</sup> T cells and B cells were decreased significantly in serous tumour tissues. DCs are the first line of defences against exogenous pathogens and are actively involved in tumour surveillance by removing damaged tissues in the TME. Serving as antigen-presenting cells, DCs can initiate and regulate the antitumour immune response.(20) DC-based vaccine immunotherapy trials are increasing and may be of great prognostic benefit to ovarian cancer patients.(21) CD8<sup>+</sup> T cells are one of the dominant components of the endogenous immune response to ovarian cancer. When degraded proteins bind to human leukocyte antigen class I molecules, CD8<sup>+</sup> T cells can be triggered to secrete proinflammatory cytokines and kill tumour cells. Thus, CD8<sup>+</sup> T cell infiltration can predict prolonged overall survival in ovarian cancer.(22, 23) B cells usually play a role in generating an antibody response against antigens and interacting with other immune cells through antigen presentation and cytokine secretion. It is reported that CD20<sup>+</sup> B-cell tumour-infiltrating lymphocytes in ovarian cancer, non-small lung carcinoma and cervical cancer are correlated with improved survival and lower relapse rates. The potential mechanism is the secretion of effector cytokines, like IFN- $\gamma$ , which could promote T-cell responses as antigen-presenting cells. From the above, it can be seen that lack of antigen presentation and tumour cell killing effects may contribute to the shorter survival of patients in the high-risk group.(24, 25) Human leukocyte antigen (HLA) plays a crucial role in activating a host immune response against pathogens and tumour cells by distinguishing self and nonself peptides. HLA can be classified into 3 groups based on function and structure. Among them, HLA class I molecules (including HLA-A, -B, -C, -E, -F) and HLA class II molecules (including HLA-DP, -DQ, -DR) are more common. Several studies have shown that downregulation of the HLA class I antigen-derived peptide complex by cancer cells can lead to tumour immune escape and poor outcomes in cancer patients. HLA class II molecules promote the switch of naïve T cells into activated T cells by presenting exogenous antigen peptides to CD4<sup>+</sup> T cells.(26) In our study, the expression of HLA class I and II molecules in the low-risk group was higher than that in the high-risk group. According to the results of gene pathway enrichment, antigen processing and presentation were significantly enriched in the low-risk group. All of the above results indicated that the better survival in the low-risk group might result from better antigen presentation and an antitumour immune response.

In order to predict the common upstream transcription factors of the prognostic IRGs, we performed an expressive correlation analysis and found NR2F1 could regulate the expression of LRP1, OGN, CXCL10 and CXCL11. NR2F1 was reported could modulate gene expression during cancer development and tumour cell dormancy.(27) Therefore, our results suggested that NR2F1 could be a potential treatment target in HGSOC.

However, there are still some limits in our study. First, the retrospective design might cause some unavoidable bias. The results should be validated by further prospective studies. Second, although the ROC values of the signature in the TCGA training cohort were gratifying, they seemed moderate for predicting the overall survival of serous ovarian cancer patients in the GEO test cohort. Third, our results were obtained merely by bioinformatics analysis. Further experimental studies should be performed to clarify the functions and mechanisms of these 9 survival-related IRGs during HGSOC progression.

## Conclusions

We established an effective, nine-IRG-based prognostic signature of HGSOC and demonstrated that it served as an independent prognostic factor of overall survival in HGSOC patients. The risk score was negatively correlated with neutrophil, DCs, CD8<sup>+</sup> T cells, CD4<sup>+</sup> T cells and B cells infiltration in the TME. Further studies should be performed on these eleven IRGs, which could be promising new therapeutic targets for HGSOC treatments.

## Abbreviations

IRGs: immune-related genes

FT: fallopian tube

HGSOC: high grade serous ovarian cancer

TCGA: The Cancer Genome Atlas

GTEX: Genome Tissue Expression

GEO: Gene Expression Omnibus

PSMC1: proteasome 26S subunit, ATPase 1

PI3: peptidase inhibitor 3

CXCL10: C-X-C motif chemokine ligand 10

CXCL9: C-X-C motif chemokine ligand 9

CXCL11: C-X-C motif chemokine ligand 11

LRP1: low-density lipoprotein receptor related protein 1

OGN: osteoglycin

STAT1: signal transducer and activator of transcription 1

HLA: human leukocyte antigen

LASSO: least absolute shrinkage and selection operator

ROC: receiver operating characteristic

GESA: Gene Set Enrichment Analysis

AUC: area under curve

ECM: extracellular matrix

TME: tumour microenvironment

NR2F1: Nuclear Receptor Subfamily 2 Group F member 1

## Declarations

**Ethics approval and consent to participate:** Any data used in our study are from open access and do not require approval. Ethics approval and consent to participate are not applicable for them.

