

# Development a prognostic risk model based on B cell-related immune genes in Ovarian Cancer by integrative analysis of single-cell and bulk RNA sequencing data

**Jing Hu**

Huangshi Hubei Medical Group of maternal and Child Health Hospital

**Xuan Huang**

Department of Ophthalmology, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China

**Qihao Wei**

Sinopharm Genomics Technology Co., Ltd., Wuhan, China

**Xianjin Xiao**

Institute of Reproductive Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

**Yaping Lu** (✉ [luyaping@sinopharm.com](mailto:luyaping@sinopharm.com))

Sinopharm Genomics Technology Co., Ltd.

---

## Research

**Keywords:** scRNA-seq, B cell, ovarian cancer, Prognostic signature, immunotherapy

**Posted Date:** April 21st, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1522879/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background.** Ovarian cancer (OV) is the most serious malignancy of the female reproductive system and is generally diagnosed at an advanced stage with peritoneal metastasis. The aim of this research was to construct a tumor-infiltrating B (TIL-B) lymphocyte-associated immune-related gene profile for prognostic assessment of ovarian cancer patients.

**Methods.** The single cell RNA-seq data of metastatic ovarian cancer in this study were obtained by GEO data, and bulk RNA-seq data was obtained from TCGA database. Identification of immune-related marker genes in infiltrating B lymphocytes in tumor tissues of patients with metastatic ovarian cancer by scRNA-seq data analysis. Subsequently, based on bulk RNA-seq data and clinical follow-up data, univariate Cox analysis was used to identify the prognostic targets associated with tumor infiltrating B lymphocytes, and least absolute shrinkage and selection operator (LASSO) regression and multivariate Cox regression were used to construct prognostic risk models. The Kaplan-Meier survival curve and the ROC curve were used to test the prognostic performance of the model. EPIC, MCPcounter and ssGSEA software were used to predict the proportion of tumor-infiltrating lymphocytes in bulk RNA-seq expression profile samples. Multivariate Cox regression analysis was used to screen independent prognostic factors and construct linear plots to predict and assess 3 and 5-year survival of ovarian cancer patients.

**Results.** Single cell data from metastatic ovarian cancer tissue were clustered into 19 subgroups. After known cell type annotation, these cell subpopulations were annotated as six known cell types. The proportion of B cells was contrary to the clinical stage of the patient's tumor. Difference analysis identified 88 immune-related genes specifically expressed by B cells. Univariate Cox regression analysis, the LASSO regression analysis and multivariate Cox regression analysis were used to identify the independent prognostic factors associated with tumor invasion B cell immunity, *ISG20* and *SLAMF7*. Based on the risk model constructed by *ISG20* and *SLAMF7*, the AUC values of the 3-year and 5-year survival in the training set were 0.619 and 0.736, the AUC values of the test set were 0.694 and 0.758, and the AUC values of the external validation set were 0.6 and 0.61. The proportion of CD8+T cells, B cells, cytotoxic lymphocytes and aDC cells in the low-risk group was higher than in the high-risk group. The prognostic model has better independence and has good prognostic evaluation effect combined with clinical characteristics ( $p=0.013$ ). At the same time, we also construct a nomogram based on the prediction model.

**Conclusions.** Our study identified immune-related prognostic marker genes *ISG20* and *SLAMF7* in TIL-B cells by analyzing single-cell and bulk RNA-seq data of metastatic ovarian cancer, and established and verified the prognostic model as an independent prognostic indicator of ovarian cancer. It can provide potential therapeutic targets for immunotherapy and chemotherapy in patients with advanced cancer and predict patients' therapeutic response.

## Introduction

Ovarian cancer is the most lethal malignancy of the female reproductive system(Siegel et al., 2022). If ovarian cancer is detected and diagnosed early, surgery and chemotherapy can better treat the disease in clinical treatment. However, for most ovarian cancer patients, they are commonly diagnosed at advanced

stage (III/IV), with metastasis and high mortality(Jayson et al., 2014). Therefore, there is an urgent need to understand the underlying mechanism of ovarian cancer metastasis and its related biomarkers, so as to enable early diagnosis and timely treatment of ovarian cancer.

Single cell RNA sequencing is a new technology that can be used for transcriptome studies of individual cells, to investigate the heterogeneity of cell populations in tissues and to explore the cells involved in tumorigenesis and metastasis. The traditional bulk sequencing method averages the level of gene expression and does not reflect the proportion of cells in the tissue. Advanced tumor microenvironment is complex and highly heterogeneous, bulk RNA-seq cannot accurately response the components of the immune cells in the tumor microenvironment, and scRNA-seq relatively higher resolution and can be transcribed from a single cell level to explore the immune cells in the immune microenvironment.

The results of immune checkpoint inhibitor therapy in patients are related to NK cells, T cells and B cells in the tumor microenvironment, which makes the role of immune cells in the tumor immune microenvironment in tumor progression and treatment response process more important. At present, studies mainly focus on the role of T cell in anti-tumor immunity, but B cells are also the main component of infiltrating lymphocytes in tumor microenvironment, and there are few studies on their role in solid tumors(Paijens et al., 2021). Tumor-associated B cells are generally considered to promote cancer progression(Gu et al., 2019; Shalpour et al., 2015; Wei et al., 2016; Zhou et al., 2019). Nevertheless, in studies including multiple cancer species, high expression levels of B cell and plasma cell-related characteristic genes were associated with enhanced overall survival of patients(Iglesia et al., 2016). Given the dual role of TIL-B lymphocytes in cancer progression, we need to better understand the tumor microenvironment and the role of B cells to enhance therapy outcomes for patients with advanced ovarian cancer.

In this study, immune-related genes in TIL-B lymphocytes were analyzed at the single-cell level, and immunohistochemistry and clinical follow-up data were combined to explore immune-related prognostic markers of ovarian cancer, and a tumor infiltrating B cell-associated immune gene prognostic score model was developed to predict survival outcomes of ovarian cancer patients. The model can be used to distinguish potentially poor prognosis patients in the population and provide guidance for immunotherapy.

## **Materials & Methods**

### **Data sources**

Contains data of 6 cases of ovarian cancer metastasis cells scRNA -seq GSE147082 (Olalekan et al., 2021) through Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) to download. Cells isolated from the omentum of patients with metastatic ovarian cancer for single cell RNA sequencing. The clinical information of bulk RNA-seq data and samples from the The Cancer Genome Atlas Ovarian Cancer (TCGA-OV) cohort were downloaded from UCSC Xena (<https://xenabrowser.net/>) and randomly assigned to the training set and test set in a 7:3 ratio. GSE18520 from the GEO database is used as an external validation dataset. All analyses in this paper were carried out under R V4.1.2.

### **Single cell RNA-seq analysis**

Seurat (Butler et al., 2018) was used for scRNA-seq data processing. First, quality control was carried out on data and cells with nFeature less than 200 and percent\_mt greater than 20% were filtered out. Then use the "FindIntegrationAnchors" function in the Seurat package to integrate the data from six samples and remove batch effects. The integrated data was normalized by LogNormalize function, subsequently principal component analysis (PCA) was performed, followed by t-SNE algorithm for dimensionality reduction and visualization of the data. The cell clusters were annotated for known cell types using SingleR (Aran et al., 2019). The "FindAllMarkers" function is used to identify differentially expressed marker genes among different cell types. Genes with  $|\logFC| > 1$  and  $p < 0.05$  were defined as differentially expressed genes.

### **Gene function enrichment analysis**

Gene Ontology(GO) and the Kyoto Encyclopedia of Genes and Genomes Analyses(KEGG) pathway enrichment analysis was performed by clusterProfiler (Yu et al., 2012) package (V3.14.3).  $p < 0.05$  was considered statistically significant.

### **Construction of immune-related prognostic model of TIL-B cells**

The endpoint of this research was overall survival (OS), and univariate Cox proportional regression model was used to screen candidate genes associated with prognosis in TIL-B cell immune-related genes from the training set. The LASSO Cox regression model of glmnet (Friedman et al., 2010) was used to further screen the prognostic candidate genes to avoid model overfitting. Then, a stepwise selection method based on Akchi Information criterion (AIC) was used to select the inclusion variables (Attaar et al., 2017) to screen the appropriate TIL-B immune-related prognostic model. We calculated the risk score using the following formula: . Then, maxstat package method was used to define the optimal cutoff value, and the patients were classified into high risk group (greater than the optimal cutoff value) and low risk group (less than the optimal cutoff value).

### **Analysis of immune checkpoint and regulators of immunogenic cell death**

Immune checkpoints (ICs) and a regulator of immunogenic cell death (ICD) related genes were gathered from published articles (Huang et al., 2021). The Wilcox test was used to analyze the expression of related genes in the training set, test set and external validation set (\* :  $P < 0.05$ ; \*\* :  $p < 0.01$ ; \*\*\* :  $p < 0.001$ ; \*\*\*\* :  $p < 0.0001$ ).

### **Analysis of tumor infiltrating immune cells**

We used three software including EPIC, MCPcounter and ssGSEA to calculate the proportion of tumor infiltrating immune cells between different risk score subgroups, and then tested the differences between groups using Wilcox.test. (\* :  $p < 0.05$ ; \*\* :  $p < 0.01$ ; \*\*\* :  $p < 0.001$ ; \*\*\*\* :  $p < 0.0001$ )

### **Construction and validation of the nomogram**

Univariate Cox regression analysis was used to screen clinic variables associated with prognosis, and univariate Hazard Ratio (HR) and p value were calculated. Variables with  $p < 0.05$  were used to develop

nomogram with the rms R software package along with the prognostic risk score. The Calibration curves describes the degree to which the actual survival time fits the survival time forecast by the nomogram.

## Statistical analysis

R software V4.1.2 (<https://www.r-project.org/>) was used for statistical analysis and visualization in this study. The Wilcoxon test was used for comparison between the two groups. Survival and survminer were used for survival analysis.  $p < 0.05$  was considered statistically significant.

