Development and Verification of an Immune-Related Gene Pairs Prognostic Signature in Ovarian Cancer

Bao Zhang
Shengjing Hospital of China Medical University
Xiaocui Nie
Shenyang women's and children's hospital
Xinxin Miao
Shengjing Hospital of China Medical University
Shuo Wang
Shengjing Hospital of China Medical University
Jing Li
Shengjing Hospital of China Medical University
Shengke Wang (skwang@cmu.edu.cn)
Shengjing Hospital of China Medical University https://orcid.org/0000-0002-5863-5012

Research

Keywords: Ovarian cancer, Immune-related gene pairs, TCGA, Prognostic signature, Bioinformatics

Posted Date: April 15th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-22420/v1

Abstract

Background: Ovarian cancer (OV) is the most common gynecological cancer and is a major cause of cancer-related death among women worldwide. Immunotherapy has recently been of great interest, and it has been proven to be an effective treatment strategy. For this reason, the work here attempts to produce a prognostic immune-related gene pair (IRGP) signature to estimate OV patient survival.

Methods: The Gene Expression Omnibus (GEO) and the Cancer Genome Atlas (TCGA) databases provided the genetic expression profiles and clinical data of OV patients, whose samples were divided into training, validation, and testing sets. The InnateDB database and the least absolute shrinkage and selection operator (LASSO) regression model, were used to build an IRGP signature. The functional enrichment analysis was performed to investigate the biological activities of IRGPs. Then, the correlation between risk scores and clinicopathological parameters was further analyzed. Finally, we compared our signature and four existing signatures for OV.

Results: We first identified a 17-IRGP signature significantly associated with survival based on LASSO regression model. The average area under the curve (AUC) values of the training, validation, and all TCGA sets were 0.869, 0.712, and 0.778, respectively. In addition, the average AUC of the independent testing set was 0.73. The 17-IRGP signature noticeably split patients into high- and low-risk groups with different prognostic outcomes. As suggested by a functional study, some biological pathways, including the Toll-like receptor and chemokine signaling pathways, were significantly negatively correlated with risk scores; however, pathways such as the p53 and apoptosis signaling pathways had a positive correlation. Moreover, tumor stage III, IV, grade G1/G2, and G3/G4 samples had significant differences in risk scores. High- and low-risk groups of the four TCGA OV subtypes also had significant prognostic differences. When compared with four other signatures for OV, our 17-IRGP signature had better prognostic prediction performance.

Conclusions: An effective 17-IRGP signature was produced to predict prognostic outcomes in OV, providing new insights into immunological biomarkers.

Background

Ovarian cancer (OV) is the most common gynecological cancer and is a major cause of cancer-related death among females worldwide. The number of new cases was estimated at approximately 295,414, and the number of deaths was approximately 184,799 in 2018 (1). A family history of OV is an important risk factor with known genetic predisposition (2). Standard treatments include platinum-based chemotherapy, but most tumors become resistant to the treatment (3). Despite improved treatment outcomes in recent years, the prognosis of patients with advanced OV remains poor. The emergence of chemoresistant diseases confined to the peritoneum is the leading cause of death (4). Thus, an in-depth understanding of the molecular functions of OV could lead to new diagnostic, predictive, prognostic, and therapeutic biomarkers.

Substantial interest in the field of immunotherapy has recently emerged, and immunotherapy has been proven effective in the treatment of human malignancies (5). Altered phenotype and function of major immune cell subsets (including bone marrow cells, macrophages, dendritic cells, and T cells) in the OV microenvironment have been reported in response to immunotherapy (6). Preclinical studies have also been conducted in the past decade, with most emphasis on the use of programmed cell death protein 1 (PD-1) or its ligand (PD-L1) to induce cell death. Research based on tumor biology is exploring new and more effective immunotherapies. There have been several combination therapies, such as check-point inhibitors, anti-VEGF therapy, PARP inhibitors, and adoptive immunotherapies in OV treatments (7). Additionally, targeting other immunosuppressive pathways may become a way to enhance responses to immunotherapy.

Recently, based on microarray and RNA-sequencing methods, there have been increasing studies on immune-related prognostic signatures in human cancers. For example, using a cohort of glioma samples with expression information from whole genome microarrays from the Chinese Glioma Genome Atlas and TCGA databases, researchers constructed a local risk signature associated with immunity that can independently identify patients with a high risk (8). Wang *et al*. (9) used the TCGA and ImmPort databases to build a 15-gene prognostic model in renal papillary cell carcinoma. Their signature was associated with tumor staging and tumor type. Other immune-related signatures have also been reported in cancers, including gastric cancer (10), anaplastic gliomas (11), breast cancer (12), and pancreatic cancer (13). However, there have been no reports of immune-related gene signatures in OV.

In the present study, we used the genetic expression profiles and clinical data of OV cases harvested according to the TCGA and GEO databases, which were divided into training, validation, and testing sets. Based on the InnateDB database and LASSO regression model, we identified a 17-IRGP signature that was significantly associated with survival. When compared with four other signatures in OV, our 17-IRGP signature had better prognostic prediction performance. In conclusion, an effective 17-IRGP signature was produced to predict prognostic outcome in OV, providing new insights into immunological biomarkers.

Methods

Gene expression data source

The OV datasets in this study were derived from the TCGA (14) and GEO (15) databases. First, we used the GDC Data Transfer Tool to download RNA-sequencing data as FPKM files and corresponding clinical information of patients from the TCGA (https://portal.gdc.cancer.gov/) database. The download time was March 2019. Here, a total of 334 OV samples were included. Second, the GEO database (https://www.ncbi.nlm.nih.gov/geo/) provided OV gene expression profiles, including GSE14764 (n = 80 samples) (16) and GSE26712 (n = 195 samples) (17). They were both performed on the GPL96 platform (Affymetrix Human Genome U133A Array).

Data processing

To ensure the analysis consistency in different datasets, we downloaded the raw data of GSE14764 and GSE26712 and used the robust multiarray average (RMA) method (18) for homogenization. Because both GEO datasets were performed by GPL96 platform, we merged them into an independent external validation set for subsequent analysis and performed batch correction to eliminate batch effects. Before constructing the prognostic signature, the original data were preprocessed by 1) removing tumor samples without clinical information and overall survival (OS) was 0 day; 2) removing normal samples; 3) removing genes with low expression (gene expression was missing or 0 in more than half of all samples); and 4) retaining only the expression profiles of immune-related genes. The clinical information of patients in two GEO datasets were shown in **Supplementary Table 7**. The preprocessed dataset ultimately contained a total of 594 OV samples.