**Availability of data and materials:** The datasets supporting the conclusion of this article are available in TCGA repository (<https://tcga-data.nci.nih.gov/tcga/>), GTEx (<https://gtexportal.org/>) and GEO database (<https://www.ncbi.nlm.nih.gov/gds/>).

**Consent for publication:** Not applicable

**Competing interests:** The authors declare that they have no competing interests.

**Funding:** This work was supported by the Shanghai Municipal Education Commission-Gaofeng Clinical Medicine Grant (Grant No. 20172003), and Guangci Distinguished Young Scholars Training Program of Shanghai Jiaotong University School of Medicine affiliated Ruijin Hospital (GCQN-2019-B12)

**Authors` contributions:** All authors contributed to the study conception and implement. YJ was contributed to write the manuscript and perform the analysis. TL, TZ and YS helped to collected the data and perform the analysis. WF designed this study and was the corresponding author of this manuscript.

**Acknowledgements:** Not applicable

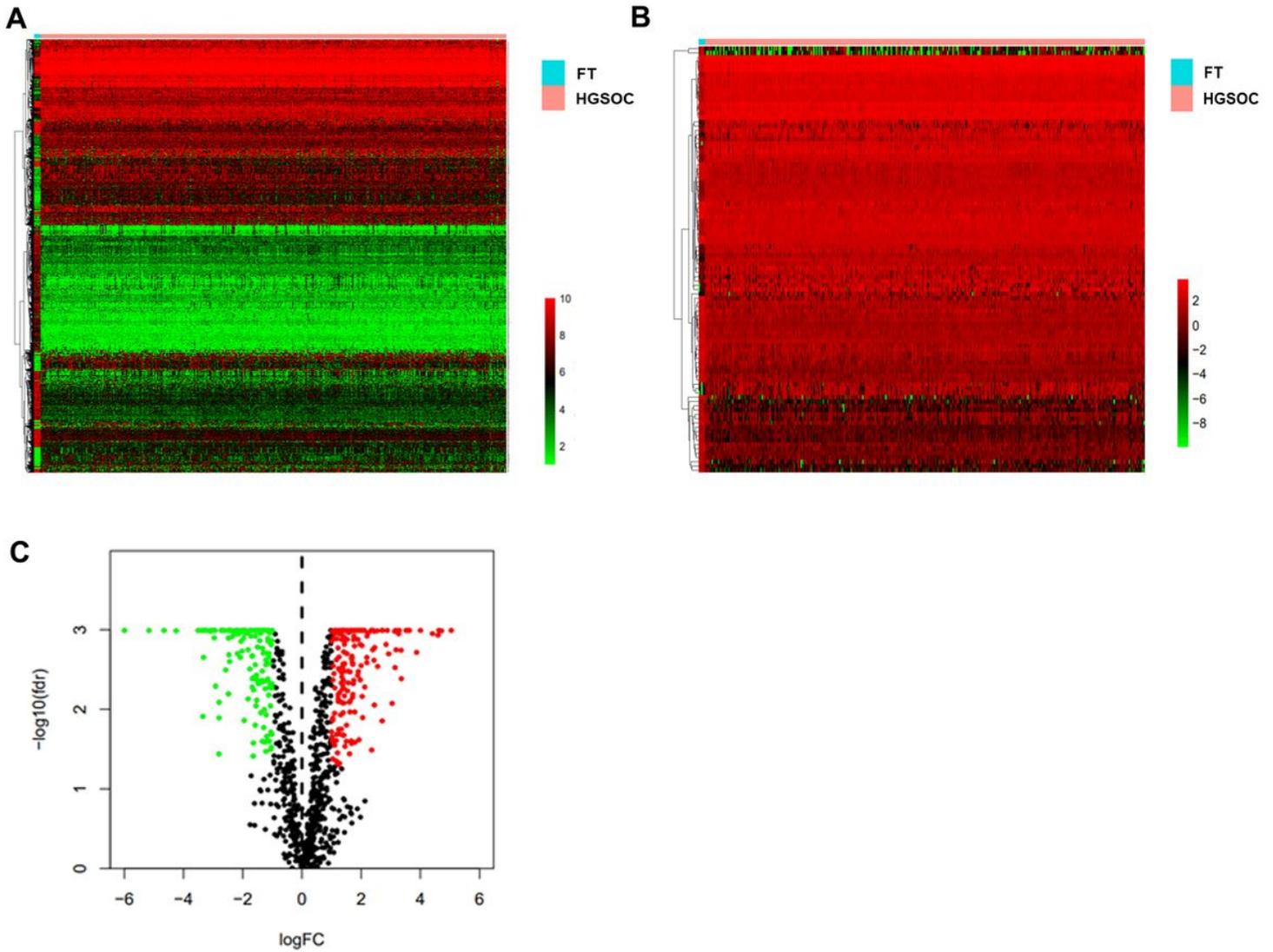
## References

1. Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, et al. Ovarian cancer statistics, 2018. *Cancer J Clin.* 2018;68(4):284–96.
2. Yang C, Xia BR, Zhang ZC, Zhang YJ, Lou G, Jin WL. Immunotherapy for Ovarian Cancer: Adjuvant, Combination, and Neoadjuvant. *Frontiers in immunology.* 2020;11:577869.

3. Zhang L, Conejo-Garcia JR, Katsaros D, Gimotty PA, Massobrio M, Regnani G, et al. Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. *N Engl J Med*. 2003;348(3):203–13.
4. Menderes G, Schwab CL, Black J, Santin AD. The Role of the Immune System in Ovarian Cancer and Implications on Therapy. *Expert Rev Clin Immunol*. 2016;12(6):681–95.
5. Fan CA, Reader J, Roque DM. Review of Immune Therapies Targeting Ovarian Cancer. *Curr Treat Options Oncol*. 2018;19(12):74.
6. Ghoneum A, Afify H, Salih Z, Kelly M, Said N. Role of tumor microenvironment in the pathobiology of ovarian cancer: Insights and therapeutic opportunities. *Cancer medicine*. 2018;7(10):5047–56.
7. Worzfeld T, Pogge von Strandmann E, Huber M, Adhikary T, Wagner U, Reinartz S, et al. The Unique Molecular and Cellular Microenvironment of Ovarian Cancer. *Front Oncol*. 2017;7:24.