## Results

### Analysis of single-cell RNA-seq data

Ovarian cancer single cell sequencing data(GSE147082) from 6 samples containing a total of 9885 cells. According to strict quality control process, quality control and filtering of scRNA-seq data were carried out, in which cells with Feature\_RNA < 200 or percent.mt > 20% were filtered. After quality control, the number of detected genes was nFeature and count was evenly distributed in 6 samples (Figure S1A, B), and the proportion of ribosomes in cells was approximately 5% (Figure S1A, C). After filtration, 9583 high-quality cells were obtained, and subsequent analyses were based on these cells. According to the correlation analysis, it was found that the number of detected genes (nFeature) had a strong positive correlation with the sequencing depth (UMIs total number, nCount) (Figure S1D). We then normalized the data using the LogNormalize method. The normalized data were analyzed for variance analysis to identify highly variable genes, and the first 2000 highly variable genes were selected for the downstream analysis (Figure S1E).

We then apply the FindIntegrationAnchors function in the Seurat package to remove the batching effect and perform a t-SNE clustering analysis. Finally, 9583 cells from 6 samples were clustered into 16 cell clusters (Figure 1B). Visual clustering results showed that after removing the batch effect, the difference of sample origin was no longer the main difference among cell clusters (Figure 1A right). At this point, we need to know the cell type of each cell cluster, and use the R software package SingleR to annotate the known cell type of the cell cluster after clustering. The annotated cells mainly contain six known cell types: Monocytes(CD16<sup>-</sup>, CD14<sup>-</sup>) T cells (CD4<sup>+</sup> Th1) NK cells B cells(naive) T cells(CD4<sup>+</sup>, memory TREG) Monocytes(CD16<sup>+</sup>, CD14<sup>+</sup>) (Figure 1C). Marker genes (*CD14*, *FCGR3A*, *COL1A1*, *CTLA4*, *FOXP3*, *IL2RA*, *CD79A*, *MS4A1*, *CD96*) of known immune cells were used to verify cell types identified (Sakaguchi et al., 2020; Abplanalp et al., 2021). As shown in Figure 1D, marker genes of immune cells are highly expressed in corresponding cell clusters. In addition, FindAllMarkers function were used to identify the differentially expressed marker genes among cell clusters. 1000 cells were randomly selected from each cell cluster and the top 10 differentially expressed genes in each cell type were displayed with heatmap (Figure 1E). The patient samples were ranked by disease score (Olalekan et al., 2021), from low to high: PT-1 to PT-6, as shown in Table1 (Olalekan et al., 2021). The proportion of cells from different sample sources indicated that the proportion of tumor infiltrating B cells varied greatly among all samples. As shown in Figure 1F, the sample with the largest proportion of tumor infiltrating B cells was GSM4416535, whose ovarian cancer stage was ypT3a Nx M1/ lvb, and the lowest disease severity score was PT-1. And GSM4416537, which had the second largest proportion of tumor-infiltrating B cells, had a disease score of PT-2. The results showed that the proportion

of tumor infiltrating B lymphocytes in the tissues of 6 patients with ovarian cancer was opposite to the seriousness of the disease.

### **Marker gene enrichment analysis of B cells**

We used GO and KEGG enrichment analysis to study the enrichment of 140 marker genes in related pathways found in tumor infiltrating B cells of ovarian cancer samples, as shown in Figure 2A-C. The results showed that the biological process (BP) of marker genes in B cells were mainly enriched in: positive regulation of cell activation, positive regulation of lymphocyte activation, positive regulation of leukocyte Activation, B cell activation, antigen receptor-mediated signaling pathway, immune response - activating cell surface Receptor signaling Pathway, B cell receptor signaling Pathway. Cellular component (CC) was mainly enriched in: Immunoglobulin complex, external side of plasma membrane, immunoglobulin complex, Circulating, Blood Microparticle, MHC Class II Protein Complex. Molecular function (MF) are mainly enriched in: Antigen binding, immunoglobulin receptor binding, MHC class II protein complex binding, MHC protein complex binding, immune Receptor activity, peptide binding, peptide antigen binding, MHC class II receptor activity, immunoglobulin Binding. Figure 3D mainly shows the top 10 pathways in KEGG enrichment analysis results, which are as follows: Cell adhesion molecules, Hematopoietic Cell lineage, Hematopoietic Cell lineage, and Intestinal immune network for IgA Production, epicardial thyroid disease, Antigen processing and presentation and other pathways. These enrichment results suggested that the marker genes we screened were closely related to the function of B cells, further indicating that the genes we screened were reliable marker genes of B cells.

### **Construction of prognostic model of marker gene in tumor infiltrating B cells**

To further assure that the maker genes we screened for TIL-B cells were immune-related genes, the ImmuneSigDB subset (V7.5.1) from the immune-related dataset C7 was downloaded from the GSEA database(Liberzon et al., 2015). 88 TIL-B cell marker genes were obtained by intersection with marker genes found in TIL-B cells (Figure 3A), which could be considered as immune gene sets associated with TIL-B cells. To investigate which of these genes are associated with patient survival prognosis, we obtained a dataset of TCGA-OV cohort of ovarian cancer patients from the TCGA database (N=379), which was randomly assigned to a training set (N=265) and a test set (N=114) in a 7:3 ratio. Based on the TCGA training set, we adapted univariate Cox regression analysis of marker genes in 88 TIL-B cells to search for potential prognostic genes. The results showed that 10 of these genes were significantly correlated with the survival and prognosis of patients (Figure 3B): *IGHM* (HR = 0.93, P=0.0314), *CCR7* (HR=0.66, P=0.015), *ISG20* (HR=0.74, P=0.0098), *IGKC* (HR=0.95, P=0.0375), *IGHG1*(HR=0.95, P=0.0277), *IGJ* (HR=0.91, P=0.0106), *CD38* (HR=0.64, P=0.003), *SLAMF7* (HR=0.76, P=0.0053), *MZB1* (HR=0.86, P=0.0184) and *FKBP11*(HR=0.74, P=0.0441). LASSO Cox regression analysis was then used to further screen these genes and lambda.min, which was cross-validated 10 times, was selected as the optimal lambda (Figure 3C, D). The corresponding model contained six non-zero coefficients (Figure 3E). Patients with high expression of these genes had better survival prognosis than those with low expression, such as *IGL*, *SLAMF7*, *CCR7*, *FKBP11*, *CD38* and *ISG20* (Figure 3F). In order to construct a prognostic model for patients, all six prognostic protective genes were selected for multivariate Cox regression analysis to construct a prognostic model for TIL-B cells. In the

process of model construction, the AIC model was selected based on the model AIC value, and the optimal AIC model was selected through the stepwise backward algorithm. The optimal model ultimately contained genes *ISG20* and *SLAMF7* (Figure 3g). Based on these two genes, a prognostic score model for TIL-B was constructed: RiskScore =  $-0.23 \times \exp(\text{ISG20}) - 0.23 \times \exp(\text{SLAMF7})$ .

### **Validation of prognostic model performance**

The risk score of each sample in the training set, test set and external test set was calculated using the TIL-B immune-related prognosis model constructed by us, and the patients were split into high and low risk groups according to the best cutoff value of risk score. Heatmap was used to demonstrate the expression of relevant genes in TIL-B immune-related prognosis model in patients, and it was found that the expression values of *ISG20* and *SLAMF7* were contrary to the trend of risk score (training set -Figure 4A, test set - Figure 5A, external test set - Figure 6A). Subsequently, the Wilcox test was used to test the difference between the expression of *ISG20* and *SLAMF7* in the high-risk and low-risk groups. As shown in the box diagram in Figure S2, *ISG20* and *SLAMF7* were up-regulated in the low-risk group in all data sets, and their expression levels were negatively correlated with risk score. Scatter plots were used to show the risk score, survival time and survival status of patients in each data set (training set-Figure 4 B and C, test set-Figure 5 B and C, external test set-Figure 6 B and C). In order to verify whether the risk score constructed by us is associated with the prognosis of patients, we used Log Rank to test the K-M survival curve of patients in the high and low score groups, and the results showed that patients in the high risk group had a worse prognosis than those in the low risk group (Log Rank test of training set  $p < 0.0001$ , HR=2.72, As shown in figure 4D; Log Rank test  $p = 0.0038$ , HR=2.31, as shown in Figure 5D; External test set log Rank tests  $P = 0.022$ , HR=1.34, as shown in Figure 6D). The 3-year and 5-year survival were predicted by the model. ROC curves (AUC) were 0.619 (3 years) and 0.736 (5 years) in the training set and 0.694 (3 years) and 0.758 (5 years) in the test set (Figure 5E). ROC curves (AUC) in the external test set were 0.6 (3 years) and 0.61 (5 years) respectively (Figure 6E). These results indicate that the model constructed based on TIL-B immune-related genes has a good prognostic value for survival.

### **Assessment of tumor immune cell infiltration**

Tumor infiltrating immune cells play an important role in tumor progression and clinical treatment response. EPIC, MCPcounter and ssGSEA were used to evaluate the tumor infiltrating immune cells using bulk RNA sequencing data. It is generally believed that patients with higher proportion of CD8+T cells, B cells, Cytotoxic lymphocytes, NK cells and aDC cells in tumor microenvironment have better clinical prognosis (Becht et al., 2016b; Fridman et al., 2012; Becht et al., 2016a). In our results, the proportion of CD8+T cells, B cells, Cytotoxic lymphocytes, and aDC cells in the low-risk group was higher than that in the high-risk group (Figure 7).