Computation of immune-related gene pairs

First, we downloaded the genes associated with immune from the InnateDB dbase (https://www.innatedb.com/) (19). This database records the innate immune-related genes of multiple species that are supported by the literature and manually corrected. Here, we obtained endogenous human IRGs. After sorting (removing genes with duplicate symbols), there were a total of 1,039 IRGs (**Supplementary Table 1**). IRGPs were constructed based on 1,039 IRGs with the calculation rules as described previously (20). In brief:

If $Expr_{IRG1} < Expr_{IRG2}$, IRGP = 1, else: IRGP = 0

$Expr_{IRG1}$ is the expression value of immune gene 1, and $Expr_{IRG2}$ is the expression value of immune gene 2. According to the above rules, the IRGP value of each dataset was calculated separately, and a total of (1,039 * 1,038)/2 IRGPs were obtained. After removing samples in all datasets with a gene pair value of 0 or 1, we left the residual IRGPs and subsequently selected them as first candidate IRGPs to conduct further analysis (**Supplementary Fig. 1**).

Construction of prognostic IRGP signatures

First, 334 samples were divided into the training or validation set in the TCGA dataset. To avoid the influence of random assignment bias on the stability of the subsequent modeling, all samples were put back into random groupings 100 times. The samples were selected according to a ratio of 1:1 in the training set and validation set. Here, the training set and validation set with no significant differences in the distributions of age, tumor stage, follow-up time, and patient survival status were selected for signature construction. The resulting training set and validation set samples are available in Table 1. The prognostic signature was constructed in two steps as follows: 1) a univariate Cox proportional hazards regression model was employed for the calculation of the relationship between each IRGP and patients' prognosis with log-rank test p-value < 0.05; 2) we used the LASSO regression to further filter the above IRGPs to reduce the numbers in the risk model.

Table 1
The clinical information distributions of training set and validation set samples

Clinical features	Training set (n = 166)	Validation set (n = 166)	Chi-square test p-value
Status			0.5005
Alive	69	62	
Dead	97	104	
Grade			1
G1_G2	21	21	
G3_G4	140	143	
GX	5	2	
Stage			0.8261
II	11	11	
III	127	131	
IV	27	23	
Unknown	1	1	
Age			0.9278
0 ~ 50	33	35	
50 ~ 60	56	51	
60 ~ 70	42	39	
70 ~ 80	29	34	
80 ~ 100	6	7	
Note GX represents grade cannot be assessed.			

Functional enrichment analysis

We used the clusterProfiler R package (21) to analyze the IRGPs for molecular function (MF), cellular component (CC), biological process (BP) enrichment by studying the Gene Ontology (GO) terms. A q-value < 0.05 was set as the threshold for significant enrichment. The dot plot of clusterProfiler displayed the enrichment results. We took the genes with differential expression (DEGs) in groups of high- and low-risk with the use of the rank test with false discovery rate (FDR) < 0.05, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment of DEGs was carried out using the GSVA R package.

Evaluation of IRGP prognostic signature

Based on the 17 IRGPs obtained by LASSO regression and its regression coefficient, we obtained the risk score of each respective OV patient by:

Risk score = $\sum$IRGP * coefficient

Here, the IRGP was the immune gene pair, and the coefficient was the regression coefficient. We split all OV samples into a low-risk group (Risk-L) or a high-risk group (Risk-H) in line with the median risk score. We used the performance of this signature to plot the receiver operating characteristic (ROC) curves.

Comparison of IRGP risk signatures with other models

To evaluate the performance of the 17-IRGP risk model and other existing prognostic models, we selected models that were also constructed based on gene expression data for comparison. After consulting the literature, four prognostic risk models were selected: Liu (7-gene signature, 2018, based on GSE32062 for Japan-cohort in GEO, GSE63885 for Poland-cohort in GEO, and E-MTAB-386 for USA-cohort in arrayexpress) (22), Liu (5-gene signature, 2016, based on TCGA database) (23), Hou (6-gene signature, 2018, based on n TCGA database with records of Taxol treatment) (24), and Zhang (2-gene signature, 2018, based on TCGA database) (25). To make the model more comparable, we used the same method for the calculation of the risk score of each OV sample and evaluated the AUC value of each model. According to the median risk score, the samples were also split into Risk-H and Risk-L groups, and the difference in overall survival between the two groups was calculated by log-rank test. Using the concordance index (C-index) and restricted mean survival (RMS), we assessed the prognosis precision of the signature (26). RMS can be interpreted as the average free-event survival time in a specific time period, which is equivalent to the area under the Kaplan-Meier (K-M) survival curves at a specific time point. The higher the RMS time ratio, the greater the difference is in prognosis. Here, we evaluated the period between 0 to 150 months RMS of each model and drew the RMS curve.

Statistical analysis

With the use of R (version 3.6.1) software, we carried out statistical analyses. We used the survival R package for univariate and multivariate risk regression analysis, the glmnet package for LASSO Cox regression model, and the survivalROC package to evaluate the model's ROC curve and calculated AUC values. Moreover, the KMsurv R package was performed to show the K-M curves of grouped samples. The C-index calculation used the survcomp R package, and RMS cure and RMS time calculation were performed by using the survival and survRM2 packages. In terms of each test, a p-value < 0.05 suggested a significant difference. * p-value < 0.05, ** p-value < 0.01, and *** p-value < 0.001 express statistically significant characteristics.

Results

Construction of IRGP signatures

In this study, in accordance with the gene expression profiles of OV from the TCGA database, we performed a series of bioinformatics analyses to build a model for prognostic evaluation of IRGPs in OV patients (Fig. 1). Here, a total of 1,039 IRGs from the InnateDB database were obtained. Of these genes, 539,241 IRGPs were established. After removing samples in all datasets with gene pair values of 0 or 1, we left the residual IRGPs and subsequently selected them as first candidate IRGPs to conduct further analysis. Then, a univariate Cox proportional hazards regression model was adopted for the calculation of the relationship between each IRGP and patient survival with a log-rank test p-value < 0.05. A total of 3,765 identified IRGPs were identified (**Supplementary Table 2**). Moreover, using the LASSO regression model, we further selected 17 IRGPs for the final risk prognostic signature (Table 2, **Supplementary Fig. 2**).