8. Guo X, Wang Y, Zhang H, Qin C, Cheng A, Liu J, et al. Identification of the Prognostic Value of Immune-Related Genes in Esophageal Cancer. *Frontiers in genetics*. 2020;11:989.
9. Shen S, Wang G, Zhang R, Zhao Y, Yu H, Wei Y, et al. Development and validation of an immune gene-set based Prognostic signature in ovarian cancer. *EBioMedicine*. 2019;40:318–26.
10. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011;474(7353):609–15.
11. Lieber S, Reinartz S, Raifer H, Finkernagel F, Dreyer T, Bronger H, et al. Prognosis of ovarian cancer is associated with effector memory CD8(+) T cell accumulation in ascites, CXCL9 levels and activation-triggered signal transduction in T cells. *Oncoimmunology*. 2018;7(5):e1424672.
12. Ignacio RMC, Lee ES, Wilson AJ, Beeghly-Fadiel A, Whalen MM, Son DS. Chemokine Network and Overall Survival in TP53 Wild-Type and Mutant Ovarian Cancer. *Immune network*. 2018;18(4):e29.
13. Josahkian JA, Saggiaro FP, Vidotto T, Ventura HT, Candido Dos Reis FJ, de Sousa CB, et al. Increased STAT1 Expression in High Grade Serous Ovarian Cancer Is Associated With a Better Outcome. *International journal of gynecological cancer: official journal of the International Gynecological Cancer Society*. 2018;28(3):459–65.
14. Li X, Wang F, Xu X, Zhang J, Xu G. The Dual Role of STAT1 in Ovarian Cancer: Insight Into Molecular Mechanisms and Application Potentials. *Frontiers in cell developmental biology*. 2021;9:636595.
15. Feng C, Ding G, Ding Q, Wen H. Overexpression of low density lipoprotein receptor-related protein 1 (LRP1) is associated with worsened prognosis and decreased cancer immunity in clear-cell renal cell carcinoma. *Biochem Biophys Res Commun*. 2018;503(3):1537–43.
16. Mei Y, Du Z, Hu C, Greenwald NF, Abedalthagafi M, Agar NYR, et al. Osteoglycin promotes meningioma development through downregulation of NF2 and activation of mTOR signaling. *Cell communication signaling: CCS*. 2017;15(1):34.
17. Hu X, Li YQ, Li QG, Ma YL, Peng JJ, Cai SJ. Osteoglycin-induced VEGF Inhibition Enhances T Lymphocytes Infiltrating in Colorectal Cancer. *EBioMedicine*. 2018;34:35–45.
18. Ding Q, Dong S, Wang R, Zhang K, Wang H, Zhou X, et al. A nine-gene signature related to tumor microenvironment predicts overall survival with ovarian cancer. *Aging*. 2020;12(6):4879–95.