### **Immune checkpoints and regulators of immunogenic cell death**

Patients with ovarian cancer are commonly diagnosed at advanced stage or with metastatic disease, and the usual surgical interventions have failed to react. A combination of radiotherapy, chemotherapy, targeted therapy, and immune checkpoint inhibitors is used to treat cancer, and even therapeutic tumor vaccines are

used. Immune checkpoints (ICs) and immunogenic cell death (ICD)(Krysko et al., 2012; Huang et al., 2021) are important in cancer immunotherapy strategies and chemotherapy(Gebremeskel and Johnston, 2015), and we studied their expression levels in patients with different risk scores. It can be seen that in the TCGA-OV cohort data set, ICs expression levels in different risk groups are significantly different. Overall, ICs expression levels in the low-risk group are higher than those in the high-risk group (Figure 8A). The expression of ICs related genes *CD86*, *LAG3*, *LAIR1* and *CD44* in the TCGA-OV dataset and the external GEO dataset were significantly different in the high and low risk groups, and the expression was up-regulated in the low risk group (Figure 8A, B). *CD86*, as a ligand of Cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), can interact with it to inhibit T cell activation (Dyck and Mills, 2017). *LAG3* is an important ICs with important implications for cancer, infectious diseases, and autoimmunity. The binding of leukocyte associated immunoglobulin-like receptor 1 (AIR-1) to its ligand results in the loss of immune function in the tumor microenvironment (TME), and reduced T cell function and immune response of antigen presenting cells (Xu et al., 2020). Patients in the high-risk group all had low ICs expression levels, suggesting that ICs inhibitors may be less effective in the high-risk group. As for ICD-related gene sets, most of them were also highly expressed in the low-risk group, and *CXCL10* and *TLR4* showed statistically significant differences in TCGA-OV and GEO data sets (Figure 8C, D), suggesting that patients in the low-risk group might have better chemotherapy effects. Consequently, the risk score can reflect the expression level of ICs and ICD regulatory factors, and can be used for drug guidance in tumor immunotherapy and chemotherapy.

### Clinical nomogram construction

Considering individual differences of clinical patients, clinically related characteristics of patients were included (Table 2), and the relationship between clinically related characteristics and risk score and prognosis of patients was studied by univariate Cox regression and multivariate Cox regression analysis. Univariate Cox regression analysis showed that risk score and clinical stage were correlated with the prognosis of patients (Figure 9A). Multivariate Cox regression analysis showed that, after adjusting other factors, there was no significant correlation between clinical stage and prognosis ( $P>0.05$ ), while risk score (HR=3.151,  $P=0.013$ ) was significantly correlated with prognosis (Figure 9B). This indicated that the risk score was an independent prognostic factor. Subsequently, we established a prognostic linear plot based on risk score to quantitatively estimate the 3,5-year survival probability of ovarian cancer patients (Figure 9C). The Calibration curves shows that TCGA-OV cohort data (Figure 9D) and external validation data GSE18520 (Figure E) are in good agreement with the optimal prediction probability (gray dotted line in figure) at 3 and 5 years. These results suggest that this comprehensive nomogram can be used as a reliable tool for predicting OS in ovarian cancer patients.

## Discussion

Each year, about 313,959 women worldwide develop ovarian cancer, and about 207,252 women die from ovarian cancer, severely endangering women's health(Sung et al., 2021). Although ovarian cancer is treated with advanced surgery and chemotherapy, patients with advanced metastatic ovarian cancer have a poor prognosis. Due to the heterogeneity of advanced tumor tissue microenvironment, there is currently a lack of effective biomarkers, and unable to predict their treatment prognosis(Hoppenot et al., 2018). In recent years,

tumor infiltrating cells have attracted considerable attention due to the clinical success of checkpoint blockers (ICBs) (Paijens et al., 2021). Tumor infiltrating B lymphocytes has been reported to be associated with the prognosis of ovarian cancer. In this study, six types of tumor-infiltrating lymphocytes were identified by analyzing single-cell RNA-seq data from patients with metastatic ovarian cancer. The higher the proportion of tumor infiltrating B lymphocytes, the milder the clinical staging of the patient's tissue. Studies have indicated that high expression of B-cell signature genes is associated with enhanced overall survival in patients with melanoma, lung adenocarcinoma, breast cancer and head and neck squamous cell carcinoma (Iglesia et al., 2016). By comparing the differentially expressed genes of TIL-B lymphocytes with other immune cells, we identified 140 marker genes that specifically express TIL-B lymphocytes. Eighty-eight marker genes related to immunity were screened. Univariate Cox, Lasso Cox and multivariate Cox regression analyses were used to identify two prognostic factors associated with B lymphocyte immunity in ovarian cancer. The results showed that the risk score calculated based on these two genes was significantly associated with a poor prognosis in ovarian cancer patients. We found that *ISG20* and *SLAMF7* genes were expressed higher in patients with low risk score than in patients with high risk score. The *ISG20* gene is located in the cytoplasm and nuclear lumen and activates 3', 5' exonuclease activity and RNA binding activity. Involved in defense responses to viruses, negative regulation of viral genome replication, and catabolic processes containing base compounds(2020; Van Tong et al., 2018). *SLAMF7* gene is located in the endoplasmic reticulum, participates in adaptive immune response, activates NK cells (2020), and regulates T cell function in tumor microenvironment(O'Connell et al., 2021). Considering that the outcome of immune checkpoint inhibitor therapy in cancer patients is related to T cells, NK cells and B cells in the tumor microenvironment(Paijens et al., 2021). We used EPIC, MCPcounter and ssGSEA to evaluate tumor infiltrating immune cells, and the results showed that the proportion of tumor infiltrating B lymphocytes, CD8+T cells, cytotoxic lymphocytes and aDC cells in the low-risk group was higher than that in the high-risk group. Risk score was an independent prognostic factor in multivariate Cox regression analysis after the inclusion of complex clinical factors. The results of K-M survival curve analysis showed that patients with low risk score had better clinical prognosis. The nomogram was developed to predict the clinical prognosis of ovarian cancer patients. We explored the differences in the expression levels of immune checkpoints and immunogenic cell death regulators in patients with different risk scores, and found that the expression levels of immune checkpoints were up-regulated based on *CD86*, *LAG3*, *LAIR1*, and *CD44* in patients with ovarian cancer with low risk scores. *LAG3* is more effective in treating melanoma, especially in other refractory immune checkpoint blockers(Ascierto et al., 2017); Studies have shown that by blocking the binding of *LAIR1* to its ligand, cytotoxic T cell infiltration and anti-tumor immune response can be enhanced to eliminate cancer cells (Xiao et al., 2016). In colorectal cancer, *CD44/OPN* acts as an immune checkpoint, inhibiting T cell activation and leading to tumor immunotherapy resistance (Klement et al., 2018). Therefore, we can develop new immunotherapy targets based on these immune checkpoints. The expression of immunogenic cell death regulators *CXCL10* and *TLR4* was significantly upregulated in patients with low risk score. We developed a new nomogram using a prognostic model, based on which we can better predict the clinical outcome of ovarian cancer patients.

## Conclusions

Overall, our study identified immune-related prognostic marker genes *ISG20* and *SLAMF7* in TIL-B cells by analyzing single-cell and bulk RNA-seq data of metastatic ovarian cancer, and established and validated a prognostic model as an independent prognostic indicator for ovarian cancer. Provide potential therapeutic targets for immunotherapy and chemotherapy in patients with advanced cancer and to predict patient response to treatment.

## Abbreviations

**OV:** Ovarian cancer

**GEO:** Gene Expression Omnibus

**PCA:** principal component analysis

**t-SNE:** t-Distributed Stochastic Neighbor Embedding

**GO:** Gene Ontology

**KEGG:** Kyoto Encyclopedia of Genes and Genomes Analyses

**LASSO:** least absolute shrinkage operator

**TIL-B:** tumor-infiltrating B

**BP:** biological process

**MF:** molecular function

**CC:** cellular component

**OS:** overall survival

**HR:** hazard ratio

**ICs:** Immune checkpoints

**ICD:** immunogenic cell death

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

## **Availability of supporting data**

Not applicable.

## **Competing interests**

No potential conflict of interest was reported by the authors.

## **Funding**

This work was supported by grants from the National Natural Science Foundation of China to Xuan Huang [grant number: 32000485]

## **Authors' contributions**

Jing Hu, Xuan Huang and Qihao Wei contributed equally to this work.

JH, WQ performed the statistical analyses and drafted the manuscript; JH, WQ and HX edited the manuscript; LY, XX guided and revised the manuscript.

## **Acknowledgements**

Throughout the writing of this dissertation I have received a great deal of support and assistance. I would also like to thank my tutors, for their valuable guidance throughout my studies. In addition, I could not have completed this dissertation without the support of my friends, who provided stimulating discussions as well as happy distractions to rest my mind outside of my research.