Table 2
The results of 17 IRGPs using LASSO regression model

IRGPs	Coef	p-value	HR	Low.95CI	High.95CI
LILRA2 vs P2RY14	1.083615	0.00014	2.955345	1.69188	5.162343
NOD2 vs LILRA2	-1.15639	0.000278	0.31462	0.168658	0.586905
CXCL14 vs SHARPIN	0.877286	0.003049	2.404364	1.345709	4.295853
TSC22D3_vs_CXCL11	1.177209	0.005145	3.245304	1.422665	7.403006
FOXA2_vs_RCAN1	-0.68767	0.007236	0.502744	0.30437	0.83041
PLA2G4A vs SCAF11	-0.84326	0.019819	0.430307	0.211683	0.874725
ABCA1 vs MST1R	0.75507	0.027164	2.127761	1.088905	4.157725
STAT4 vs IL1R2	-0.85268	0.031899	0.426269	0.195622	0.928859
AP3B1 vs BTN3A3	0.525508	0.040716	1.691318	1.022447	2.797755
MID1 vs THBS1	-1.69541	0.069621	0.183524	0.029397	1.145736
IFNGR1 vs CASP6	0.483861	0.250733	1.622325	0.710478	3.704465
MSR1 vs CXCL11	-0.33671	0.265347	0.714119	0.394875	1.291463
BTN3A2 vs IRF2	-0.68764	0.281319	0.502761	0.143903	1.756516
IL1B vs CXCR3	0.37003	0.284697	1.447778	0.735004	2.851768
SNX27 vs CXCL11	0.364458	0.388464	1.439734	0.628867	3.296136
CASP7 vs CXCL11	0.274334	0.509042	1.315654	0.582788	2.970113
BTN3A3 vs TPST1	-0.00035	0.998939	0.999646	0.59306	1.684976

Note Coef: coefficient by LASSO analysis. HR: Hazard Ratio. Low.95CI: low 95% confidence interval (CI). High.95CI: high 95% CI.

The evaluation and validation of IRGP signatures for survival prediction

Based on the above 17 IRGPs, we constructed a prognostic risk model for OV patients. Since the overall survival time of patients was distributed over more than 2 years (**Supplementary Fig. 3**), the predictive effect of this model on datasets for 1, 3, and 5 years was evaluated. Next, the IRGPs were adopted for the calculation of the risk score for the respective case in the TCGA training group. The average AUC of the training set was 0.869, and the average AUC of the validation set was 0.712 (**Supplementary Table 3**). In addition, the average AUC of all TCGA datasets was 0.778, and the average AUC of the independent testing set was 0.73 (Fig. 2A-D). By dividing OV patients into low- (Risk-L) and high-risk groups (Risk-H) based on the median risk score, we observed that the Risk-H group exhibited noticeably poorer prognosis than the Risk-L group on the training set, validation set, all TCGA set and independent GEO testing set (log-rank test p-value < 0.05, Fig. 3A-D).

Functional analysis of immune-related gene pairs

In our signature, the 17 IRGPs contained a total of 29 immune genes. These genes were significantly enriched in the “caspase” and “interferon receptor” families (p-value < 0.01, Table 3). The GO annotation results of 29 genes suggested that 4 genes were significantly enriched in interleukin-1 secretion and inflammatory bowel disease biological processes (Fig. 4A). Seven of the 17 IRGPs showed a negative correlation with risk scores, and ten showed a positive correlation (Fig. 4B, Table 4). Here, IRGPs reflected the level of relative expression between two genes. We found that genes related to immune activation/response (P2RY14, CXCL11, CXCR3, MST1R, and NOD2) showed relatively low expression levels (**Supplementary Table 4**). The relatively highly expressed genes (CASP6, CASP7, IL1B, THBS1, and TPST1) were mainly related to the apoptotic process and inflammation response (**Supplementary Table 5**). Further enrichment analysis of DEGs in the high- and low-risk groups was performed (**Supplementary Table 6**). The results suggested that the immune response-related pathways (Toll-like receptor and chemokine signal pathway, and others) were significantly negatively correlated with risk scores; however, the pathways such as p53 signaling pathway and apoptosis had a positive correlation with the risk scores (Fig. 4C), which seems to indicate that samples from the low-risk group may have higher immune activity (or immune activation status).

Table 3

The gene family enrichment results of 29 immune-related genes

Gene family	Genes	p-value	FDR
Caspases	CASP7/CASP6	0.000168	0.005042
Interferon receptors	IFNGR1	0.007737	0.232112
P2Y receptors	P2RY14	0.011584	0.347516
V-set domain containing	BTN3A2/BTN3A3	0.019208	0.576234
ATP binding cassette subfamily A	ABCA1	0.019234	0.577029
Clathrin/coatomer adaptor, adaptin-like, N-terminal domain containing	AP3B1	0.019234	0.577029
Zinc fingers RANBP2-type	SHARPIN	0.028087	0.842623
NLR family	NOD2	0.033112	0.993349
Scavenger receptors	MSR1	0.035614	1
Sorting nexins	SNX27	0.038111	1
Sulfotransferases, membrane bound	TPST1	0.048034	1
Receptor tyrosine kinases	MST1R	0.05173	1
Phospholipases	PLA2G4A	0.054186	1
Interleukin receptors	IL1R2	0.054186	1
Interleukins	IL1B	0.055412	1
Forkhead boxes	FOXA2	0.055412	1
Chemokine ligands	CXCL14	0.057858	1
CD molecules	LILRA2/CXCR3	0.092531	1
Tripartite motif containing	MID1	0.117076	1
SH2 domain containing	STAT4	0.123936	1
Endogenous ligands	CXCL11	0.258581	1
Ring finger proteins	SCAF11	0.327935	1
unknown	IRF2:RCAN1:THBS1:TSC22D3	1	1

Table 4

Analysis of correlation between gene pair values and risk scores

IRGPs	p-value	Pearson correlation	Type
TSC22D3_vs_CXCL11	1.08E-27	0.54906008	Positive
CASP7_vs_CXCL11	4.08E-23	0.50599649	Positive
SNX27_vs_CXCL11	1.42E-19	0.467906634	Positive
AP3B1_vs_BTN3A3	5.90E-16	0.42334027	Positive
IL1B_vs_CXCR3	6.39E-14	0.39504485	Positive
MSR1_vs_CXCL11	1.10E-12	0.37640839	Positive
LILRA2_vs_P2RY14	2.03E-12	0.372271702	Positive
ABCA1_vs_MST1R	9.10E-10	0.32707845	Positive
CXCL14_vs_SHARPIN	9.39E-07	0.264532336	Positive
IFNGR1_vs_CASP6	2.01E-03	0.168462127	Positive
PLA2G4A_vs_SCAF11	2.69E-04	-0.198123738	Negative
MID1_vs_THBS1	1.16E-08	-0.305853734	Negative
FOXA2_vs_RCAN1	3.98E-11	-0.3511514	Negative
BTN3A3_vs_TPST1	3.63E-11	-0.35182302	Negative
STAT4_vs_IL1R2	1.72E-11	-0.357242428	Negative
BTN3A2_vs_IRF2	1.73E-12	-0.37335433	Negative
NOD2_vs_LILRA2	1.09E-12	-0.376498688	Negative
Figures Legends			