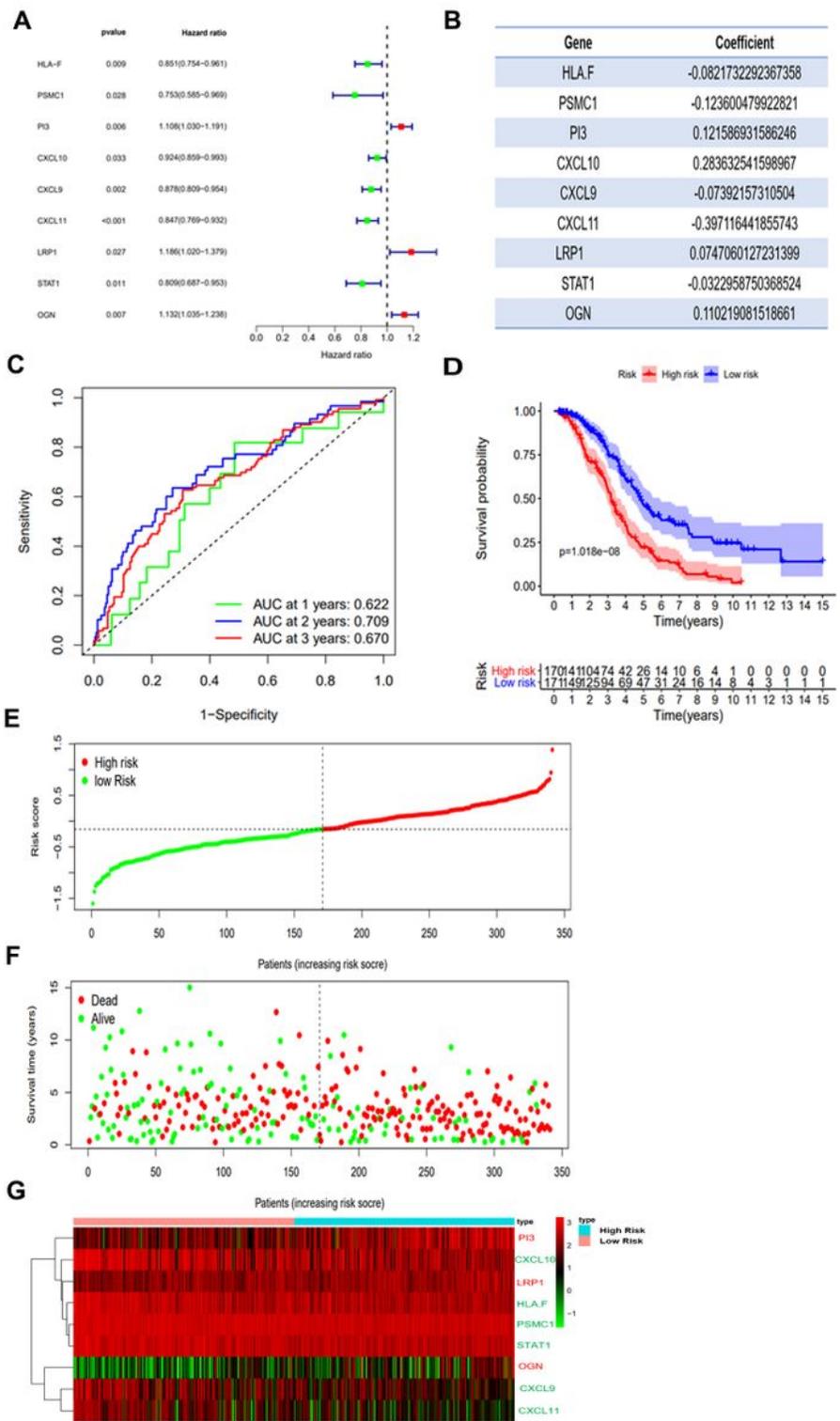
19. Curtis M, Mukherjee A, Lengyel E. The Tumor Microenvironment Takes Center Stage in Ovarian Cancer Metastasis. *Trends in cancer*. 2018;4(8):517–9.
20. Chiang CL, Kandalaft LE. In vivo cancer vaccination: Which dendritic cells to target and how? *Cancer treatment reviews*. 2018;71:88–101.
21. Drakes ML, Stiff PJ. Understanding dendritic cell immunotherapy in ovarian cancer. *Expert Rev Anticancer Ther*. 2016;16(6):643–52.
22. Goode EL, Block MS, Kalli KR, Vierkant RA, Chen W, Fogarty ZC, et al. Dose-Response Association of CD8 + Tumor-Infiltrating Lymphocytes and Survival Time in High-Grade Serous Ovarian Cancer. *JAMA oncology*. 2017;3(12):e173290.
23. Sato E, Olson SH, Ahn J, Bundy B, Nishikawa H, Qian F, et al. Intraepithelial CD8 + tumor-infiltrating lymphocytes and a high CD8+/regulatory T cell ratio are associated with favorable prognosis in ovarian cancer. *Proc Natl Acad Sci USA*. 2005;102(51):18538–43.
24. Gupta P, Chen C, Chaluvally-Raghavan P, Pradeep S. B Cells as an Immune-Regulatory Signature in Ovarian Cancer. *Cancers*. 2019;11(7).
25. Sarvaria A, Madrigal JA, Saudemont A. B cell regulation in cancer and anti-tumor immunity. *Cell Mol Immunol*. 2017;14(8):662–74.
26. Sabbatino F, Liguori L, Polcaro G, Salvato I, Caramori G, Salzano FA, et al. Role of Human Leukocyte Antigen System as A Predictive Biomarker for Checkpoint-Based Immunotherapy in Cancer Patients. *International journal of molecular sciences*. 2020;21(19).
27. Gao XL, Zheng M, Wang HF, Dai LL, Yu XH, Yang X, et al. NR2F1 contributes to cancer cell dormancy, invasion and metastasis of salivary adenoid cystic carcinoma by activating CXCL12/CXCR4 pathway. *BMC Cancer*. 2019;19(1):743.
28. Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, et al. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Scientific data*. 2018;5:180015.
29. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic acids research*. 2020;48(W1):W509-w14.
30. 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. *Proc Natl Acad Sci USA*. 2005;102(43):15545–50.