## **Authors' information**

Affiliations

**Huangshi Hubei Medical Group of maternal and Child Health Hospital, Hubei, China**

Jing Hu

**Department of Ophthalmology, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China**

**Medical Research Center, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China**

Xuan Huang

**Sinopharm Genomics Technology Co., Ltd., Wuhan 430030, China**

Yaping Lu, Qihao Wei

**Institute of Reproductive Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China**

## References

1. 2020. Alliance of Genome Resources Portal: unified model organism research platform. *Nucleic Acids Res*, 48, D650-d658. <http://dx.doi.org/10.1093/nar/gkz813>.
2. Abplanalp, W. T., John, D., Cremer, S., Assmus, B., Dorsheimer, L., Hoffmann, J., et al. 2021. Single-cell RNA-sequencing reveals profound changes in circulating immune cells in patients with heart failure. *Cardiovasc Res*, 117, 484-494. <http://dx.doi.org/10.1093/cvr/cvaa101>.
3. Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., et al. 2019. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*, 20, 163-172. <http://dx.doi.org/10.1038/s41590-018-0276-y>.
4. Ascierto, P. A., Melero, I., Bhatia, S., Bono, P., Sanborn, R. E., Lipson, E. J., et al. 2017. Initial efficacy of anti-lymphocyte activation gene-3 (anti-LAG-3; BMS-986016) in combination with nivolumab (nivo) in pts with melanoma (MEL) previously treated with anti-PD-1/PD-L1 therapy. *Journal of Clinical Oncology*, 35, 9520-9520. [http://dx.doi.org/10.1200/JCO.2017.35.15\\_suppl.9520](http://dx.doi.org/10.1200/JCO.2017.35.15_suppl.9520).
5. Attaar, A., Winger, D. G., Luketich, J. D., Schuchert, M. J., Sarkaria, I. S., Christie, N. A., et al. 2017. A clinical prediction model for prolonged air leak after pulmonary resection. *J Thorac Cardiovasc Surg*, 153, 690-699.e2. <http://dx.doi.org/10.1016/j.jtcvs.2016.10.003>.
6. Becht, E., Giraldo, N. A., Dieu-Nosjean, M. C., Sautès-Fridman, C. and Fridman, W. H. 2016a. Cancer immune contexture and immunotherapy. *Curr Opin Immunol*, 39, 7-13. <http://dx.doi.org/10.1016/j.coi.2015.11.009>.
7. Becht, E., Giraldo, N. A., Germain, C., de Reyniès, A., Laurent-Puig, P., Zucman-Rossi, J., et al. 2016b. Immune Contexture, Immunoscore, and Malignant Cell Molecular Subgroups for Prognostic and Theranostic Classifications of Cancers. *Adv Immunol*, 130, 95-190. <http://dx.doi.org/10.1016/bs.ai.2015.12.002>.
8. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*, 36, 411-420. <http://dx.doi.org/10.1038/nbt.4096>.
9. Dyck, L. and Mills, K. H. G. 2017. Immune checkpoints and their inhibition in cancer and infectious diseases. *Eur J Immunol*, 47, 765-779. <http://dx.doi.org/10.1002/eji.201646875>.
10. Fridman, W. H., Pagès, F., Sautès-Fridman, C. and Galon, J. 2012. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer*, 12, 298-306. <http://dx.doi.org/10.1038/nrc3245>.
11. Friedman, J. H., Hastie, T. and Tibshirani, R. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33, 1 - 22. <http://dx.doi.org/10.18637/jss.v033.i01>.
12. Gebremeskel, S. and Johnston, B. 2015. Concepts and mechanisms underlying chemotherapy induced immunogenic cell death: impact on clinical studies and considerations for combined therapies. *Oncotarget*, 6, 41600-19. <http://dx.doi.org/10.18632/oncotarget.6113>.

13. Gu, Y., Liu, Y., Fu, L., Zhai, L., Zhu, J., Han, Y., et al. 2019. Tumor-educated B cells selectively promote breast cancer lymph node metastasis by HSPA4-targeting IgG. *Nat Med*, 25, 312-322. <http://dx.doi.org/10.1038/s41591-018-0309-y>.
14. Hoppenot, C., Eckert, M. A., Tienda, S. M. and Lengyel, E. 2018. Who are the long-term survivors of high grade serous ovarian cancer? *Gynecol Oncol*, 148, 204-212. <http://dx.doi.org/10.1016/j.ygyno.2017.10.032>.
15. Huang, X., Zhang, G., Tang, T. and Liang, T. 2021. Identification of tumor antigens and immune subtypes of pancreatic adenocarcinoma for mRNA vaccine development. *Mol Cancer*, 20, 44. <http://dx.doi.org/10.1186/s12943-021-01310-0>.
16. Iglesia, M. D., Parker, J. S., Hoadley, K. A., Serody, J. S., Perou, C. M. and Vincent, B. G. 2016. Genomic Analysis of Immune Cell Infiltrates Across 11 Tumor Types. *J Natl Cancer Inst*, 108. <http://dx.doi.org/10.1093/jnci/djw144>.
17. Jayson, G. C., Kohn, E. C., Kitchener, H. C. and Ledermann, J. A. 2014. Ovarian cancer. *Lancet*, 384, 1376-88. [http://dx.doi.org/10.1016/s0140-6736\(13\)62146-7](http://dx.doi.org/10.1016/s0140-6736(13)62146-7).
18. Klement, J. D., Paschall, A. V., Redd, P. S., Ibrahim, M. L., Lu, C., Yang, D., et al. 2018. An osteopontin/CD44 immune checkpoint controls CD8+ T cell activation and tumor immune evasion. *J Clin Invest*, 128, 5549-5560. <http://dx.doi.org/10.1172/jci123360>.
19. Krysko, D. V., Garg, A. D., Kaczmarek, A., Krysko, O., Agostinis, P. and Vandenabeele, P. 2012. Immunogenic cell death and DAMPs in cancer therapy. *Nat Rev Cancer*, 12, 860-75. <http://dx.doi.org/10.1038/nrc3380>.
20. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P. and Tamayo, P. 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems*, 1, 417-425. <http://dx.doi.org/10.1016/j.cels.2015.12.004>.
21. O'Connell, P., Hyslop, S., Blake, M. K., Godbehare, S., Amalfitano, A. and Aldhamen, Y. A. 2021. SLAMF7 Signaling Reprograms T Cells toward Exhaustion in the Tumor Microenvironment. *J Immunol*, 206, 193-205. <http://dx.doi.org/10.4049/jimmunol.2000300>.
22. Olalekan, S., Xie, B., Back, R., Eckart, H. and Basu, A. 2021. Characterizing the tumor microenvironment of metastatic ovarian cancer by single-cell transcriptomics. *Cell Rep*, 35, 109165. <http://dx.doi.org/10.1016/j.celrep.2021.109165>.
23. Paijens, S. T., Vledder, A., de Bruyn, M. and Nijman, H. W. 2021. Tumor-infiltrating lymphocytes in the immunotherapy era. *Cell Mol Immunol*, 18, 842-859. <http://dx.doi.org/10.1038/s41423-020-00565-9>.
24. Sakaguchi, S., Mikami, N., Wing, J. B., Tanaka, A., Ichiyama, K. and Ohkura, N. 2020. Regulatory T Cells and Human Disease. *Annu Rev Immunol*, 38, 541-566. <http://dx.doi.org/10.1146/annurev-immunol-042718-041717>.
25. Shalapour, S., Font-Burgada, J., Di Caro, G., Zhong, Z., Sanchez-Lopez, E., Dhar, D., et al. 2015. Immunosuppressive plasma cells impede T-cell-dependent immunogenic chemotherapy. *Nature*, 521, 94-8. <http://dx.doi.org/10.1038/nature14395>.
26. Siegel, R. L., Miller, K. D., Fuchs, H. E. and Jemal, A. 2022. Cancer statistics, 2022. 72, 7-33. <http://dx.doi.org/https://doi.org/10.3322/caac.21708>.

27. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*, 71, 209-249. <http://dx.doi.org/10.3322/caac.21660>.
28. Van Tong, H., Hoan, N. X., Binh, M. T., Quyen, D. T., Meyer, C. G., Song, L. H., et al. 2018. Interferon-stimulated gene 20 kDa protein serum levels and clinical outcome of hepatitis B virus-related liver diseases. *Oncotarget*, 9, 27858-27871. <http://dx.doi.org/10.18632/oncotarget.25559>.
29. Wei, X., Jin, Y., Tian, Y., Zhang, H., Wu, J., Lu, W., et al. 2016. Regulatory B cells contribute to the impaired antitumor immunity in ovarian cancer patients. *Tumour Biol*, 37, 6581-8. <http://dx.doi.org/10.1007/s13277-015-4538-0>.
30. Xiao, X., Lao, X. M., Chen, M. M., Liu, R. X., Wei, Y., Ouyang, F. Z., et al. 2016. PD-1hi Identifies a Novel Regulatory B-cell Population in Human Hepatoma That Promotes Disease Progression. *Cancer Discov*, 6, 546-59. <http://dx.doi.org/10.1158/2159-8290.Cd-15-1408>.
31. Xu, L., Wang, S., Li, J., Li, J. and Li, B. 2020. Cancer immunotherapy based on blocking immune suppression mediated by an immune modulator LAIR-1. *Oncoimmunology*, 9, 1740477. <http://dx.doi.org/10.1080/2162402x.2020.1740477>.
32. Yu, G., Wang, L. G., Han, Y. and He, Q. Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*, 16, 284-7. <http://dx.doi.org/10.1089/omi.2011.0118>.
33. Zhou, M., Wang, H., Zeng, X., Yin, P., Zhu, J., Chen, W., et al. 2019. Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*, 394, 1145-1158. [http://dx.doi.org/10.1016/s0140-6736\(19\)30427-1](http://dx.doi.org/10.1016/s0140-6736(19)30427-1).

## Tables

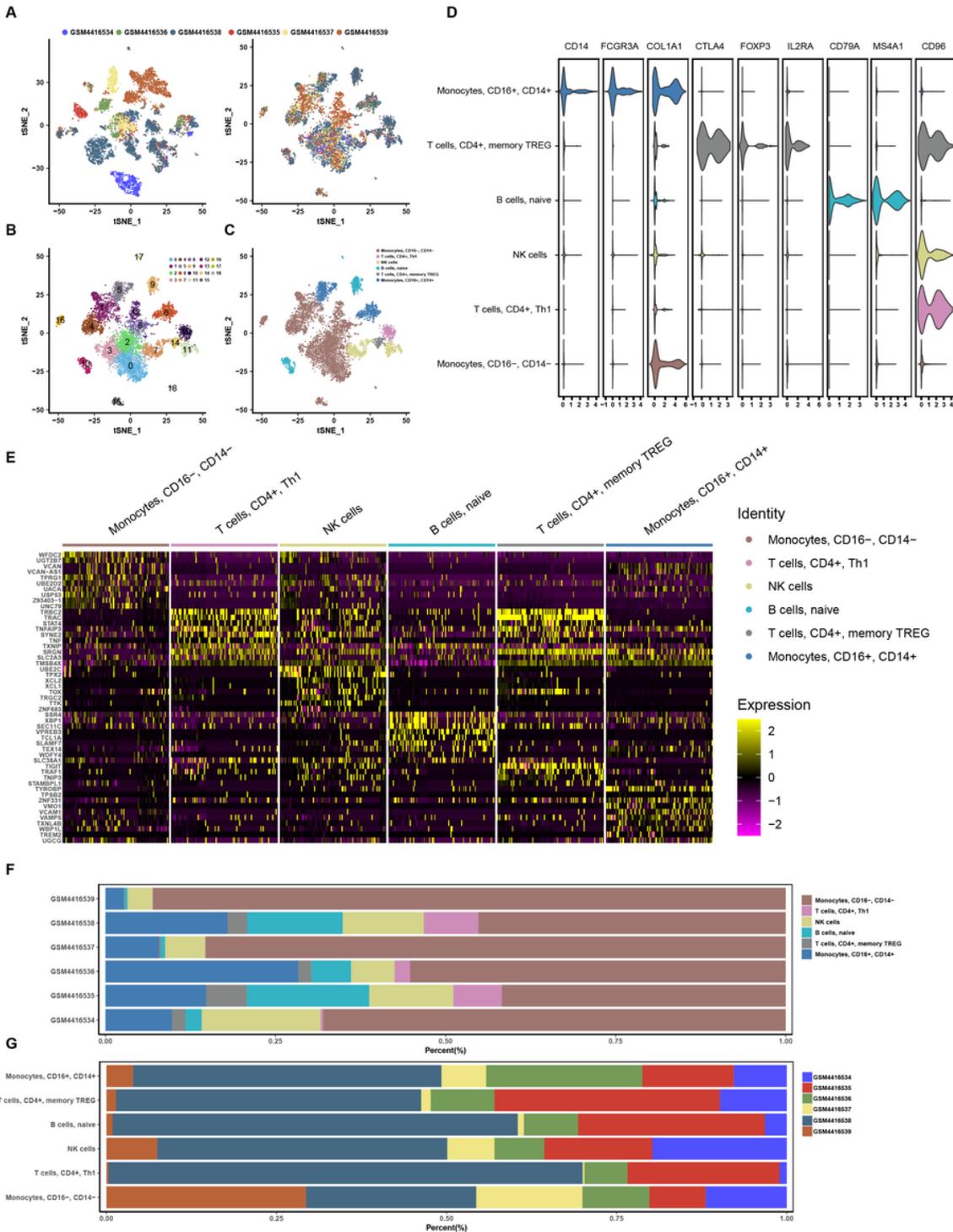
**Table 1 Metadata of metastatic omental tumors in 6 patients with ovarian cancer**