Relationship between prognostic risk signature and clinical features

Using clinical information such as age, tumor stage, and grade from the TCGA database, we analyzed the relationship between the 17-IRGP risk signature and clinical characteristics. Here, samples from the tumor stage II, III and IV groups showed significant differences in risk scores (Fig. 5A), but no significant difference was observed for prognosis of the high- and low-risk group samples in stage II (p-value > 0.05). However, significant differences were shown in stages III and IV, indicating that the model may be more suitable for stage III/IV OV patients (Fig. 5B-D). For tumor grades, G1/G2 and G3/G4 samples had no significant difference in risk scores (Fig. 5E). However, the prognosis of the high- and low-risk group

samples showed significant differences in G1/G2 and G3/G4 (Fig. 5F-G). We also did not observe a significant correlation between age and risk scores (Fig. 5H).

The performance of prognostic risk signature in OV subtypes

The TCGA project revealed that surviving gene expression characteristics can predict clinical outcomes and divide OV patients into four transcription subtypes, including differentiated, immunoreactive, mesenchymal, and proliferative (27). We next compared the prognostic performance of our model on these four molecular subtypes. Low- and high-risk groups of the four subtypes were identified to have significant prognostic differences (Fig. 6A-D). In addition, we also found the best prognosis in immunoreactive subtype, Risk-L samples, while the worst prognosis mesenchymal subtype, Risk-H samples. Moreover, OV was divided into four immune subtypes (C1-C4) (28) based on immune molecular tags. We further compared the model's performance on different immune subtypes (the C3 immune subtype had only 3 samples and was not added to the analysis). Among the above three immune subtypes, there were also different survival outcomes between the high- and low-risk groups in both the C1 and C2 immune subtypes (Fig. 6E-G).

Comparison of our prognostic risk signature with other models

Finally, using the same method, we evaluated the AUC values of four existing OV prognostic models at 1, 3, and 5 years. The average AUC values of these 4 models were all < 0.7, which were lower than the AUC of our 17-IRGPs signature, indicating that our model has better prediction performance. Among the 4 models, only the Risk-H and Risk-L groups calculated by the 5-gene signature model have no significant difference in prognosis, and other 3 models showed significant differences in prognosis (Fig. 7A-D). Based on the C-index of above five prognostic models, the 17-IRGPs model has the largest C-index (Fig. 8A), indicating that the overall performance of our model was better than the other four models. The RMST curves of the five models also show significant differences. The risk scores of these models have a very significant relationship with the prognosis (HR > 1, p-value < 0.0001), but we see that the RMST cure of 17-IRGPs was better than the other four models, which has a steeper slope (Fig. 8B), indicating that our model can better evaluate the survival rate in OV.

Moreover, we examined using GSE14764 and GSE26712, separately. As shown in **Supplementary Fig. 4**, we first used the same method in the GSE14764 dataset to evaluate the AUC values of four existing OV prognostic models. We can see that AUC values of all four existing OV prognostic models are very small. The results of most K-M curves are not statistically significant. Similar results can still be revealed in GSE26712 dataset (**Supplementary Fig. 5**). Therefore, combining the above analysis results, we can find that in TCGA, GSE14764, or GSE26712 dataset, the AUC values of four existing OV prognostic models are

very small, indicating that our 17-IRGPs model can better evaluate the survival rate in OV (**Supplementary Fig. 6**).

Discussion

In this study, based on the TCGA database and LASSO regression model, we identified a 17-IRGP signature that was significantly associated with survival. This robust 17-IRGP signature can estimate prognosis in OV and provide new insights into immunological biomarkers.

Immunotherapy strategies in cancers aim to develop combination methods to enhance immunity and prevent local immunosuppression. Chimeric antigen receptor-modified T cells, cancer vaccines, immune checkpoint blockade, and antibody-based therapies have shown preclinical success and have been clinically tested in OV (29). Supported by the methods of RNA-sequencing and microarray, as well as available gene expression databases such as TCGA and GEO, an increasing number of reports of gene prognosis models of cancers have appeared recently. For example, studies have been reported in melanoma (30), breast cancer (31), clear cell renal cell carcinoma (32), and other cancers. However, there are few reports on immune-related gene signature associated with cancer prognosis. In colorectal cancer (CRC), Wu *et al*. (33) used genetic expression profiles and clinical data of cases to construct a 19-IRGP signature that covers 36 individual genes. Their IRGP signature can stratify CRC cases into low- and high-risk groups by prognostic outcome. This effective IRGP signature that predicts prognostic outcomes in CRC, covering early-stage disease, is capable of providing novel knowledge of identifying cases at a high risk of mortality. However, there is no prognostic model of IRGPs reported in OV.

Based on the functional analysis of IRGPs, the Toll-like receptor and chemokine signal pathways were significantly negatively correlated with risk scores. Toll-like receptors (TLRs), as the most important pattern recognition receptors in innate immunity, play an important role in inducing immune responses by recognizing microbial invaders or specific agonists (34). The antitumor effect of TLRs can directly induce tumor cell death and activate an effective antitumor immune response (35). It can trigger an inflammatory response and cell survival in the tumor microenvironment. TLR2, TLR3, TLR4 and TLR5 were reported to be highly expressed in normal and neoplastic ovarian epithelium (36). However, pathways such as the p53 signaling and apoptosis pathways had a positive correlation with the risk scores. Thus, the above signaling pathways were shown to be closely related to the risk scores of our signature and may be involved in the immune response to OV.

The prognostic signature in our study consists of 29 unique IRGs. CXCL11 (CXC chemokine ligand 11) is a chemokine involved in the progression of various cancers. CXCL11 is overexpressed in CRC tissues and cell lines. Repression of CXCL11 significantly inhibited cell migration, invasion and epithelial-mesenchymal transition (EMT) (37). It was also reported that its downregulation can inhibit tumor angiogenesis in epithelial OV (38). High CXCL11 expression was determined to predict worse OS in high-grade serous OV (39). STAT4 (signal transducer and activator of transcription 4) is a member of the STAT family. Its overexpression was shown to be associated with poor prognosis in OV patients (40). It has

also been reported to be involved in the occurrence and development of gastric cancer (41) and hepatocellular cancer (42). In addition, the expression of forkhead box A2 (FOXA2) in colon cancer tissues is upregulated and related to the metabolism and clinical stages (43). Moreover, FOXA2 is capable of facilitating EMT, inhibiting apoptosis and enhancing colon cancer cell invasion ability. In OV, miR-590-3p can promote growth and metastasis via the FOXA2-Versican pathway (44). According to the above results, the genes involved in the IRGP signature play a significant role in human cancers.