## Figures



**Figure 1**

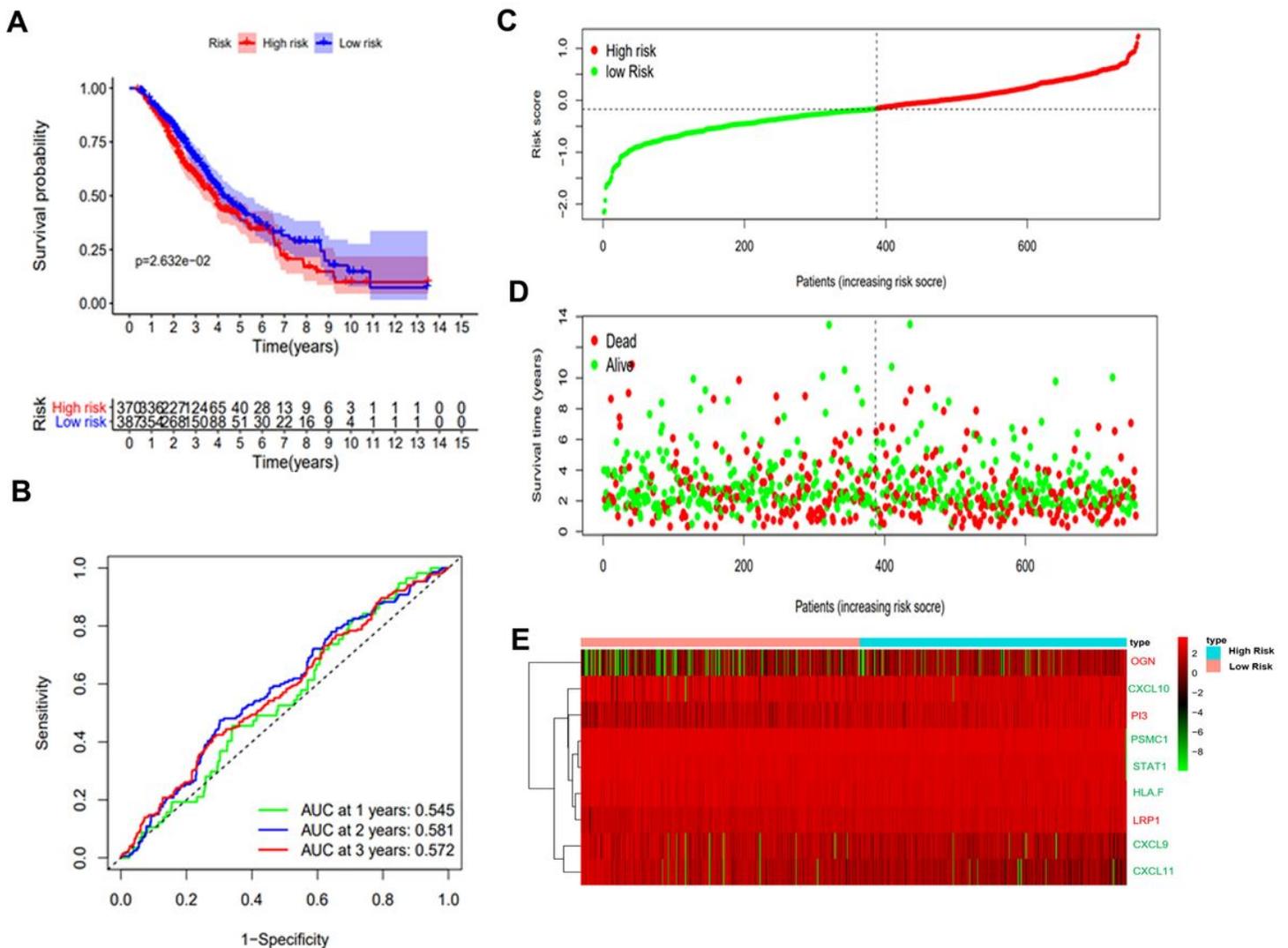
Differentially expressed genes between normal ovarian tissues and HGSOC tissues. (A) Heatmap of differentially expressed genes between FT tissues and HGSOC tissues. The gene expression values ranged from 0 to 10, which visualized by a transition from green to red. (B) Heatmap of differentially expressed IRGs between FT tissues and HGSOC tissues. The gene expression values ranged from -10 to 4, which visualized by a transition from green to red. (C) Volcano plot of IRGs between FT tissues and HGSOC tissues. The green and the red spots represent IRGs that significantly decreased and increased in HGSOC compared with FT. The black spots represent IRGs that have no statistical difference in FT and HGSOC.