Disease score	Patient	Age	Race	Origin	Histologic grade	Neoadjuvant therapy	Stage (PMN/FIGO)
PT-1	GSM4416535	62	White	undetermined	Serous	Yes	ypT3a Nx M1/ Ivb
PT-2	GSM4416538	56	White	Left ovary	High grade serous carcinoma	No	pT3c Nx Mx/ IIIc
PT-3	GSM4416536	66	Black	Left fallopian (STIC)	High grade serous carcinoma	Yes	ypT3c N1a / IIIc
PT-4	GSM4416534	46	Asian	Left fallopian (STIC)	High grade serous carcinoma	No	pT3c Nx / IIIc
PT-5	GSM4416537	71	Black	Left fallopian (STIC)	High grade serous carcinoma	Yes	pT3c, N1, M1/ IIIc
PT-6	GSM4416539	66	Asian	Fallopian	Malignant mixed Mullerian tumor	Yes	ypT3c Nx/ IIIc

**Table 2 Clinical information statistics of patients in TCGA-OV cohort**

	Overall
<b>N</b>	378
<b>status = Live (%)</b>	146 (38.6)
<b>time (mean (SD))</b>	1194.62 (946.94)
<b>clinical_stage (%)</b>	
Stage II	23 (6.1)
Stage III	294 (78.0)
Stage IV	57 (15.1)
Not report	3 (0.8)
<b>neoplasm_histologic_grade (%)</b>	
G1	1 (0.3)
G2	45 (11.9)
G3	321 (84.9)
G4	1 (0.3)
GB	2 (0.5)
GX	6 (1.6)
Not report	2 (0.5)
<b>lymphatic_invasion (%)</b>	
YES	101 (26.7)
NO	48 (12.7)
Not report	229 (60.6)
<b>age (mean (SD))</b>	59.55 (11.42)

## Figures



**Figure 1**

single cell RNA sequencing analysis results of 6 cases of metastatic ovarian cancer

A. As shown in the t-SNE plot, each color represents a sample, where (figure left) is before batch effect removal, (figure right) is after batch effect removal; B. As shown in the t-SNE plot after removing the batch effect, the t-SNE clustering according to the cluster of cells, where each color represents a cell cluster; C. After the batch effect is removed, t-SNE clusters are grouped by annotated known cell types, where each

color represents a cell type; D. Violin diagram showing the expression of marker genes of known immune cells in various cell types; E. Heatmap to display the expression of the top 10 marker genes among the six known cell types identified; F. Distribution of immune cells in each patient's tissues; G. Proportion of immune cells in each patient's tissue.

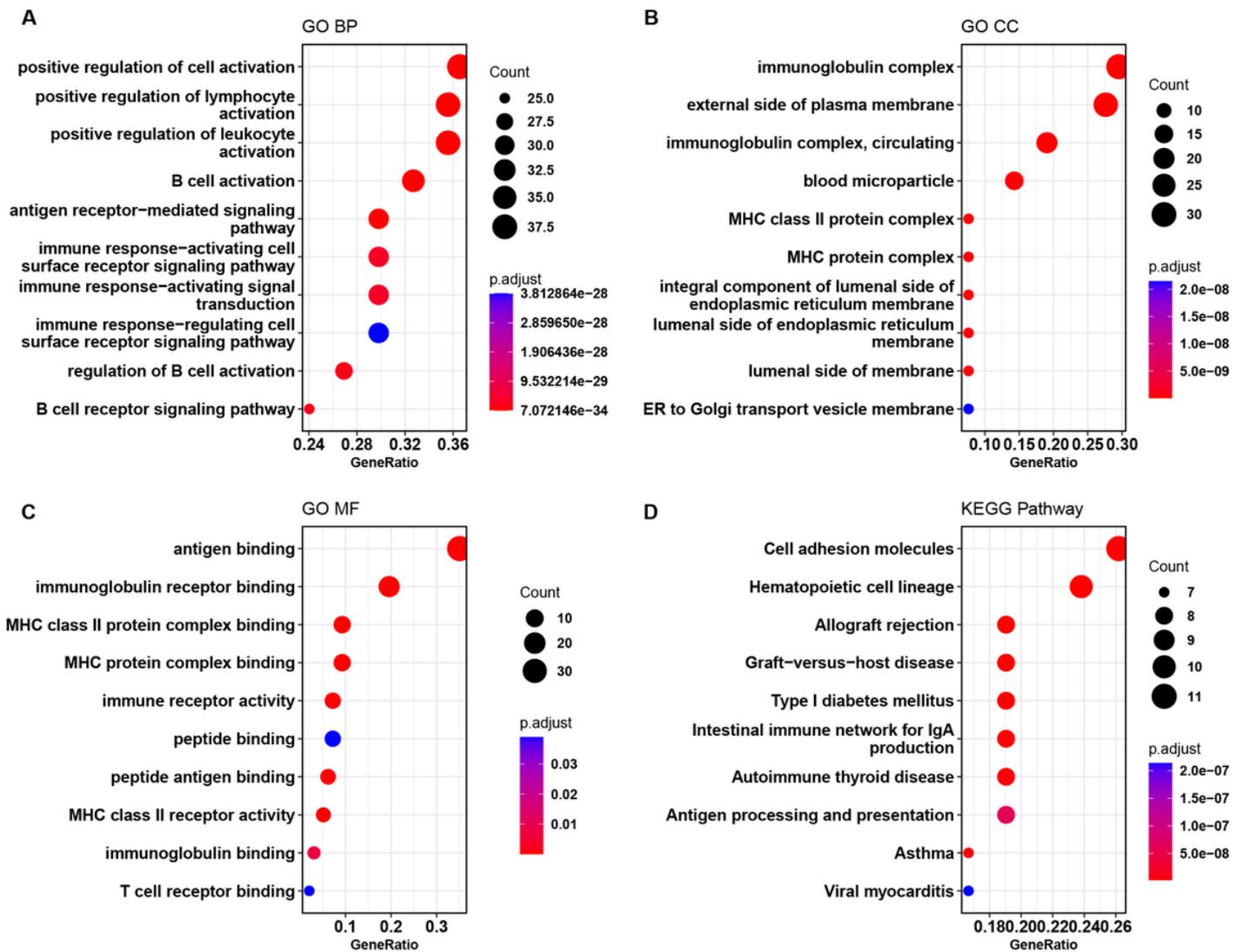
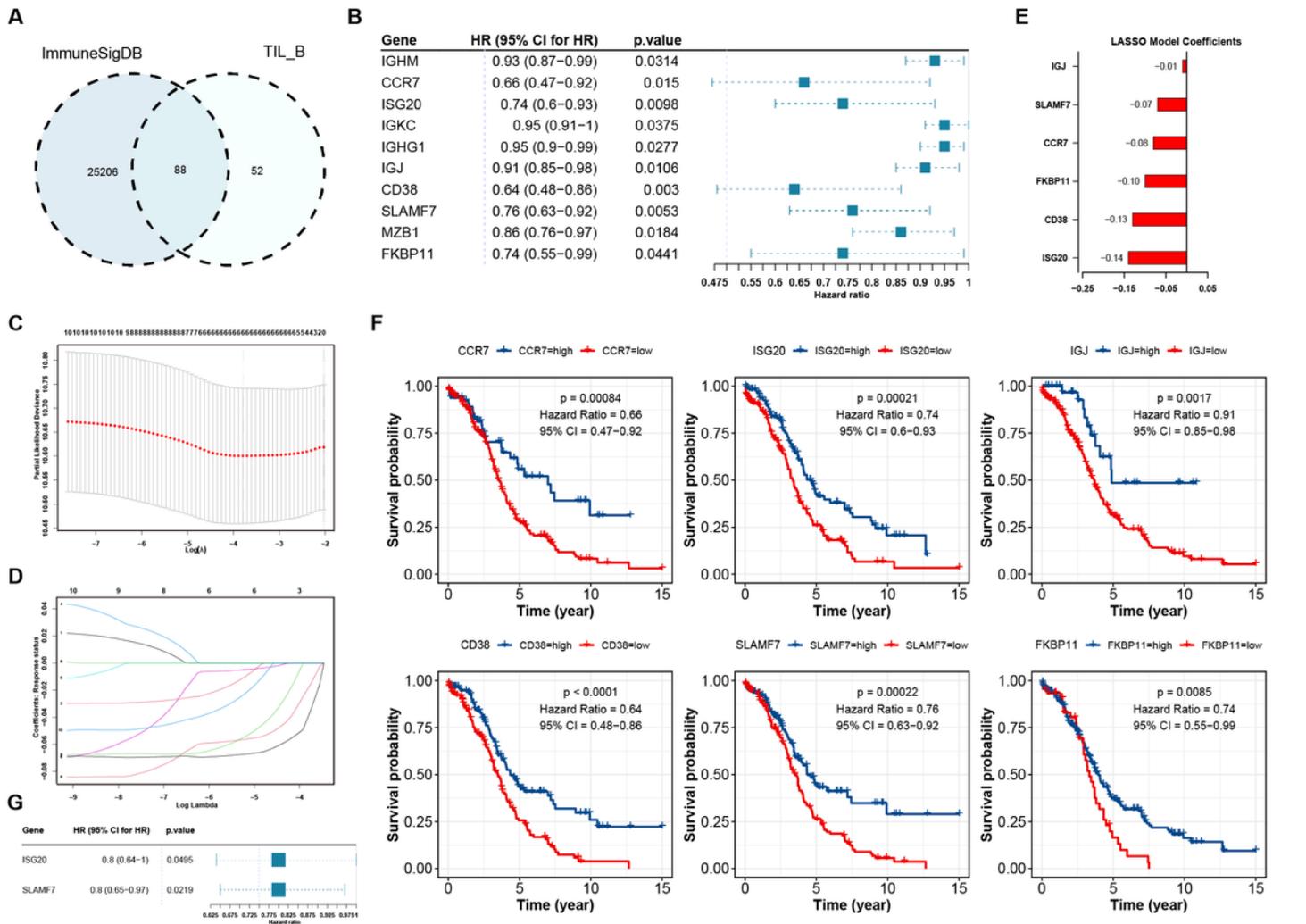