There are also some limitations to our study. First, the robustness of IRGPs was based on the gene expression profiles produced by RNA sequencing and microarray data and must be verified in large clinical samples of OV. Second, further experimental validation is required.

Conclusion

In conclusion, based on comprehensive bioinformatics analysis method, an effective 17-IRGP signature was produced to predict prognostic outcomes in OV, providing new insights into immunological biomarkers.

Abbreviations

OV: Ovarian cancer; IRGP: immune-related gene pair; GEO: Gene Expression Omnibus; TCGA: the Cancer Genome Atlas; LASSO: least absolute shrinkage and selection operator; AUC: average area under the curve; RMA: robust multiarray average; OS: overall survival; MF: molecular function; CC: cellular component; BP: biological process; DEGs: differential expression genes; FDR: false discovery rate; KEGG: Kyoto Encyclopedia of Genes and Genomes; ROC: receiver operating characteristic; C-index: concordance index; RMS: restricted mean survival; K-M: Kaplan-Meier.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Authors' contributions

BZ and XCN designed experiments. XXM, SW, and JL contributed to the literature review. BZ wrote the initial draft of the manuscript. SKW designed the study and edited the paper. All authors have approved the final version of the manuscript.

Competing interests

The authors declare no conflicts of interest.

Acknowledgements

We thank the Department of Obstetrics and Gynecology in Shengjing Hospital of China Medical University for technical advice. We also gratefully thank American Journal Experts (https://www.aje.cn/) for editing the present paper (F7C8-397B-2647-877F-4324).

Funding

This work was supported by the science and technology project of Shenyang (Grant No. 2019-ZD-0984).

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
2. Torre LA, Islami F, Siegel RL, Ward EM, Jemal A. Global Cancer in Women: Burden and Trends. Cancer Epidemiol Biomarkers Prev. 2017;26(4):444–57.
3. Binju M, Amaya-Padilla MA, Wan G, Gunosewoyo H, Suryo Rahmanto Y, Yu Y. Therapeutic Inducers of Apoptosis in Ovarian Cancer. Cancers (Basel). 2019;11(11).
4. Nath S, Saad MA, Pigula M, Swain JWR, Hasan T. Photoimmunotherapy of Ovarian Cancer: A Unique Niche in the Management of Advanced Disease. Cancers (Basel). 2019;11(12).

5. Doo DW, Norian LA, Arend RC. Checkpoint inhibitors in ovarian cancer: A review of preclinical data. Gynecol Oncol Rep. 2019;29:48–54.
6. Wang W, Liu JR, Zou W. Immunotherapy in Ovarian Cancer. Surg Oncol Clin N Am. 2019;28(3):447–64.
7. Ghisoni E, Imbimbo M, Zimmermann S, Valabrega G. Ovarian Cancer Immunotherapy: Turning up the Heat. Int J Mol Sci. 2019;20(12).
8. Cheng W, Ren X, Zhang C, Cai J, Liu Y, Han S, et al. Bioinformatic profiling identifies an immune-related risk signature for glioblastoma. Neurology. 2016;86(24):2226–34.
9. Wang Z, Song Q, Yang Z, Chen J, Shang J, Ju W. Construction of immune-related risk signature for renal papillary cell carcinoma. Cancer Med. 2019;8(1):289–304.
10. Jiang B, Sun Q, Tong Y, Wang Y, Ma H, Xia X, et al. An immune-related gene signature predicts prognosis of gastric cancer. Med (Baltim). 2019;98(27):e16273.
11. Wang W, Zhao Z, Yang F, Wang H, Wu F, Liang T, et al. An immune-related lncRNA signature for patients with anaplastic gliomas. J Neurooncol. 2018;136(2):263–71.
12. Zhao J, Wang Y, Lao Z, Liang S, Hou J, Yu Y, et al. Prognostic immune-related gene models for breast cancer: a pooled analysis. Onco Targets Ther. 2017;10:4423–33.
13. Wei C, Liang Q, Li X, Li H, Liu Y, Huang X, et al. Bioinformatics profiling utilized a nine immune-related long noncoding RNA signature as a prognostic target for pancreatic cancer. J Cell Biochem. 2019;120(9):14916–27.
14. Cancer Genome Atlas Research N. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45(10):1113–20.
15. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207–10.
16. Denkert C, Budczies J, Darb-Esfahani S, Gyorffy B, Sehouli J, Konsgen D, et al. A prognostic gene expression index in ovarian cancer - validation across different independent data sets. J Pathol. 2009;218(2):273–80.
17. Vathipadiekal V, Wang V, Wei W, Waldron L, Drapkin R, Gillette M, et al. Creation of a Human Secretome: A Novel Composite Library of Human Secreted Proteins: Validation Using Ovarian Cancer Gene Expression Data and a Virtual Secretome Array. Clin Cancer Res. 2015;21(21):4960–9.
18. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4(2):249–64.
19. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, et al. InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. Nucleic Acids Res. 2013;41(Database issue):D1228-33.
20. Li B, Cui Y, Diehn M, Li R. Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer. JAMA Oncol. 2017;3(11):1529–

37.

21. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.
22. Liu G, Chen L, Ren H, Liu F, Dong C, Wu A, et al. Seven Genes Based Novel Signature Predicts Clinical Outcome and Platinum Sensitivity of High Grade IIIc Serous Ovarian Carcinoma. Int J Biol Sci. 2018;14(14):2012–22.
23. Liu LW, Zhang Q, Guo W, Qian K, Wang Q. A Five-Gene Expression Signature Predicts Clinical Outcome of Ovarian Serous Cystadenocarcinoma. Biomed Res Int. 2016;2016:6945304.
24. Hou S, Dai J. Transcriptome-based signature predicts the effect of taxol in serous ovarian cancer. PLoS One. 2018;13(3):e0192812.
25. Zhang J, Xu M, Gao H, Guo JC, Guo YL, Zou M, et al. Two protein-coding genes act as a novel clinical signature to predict prognosis in patients with ovarian serous cystadenocarcinoma. Oncol Lett. 2018;15(3):3669–75.
26. Eng KH, Schiller E, Morrell K. On representing the prognostic value of continuous gene expression biomarkers with the restricted mean survival curve. Oncotarget. 2015;6(34):36308–18.
27. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474(7353):609–15.
28. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. Immunity. 2018;48(4):812–30. e14.
29. Krishnan V, Berek JS, Dorigo O. Immunotherapy in ovarian cancer. Curr Probl Cancer. 2017;41(1):48–63.
30. Sun L, Li P, Ren H, Liu G, Sun L. A four-gene expression-based signature predicts the clinical outcome of melanoma. J BUON. 2019;24(5):2161–7.
31. Bao X, Anastasov N, Wang Y, Rosemann M. A novel epigenetic signature for overall survival prediction in patients with breast cancer. J Transl Med. 2019;17(1):380.
32. Zhang C, He H, Hu X, Liu A, Huang D, Xu Y, et al. Development and validation of a metastasis-associated prognostic signature based on single-cell RNA-seq in clear cell renal cell carcinoma. Aging (Albany NY). 2019;11.
33. Wu J, Zhao Y, Zhang J, Wu Q, Wang W. Development and validation of an immune-related gene pairs signature in colorectal cancer. Oncoimmunology. 2019;8(7):1596715.
34. Shi M, Chen X, Ye K, Yao Y, Li Y. Application potential of toll-like receptors in cancer immunotherapy: Systematic review. Med (Baltim). 2016;95(25):e3951.
35. Dajon M, Iribarren K, Cremer I. Toll-like receptor stimulation in cancer: A pro- and anti-tumor double-edged sword. Immunobiology. 2017;222(1):89–100.
36. Husseinzadeh N, Davenport SM. Role of toll-like receptors in cervical, endometrial and ovarian cancers: a review. Gynecol Oncol. 2014;135(2):359–63.