**Figure 2**

Construction of a nine-IRG-based prognostic signature for HGSOC patients in the TCGA training cohort. (A) Forest plot of hazard ratios belonging to 9 prognosis-related IRGs obtained from univariate Cox proportional hazard regression analysis. A hazard ratio of a certain IRG more than 1 means the expression of this IRG has a negative effect on overall survival. A hazard ratio of a certain IRG less than 1 means the expression of this IRG has a positive effect on overall survival. (B) Coefficient value of 9 IRGs

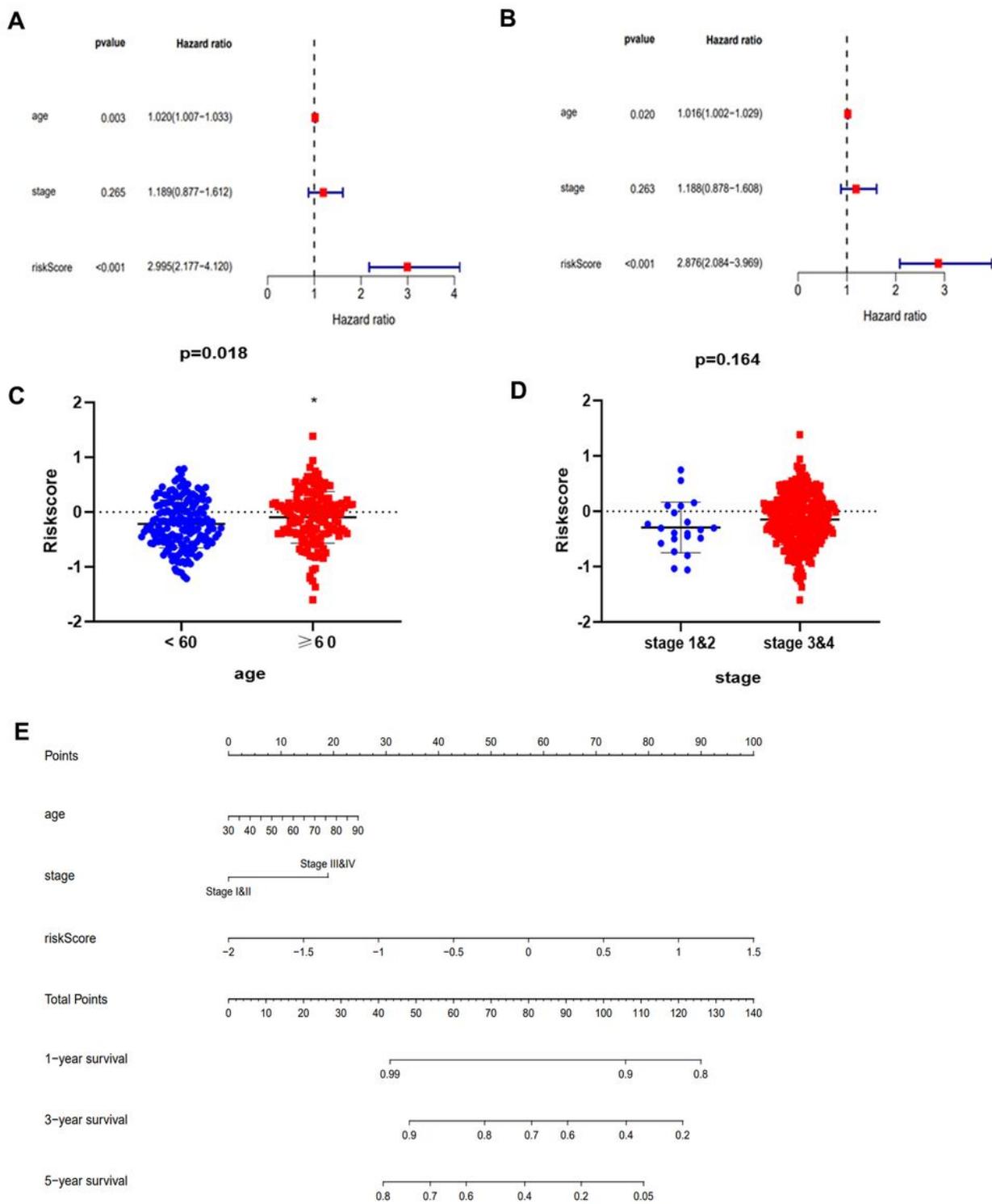
involved in the prognostic signature. (C) ROC curves of the prognostic signature with AUCs of 0.622, 0.709 and 0.670 at 1, 2, and 3 years, indicating that the risk score had good effectiveness in predicting the overall survival of HGSOC patients. (D) The Kaplan-Meier plot for overall survival of the low-risk group and high-risk group divided by risk score. (E-F) The risk plot of HGSOC patients in the TCGA training cohort. (E) shows the risk score contribution. Green spots represent the risk scores in the low-risk group. Red spots represent the risk scores in the high-risk group. (F) shows the survival years and status of patients with different risk scores. Green spots represent a dead outcome. Red spots represent an alive outcome. (G) The heatmap demonstrates the expression of 9 IRGs in the prognostic signature among patients in the low-risk group and high-risk group. The IRGs in red font colour mean that the expression of these genes is higher in high-risk group compared with it in low-risk group. Green font colour mean that the expression of these genes is lower in high-risk group.



**Figure 3**

Validation of the nine-IRG-based prognostic signatures in the GEO testing cohort. (A) The Kaplan-Meier plot for overall survival of the low-risk group and high-risk group divided by risk score. (B) ROC curves of