Figure 2

### GO and KEGG pathway enrichment analysis

A. Bubble map of BP (top 10 items) GO enrichment analysis; B. Bubble map of CC (top 10 items) GO enrichment analysis; C. Molecular function MF (top 10 items) GO enrichment analysis bubble map display; D. KEGG pathway (top 10 items) enrichment bubble map.



**Figure 3**

### TIL-B cell immune-related prognostic model construction

A. Venn diagram showing the intersection of ImmuneSigDB immune-related gene dataset and TIL-B cell marker in MSigDB; B. Forestplot showing hazard ratio (HR) and P values of 10 genes associated with prognosis of ovarian cancer patients by univariate Cox regression analysis; C. 10-fold cross-validation to screen the best parameters of lambda; D. LASSO model coefficients are determined by the optimal lambda; E. Cox regression model coefficients of six prognostic genes screened by univariate Cox regression model; F. The Kaplan-Meier survival curve of CCR7, ISG20, IGJ, CD38, SLAMF7 and FKBP11 were divided into high expression group and low expression group; G. Forestplot shows the HR and p values of independent prognostic genes ISG20 and SLAMF7 screened by multivariate Cox regression model.

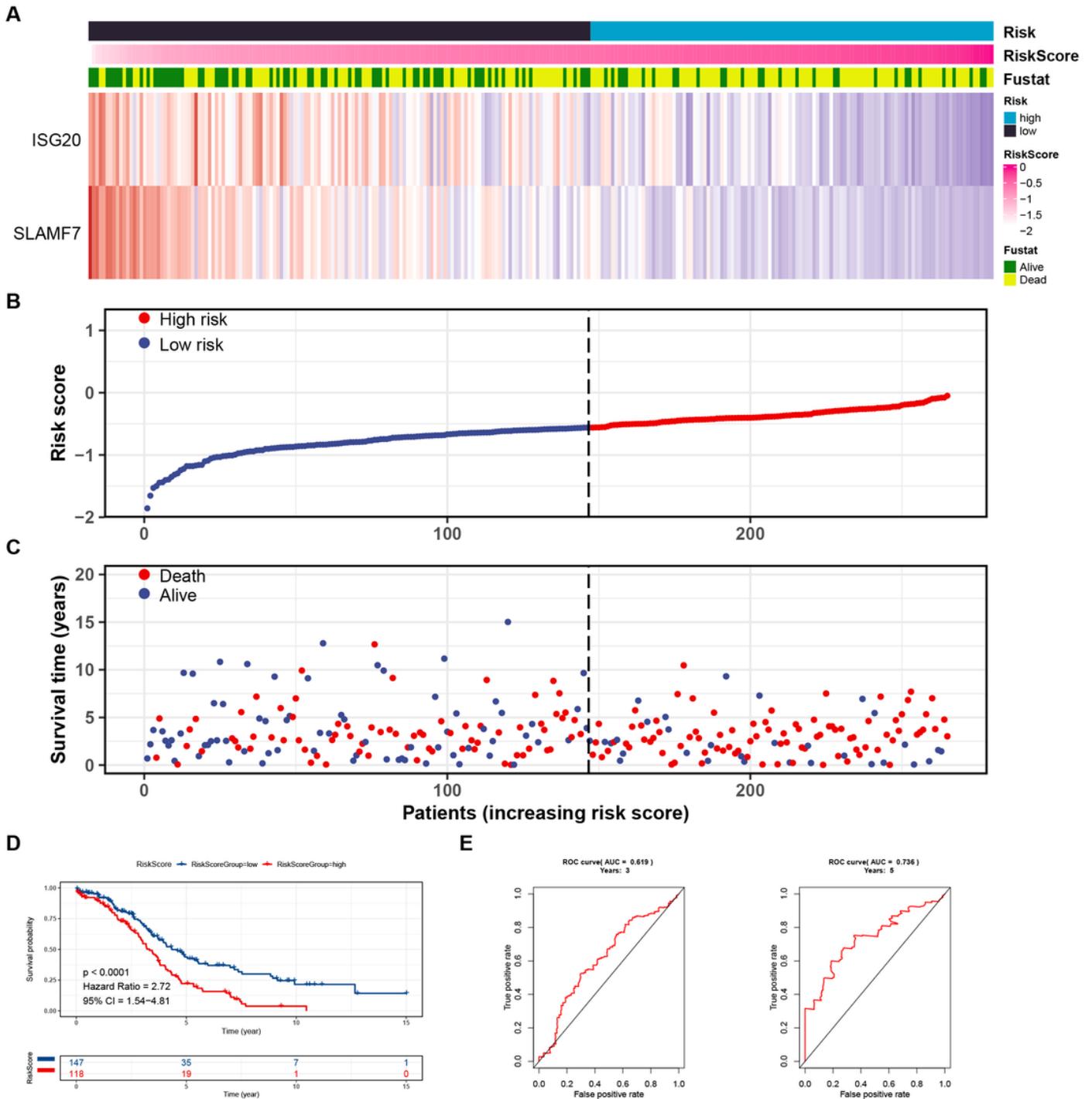


Figure 4

### Performance of the prognostic model verified by training set

A. Heatmap to show the expression of ISG20 and SLAMF7 genes in the training set, the survival status of patients and the grouping information of risk score; B. Scatter plot of training set sample risk score distribution; C. Scatter plot shows the distribution of survival time and survival state of patients in the training set; D. The Kaplan-Meier survival curve showed that the survival time of patients in the high and low

risk groups of the training set was significantly different; E. The ROC curve shows the ROC of the model to predict the 3- and 5-year survival rate.

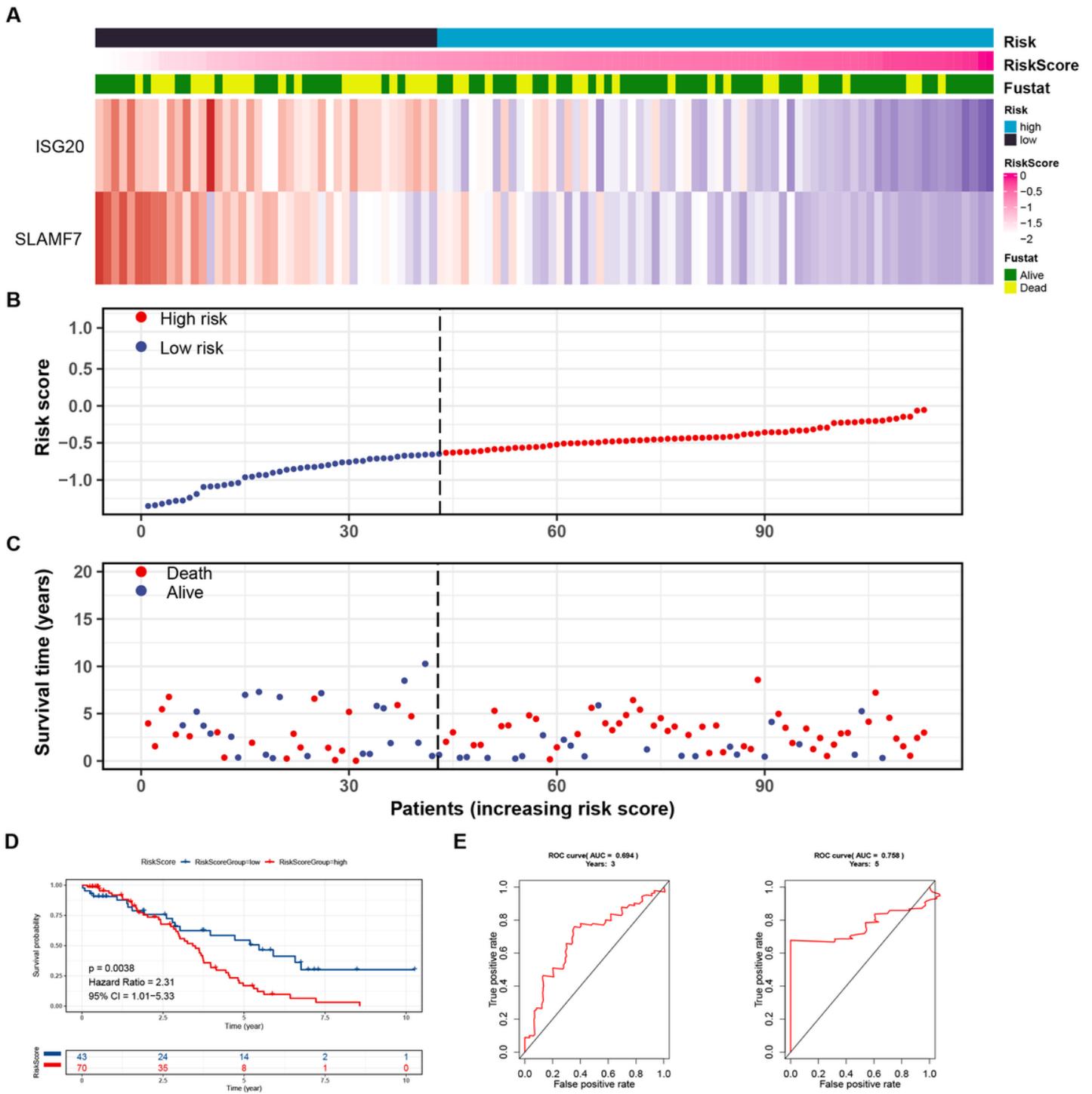


Figure 5

Test set validation of prognostic model performance

A. Heatmap to show the expression of ISG20 and SLAMF7 genes in the test set, the survival status of patients and the grouping information of risk score; B. Scatter plot of test set sample risk score distribution;

C. Scatter plot shows the distribution of survival time and survival state of patients in the test set; D. The Kaplan-Meier survival curve showed that the survival time of patients in the high and low risk groups of the test set was significantly different; E. The ROC curve shows the ROC of the model to predict the 3- and 5-year survival rate.