37. Gao YJ, Liu L, Li S, Yuan GF, Li L, Zhu HY, et al. Down-regulation of CXCL11 inhibits colorectal cancer cell growth and epithelial-mesenchymal transition. Onco Targets Ther. 2018;11:7333–43.
38. Koo YJ, Kim TJ, Min KJ, So KA, Jung US, Hong JH. CXCL11 mediates TWIST1-induced angiogenesis in epithelial ovarian cancer. Tumour Biol. 2017;39(5):1010428317706226.
39. Jin C, Xue Y, Li Y, Bu H, Yu H, Zhang T, et al. A 2-Protein Signature Predicting Clinical Outcome in High-Grade Serous Ovarian Cancer. Int J Gynecol Cancer. 2018;28(1):51–8.
40. Zhao L, Ji G, Le X, Luo Z, Wang C, Feng M, et al. An integrated analysis identifies STAT4 as a key regulator of ovarian cancer metastasis. Oncogene. 2017;36(24):3384–96.
41. Nishi M, Batsaikhan BE, Yoshikawa K, Higashijima J, Tokunaga T, Takasu C, et al. High STAT4 Expression Indicates Better Disease-free Survival in Patients with Gastric Cancer. Anticancer Res. 2017;37(12):6723–9.
42. Wang G, Chen JH, Qiang Y, Wang DZ, Chen Z. Decreased STAT4 indicates poor prognosis and enhanced cell proliferation in hepatocellular carcinoma. World J Gastroenterol. 2015;21(13):3983–93.
43. Wang B, Liu G, Ding L, Zhao J, Lu Y. FOXA2 promotes the proliferation, migration and invasion, and epithelial mesenchymal transition in colon cancer. Exp Ther Med. 2018;16(1):133–40.
44. Salem M, O'Brien JA, Bernaudo S, Shawer H, Ye G, Brkic J, et al. miR-590-3p Promotes Ovarian Cancer Growth and Metastasis via a Novel FOXA2-Versican Pathway. Cancer Res. 2018;78(15):4175–90.

Figures

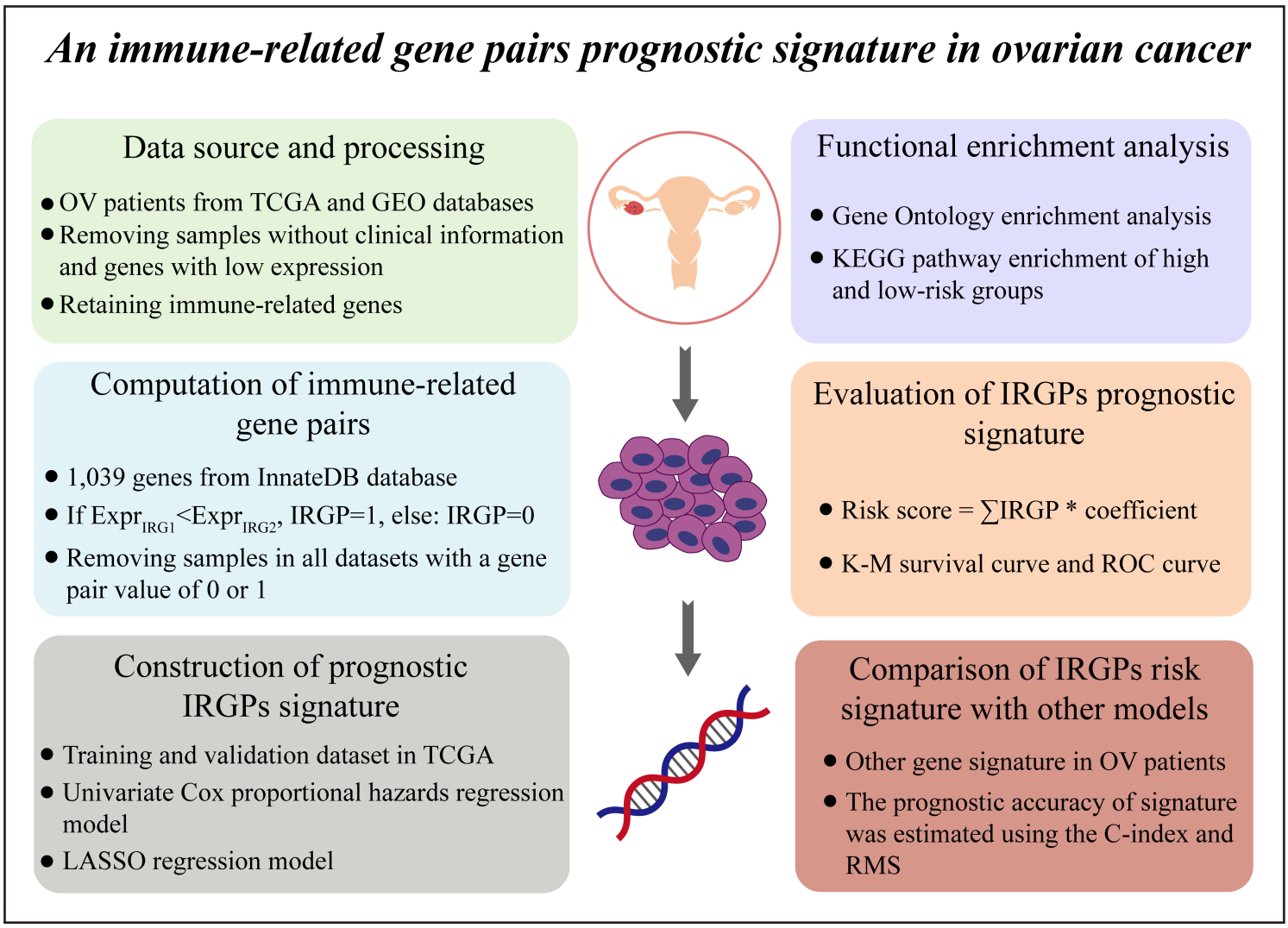

Figure 1

The workflow of this study. Production and verification of a gene pair prognostic signature related to immunity in ovarian cancer.