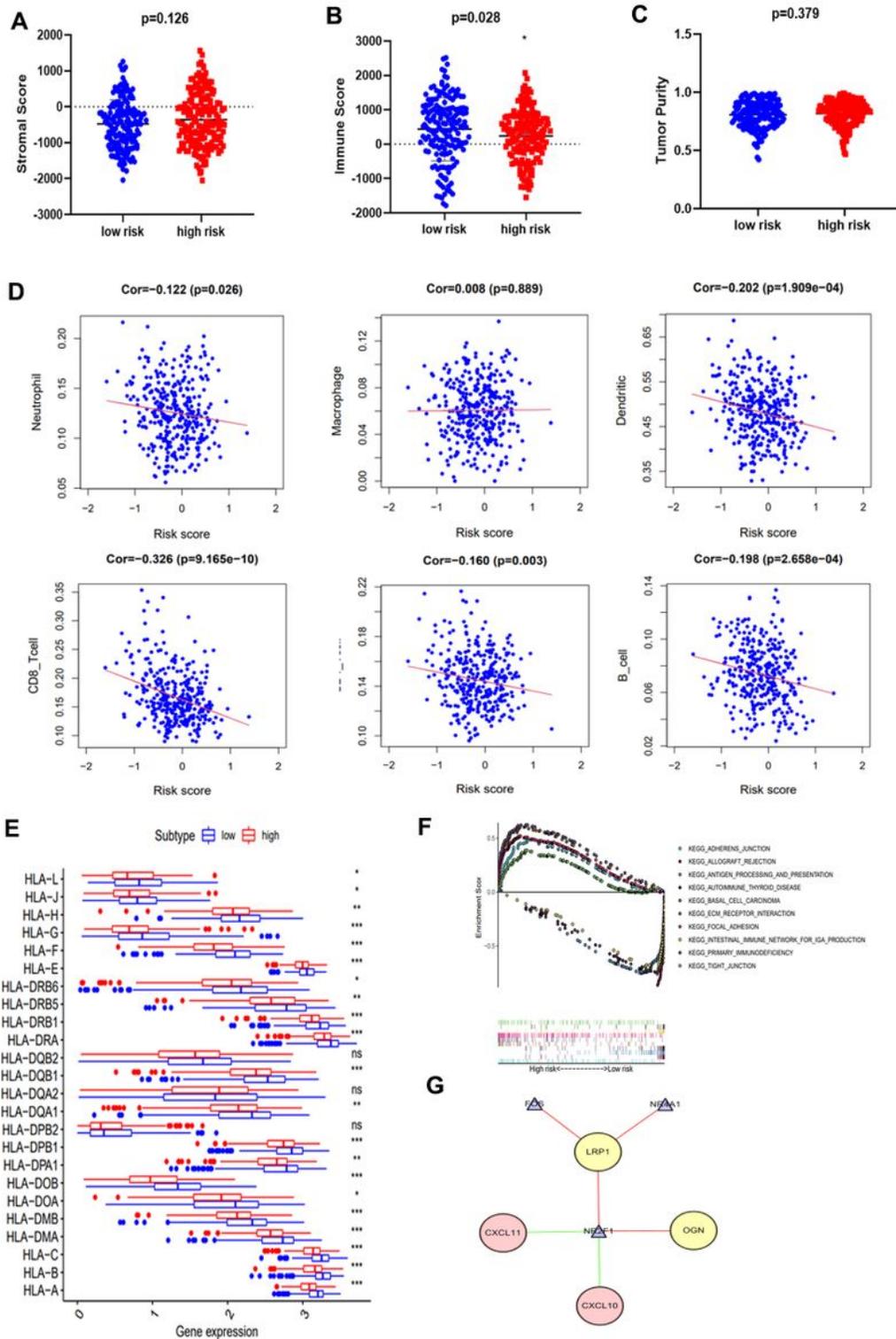
the prognostic signature with AUCs of 0.545, 0.581 and 0.572 at 1, 2, and 3 years in the GEO test cohort. (C-D) The risk plot of HGSOC patients in the TCGA training cohort. (C) shows the risk score contribution. Green spots represent the risk scores in the low-risk group. Red spots represent the risk scores in the high-risk group. (D) shows the survival years and status of patients with different risk scores. Green spots represent a dead outcome. Red spots represent an alive outcome. (E) The heatmap demonstrates the expression of 9 IRGs in the prognostic signature among patients in the low-risk group and high-risk group. The IRGs in red font colour mean that the expression of these genes is higher in high-risk group compared with it in low-risk group. Green font colour mean that the expression of these genes is lower in high-risk group.



**Figure 4**

Prognostic value of the nine-IRG-based prognostic signature. (A-B) The forest plot shows the prognostic value risk scores and clinical characteristics by univariate (A) and multivariate (B) regression analysis. Age and risk scores were independent risk factors for overall survival in the TCGA training cohort. (C-D) The relationship between risk scores and age (C) or pathological stage (D) of HGSOC patients. (E) A nomogram predicts the outcome of HGSOC patients based on their risk scores and clinical

characteristics. The total points of age, stage and risk score can predict the overall survival of HGSOC patients at 1, 3, 5 years.



**Figure 5**

Immunological feature differences between the low-risk group and the high-risk group. (A-C) Differences in stromal score (A), immune score (B) and tumour purity (C) between the low-risk group and the high-risk group. (D) The correlation between risk scores and immune components (including B cells, CD4+ T cells,

CD8+ T cells, dendrites, macrophages and neutrophils). (E) The expression of HLA-related genes in the low-risk group and high-risk group. “\*” means  $p < 0.05$ . “\*\*” means  $p < 0.01$ . “\*\*\*” means  $p < 0.001$ . “ns” means no significance. (F) KEGG analysis of GSEA in the low-risk group and high-risk group. The KEGG pathways above the x-axis are enriched in the high-risk group. And the KEGG pathways below the x-axis are enriched in the low-risk group. (G) The regulatory network of TFs and IRGs in the nine-IRG-based prognostic signature. The genes in the purple triangles are TFs. The genes in the yellow and pink circles are high-risk IRGs and low-risk IRGs, respectively. Green line means down-regulation and red line means up-regulation.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionaltable1.xlsx](#)
- [Additionaltable2.xlsx](#)
- [Additionaltable3.xlsx](#)