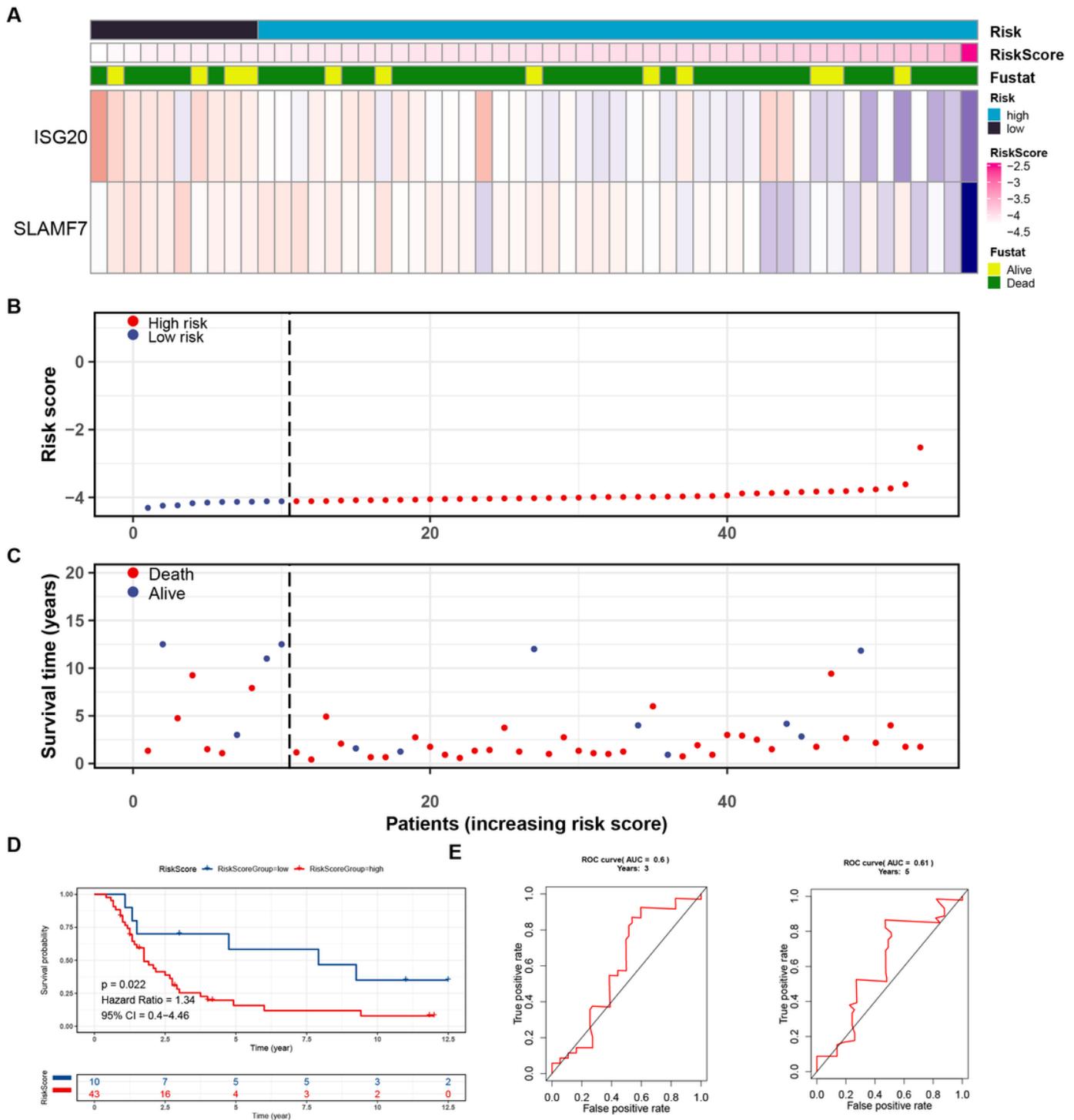


Figure 6

External validation set validation of prognostic model performance

A. Heatmap to show the expression of ISG20 and SLAMF7 genes in the external validation set, the survival status of patients and the grouping information of risk score; B. Scatter plot of external validation set sample risk score distribution; C. Scatter plot shows the distribution of survival time and survival state of patients in the external validation set; D. The Kaplan-Meier survival curve showed that the survival time of patients in the high and low risk groups of the external validation set was significantly different; E. The ROC curve shows the ROC of the model to predict the 3- and 5-year survival rate.

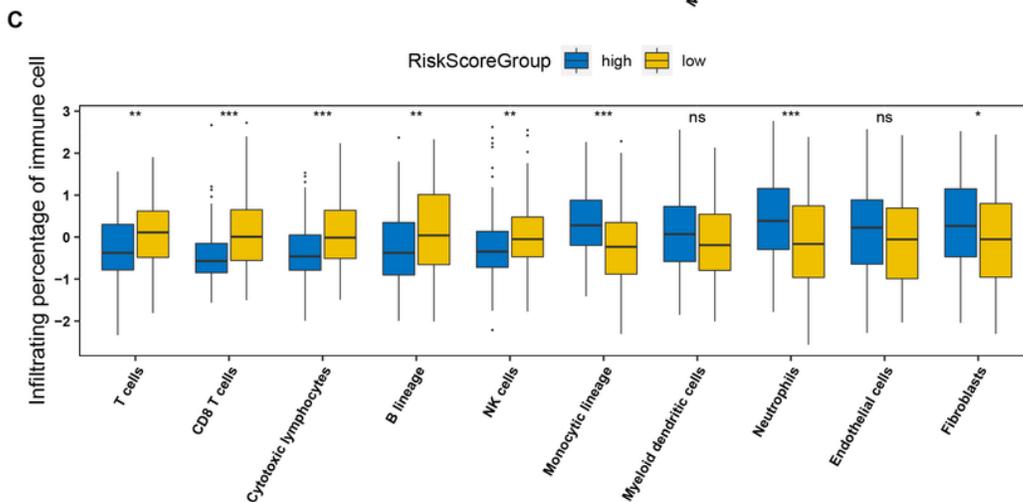
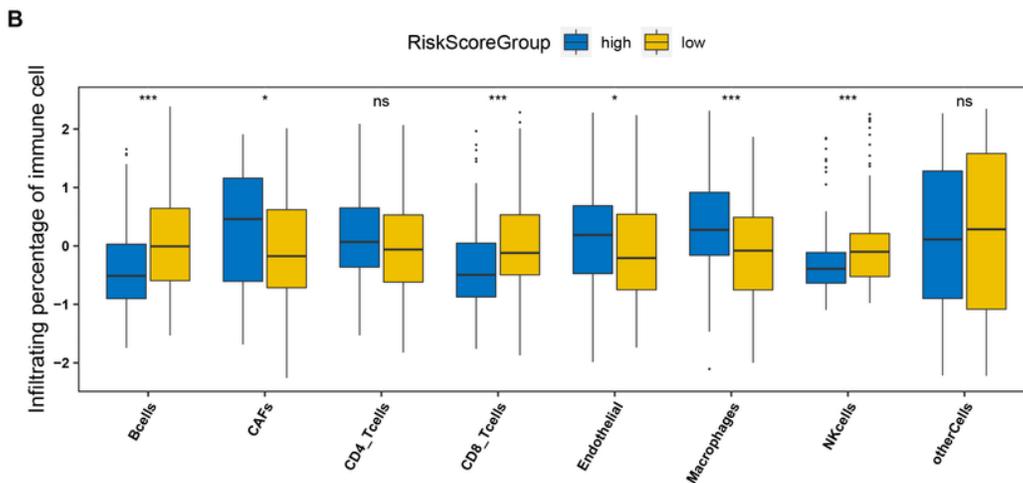
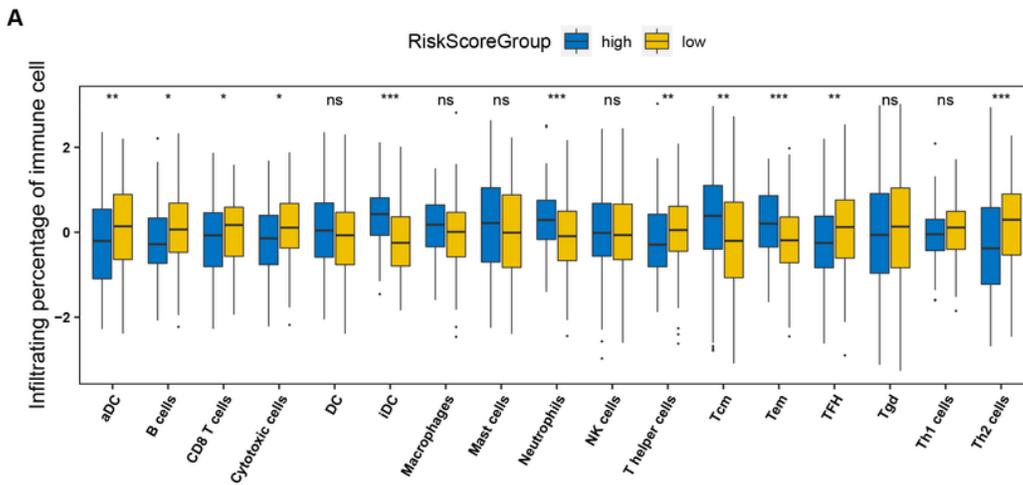


Figure 7

Tumor immune cell infiltration estimate

A. The plot shows the proportion of tumor immune cell infiltration by EPIC estimate; B. The plot shows the proportion of tumor immune cell infiltration by MCPcounter estimate; C. The plot shows the proportion of tumor immune cell infiltration by ssGSEA estimate (NS: no significant; \* :  $p < 0.05$ ; \*\* :  $p < 0.01$ ; \*\*\* :  $p < 0.001$ ).