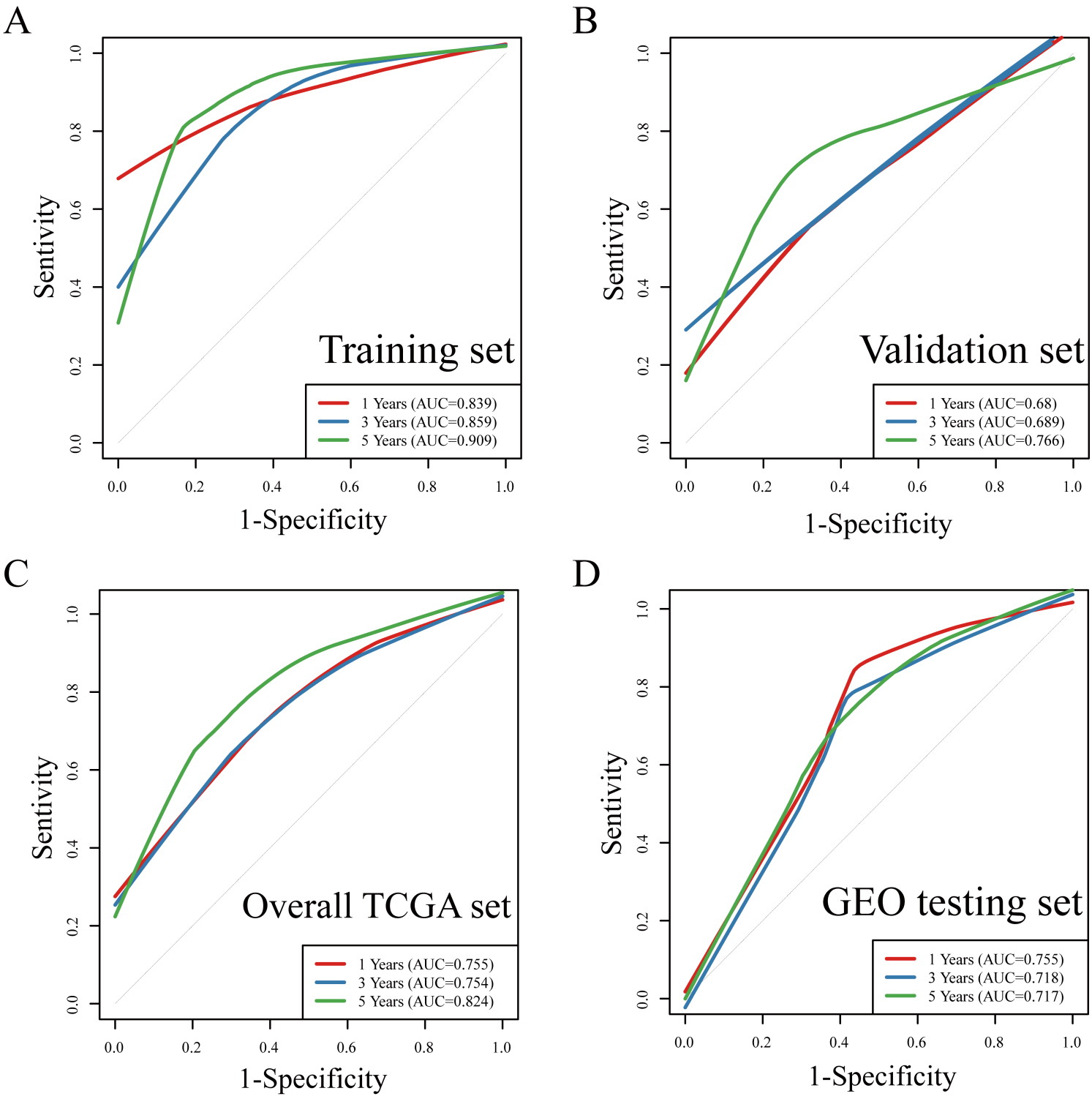

Figure 2

The time-dependent ROC curve of OV patients based on the IRGPs. (A) Training set. (B) Validation set. (C) Overall TCGA set. (D) GEO testing set.