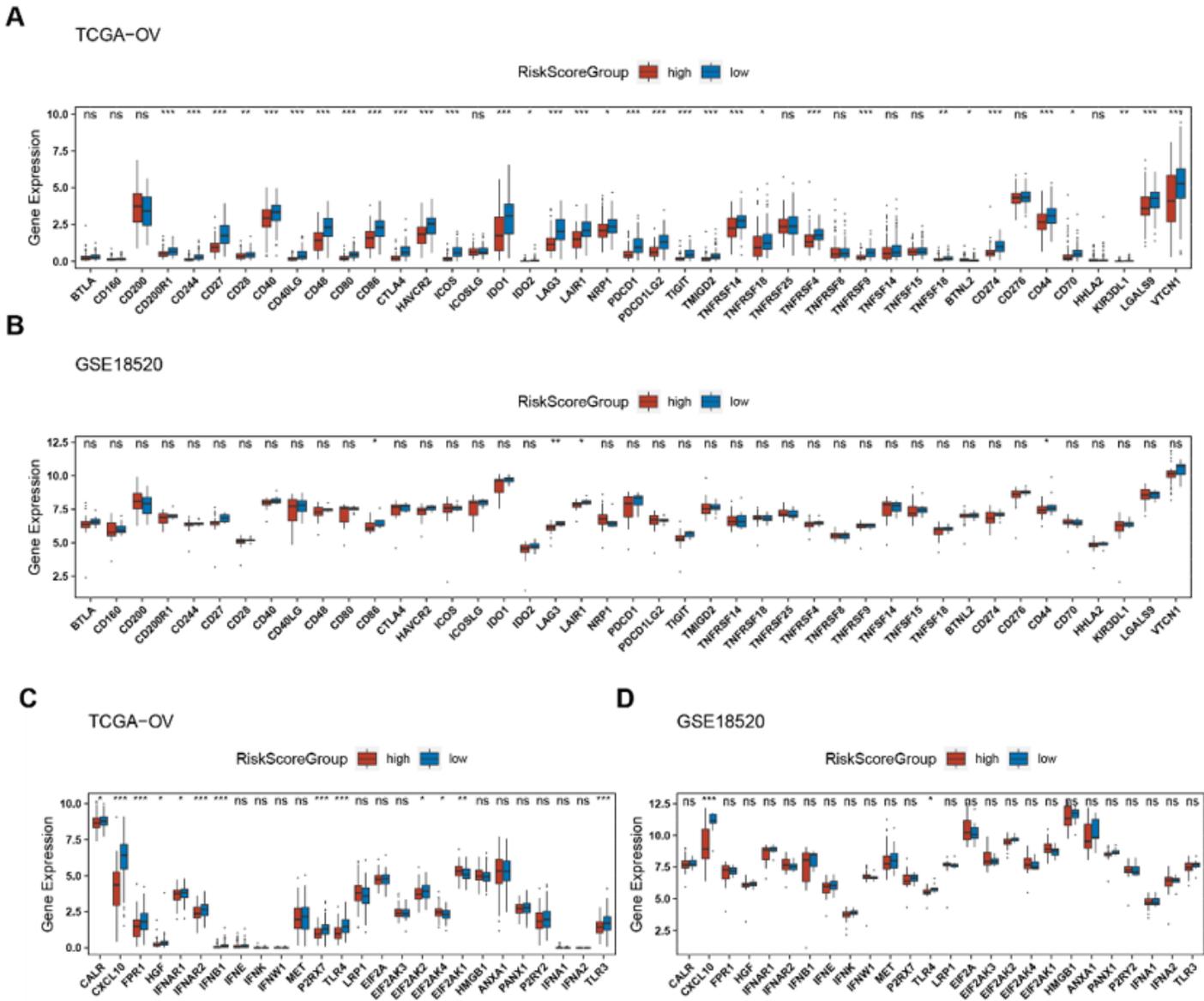
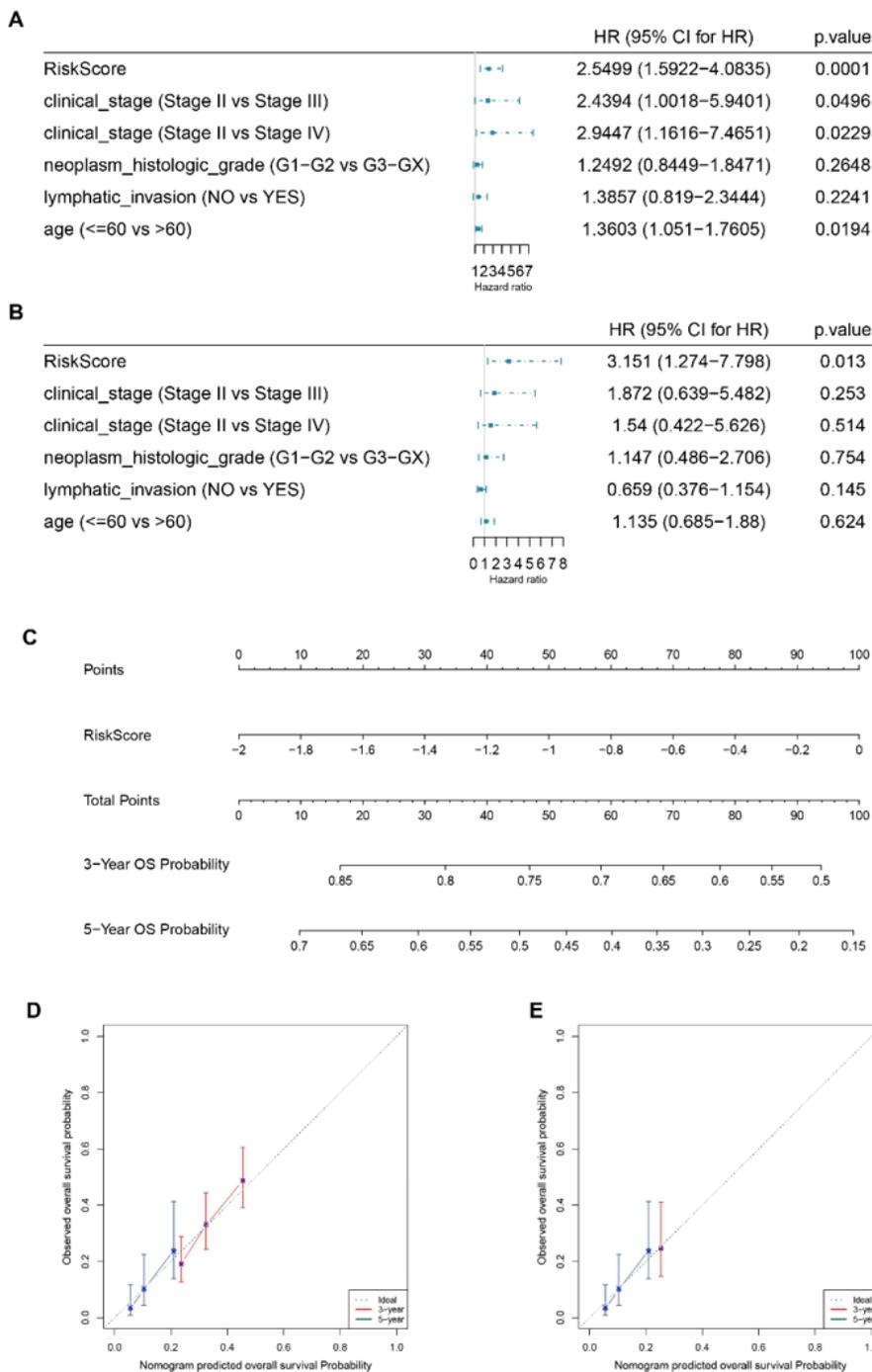


Figure 8

Expression of ICs and IDP-related genes

A. The plot shows the expression of ICs-related genes in TCGA-OV cohort; B. The plot shows the expression of ICs-related genes in samples from the external validation set GSE18520; C. The plot shows the expression

of IDP-related genes in TCGA-OV cohort; D. The plot shows the expression of IDP-related genes in the external validation set. (NS: no significant; \* :  $p < 0.05$ ; \* \* :  $p < 0.01$ ; \* \* \* :  $p < 0.001$ )



**Figure 9**

### Construction and validation of clinical nomogram

A. Forestplot shows the HR and P values of univariate Cox risk regression analysis with riskscore and clinically relevant features; B. Forestplot showing HR and P values of multivariate Cox risk regression

analysis with riskcore and clinically relevant features; C. nomogram based on risk score for predicting 3 and 5-year overall survival of ovarian cancer patients; D. Calibration curves validation of consistency between predicted 3 and 5-year overall survival and actual overall survival based on TCGA-OV cohort; E. Calibration curves for evaluating the conformance of predicted 3, 5-year overall survival with actual survival based on external validation data GSE18520. (HR hazard ratio, CI Confidence interval)

## Supplementary Files

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

- [SupplementaryFigures.docx](#)