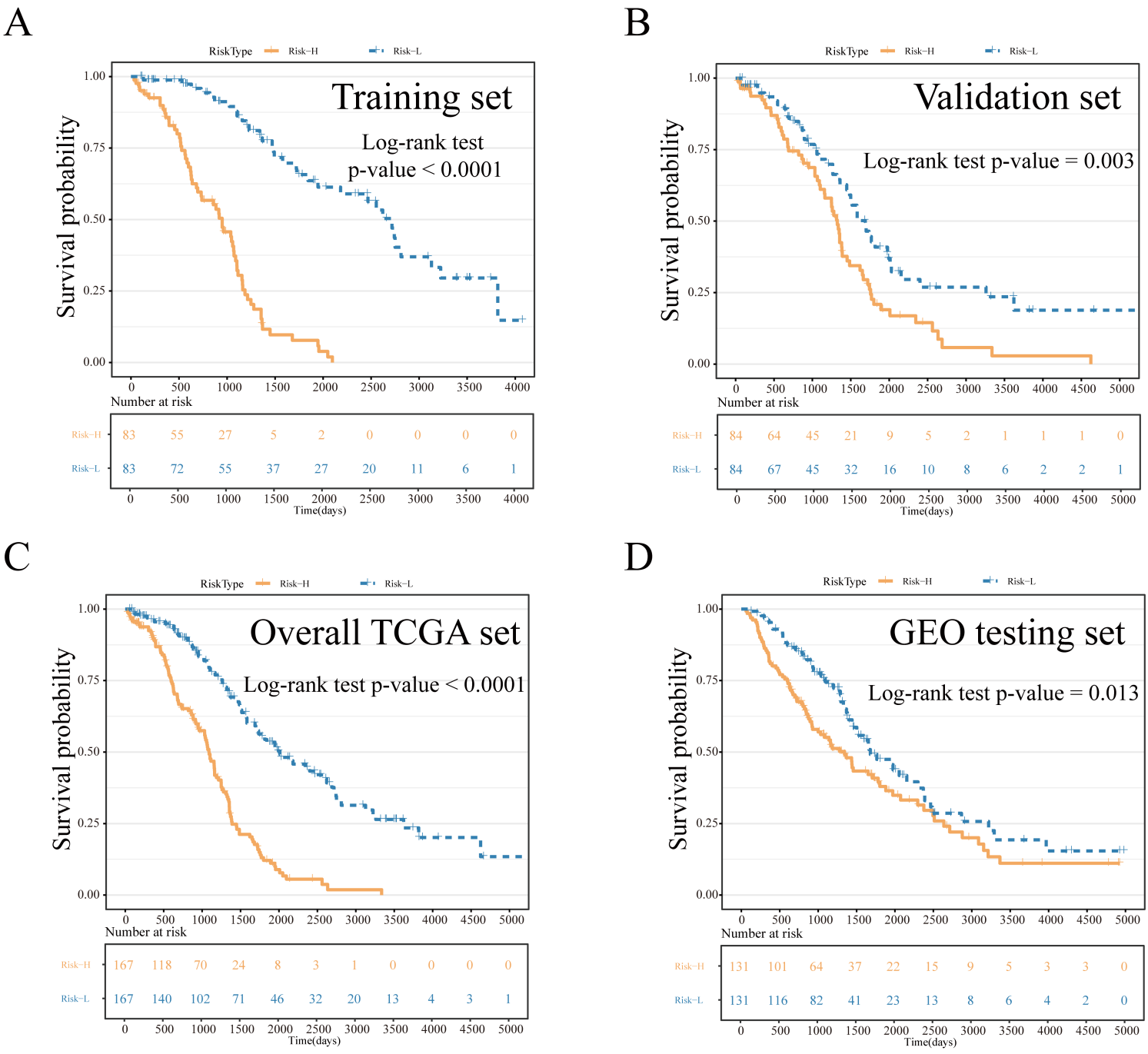

Figure 3

The Kaplan-Meier curves of total survival of various IRGP signature risk groups. OV cases were stratified by median risk scores (Risk-H and Risk-L groups). (A) Training set. (B) Validation set. (C) Overall TCGA set. (D) GEO testing set.

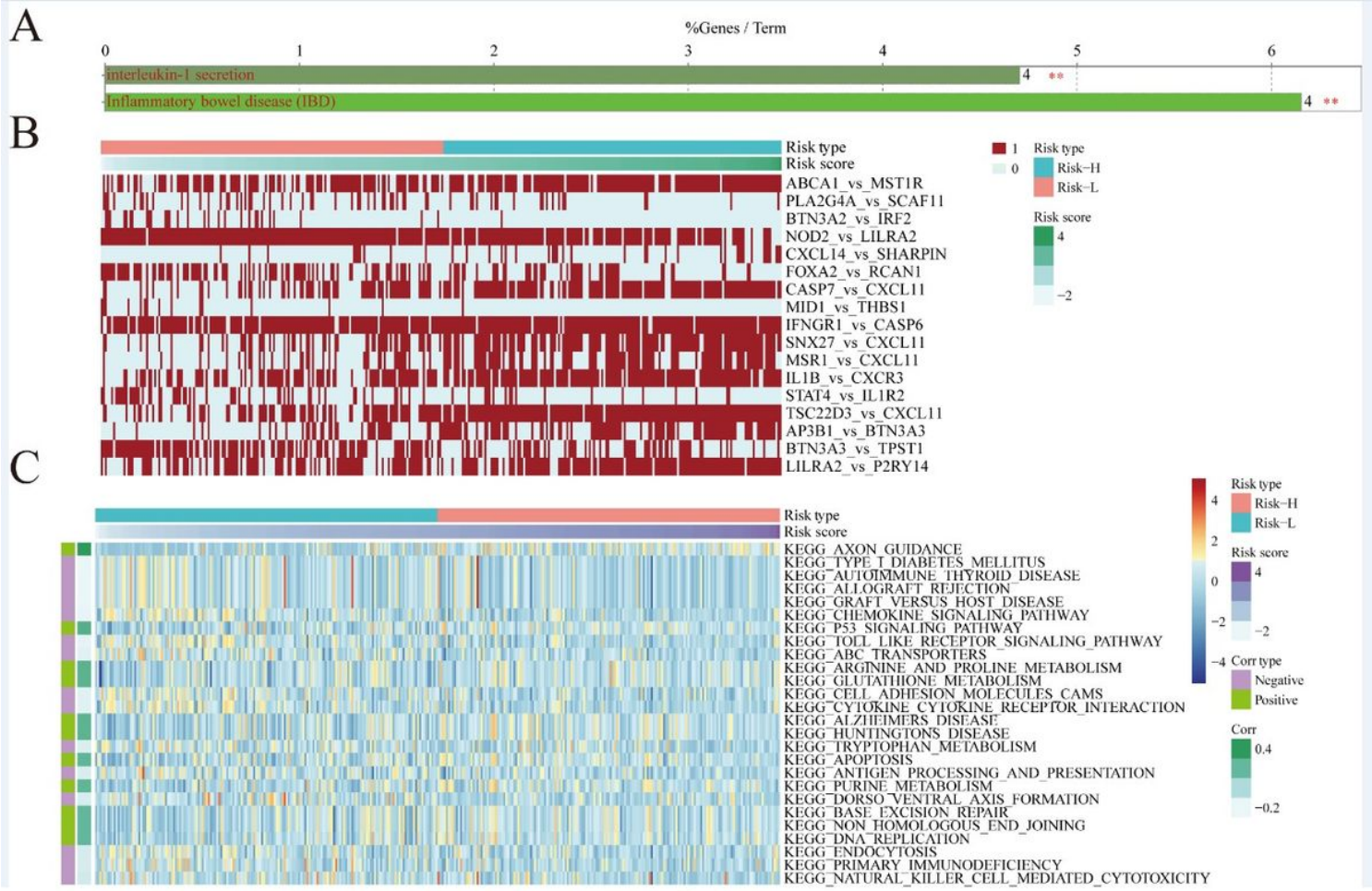

Figure 4

Functional analysis of 29 immune-related genes. (A) Gene family enrichment results for 29 immune-related genes. (B) Relationship between 17 IRGPs values and risk scores. (C) GSVA pathway enrichment results for DEGs in high- and low-risk groups. Corr represents the correlation coefficient between the enrichment scores and the sample risk scores with FDR < 0.05.

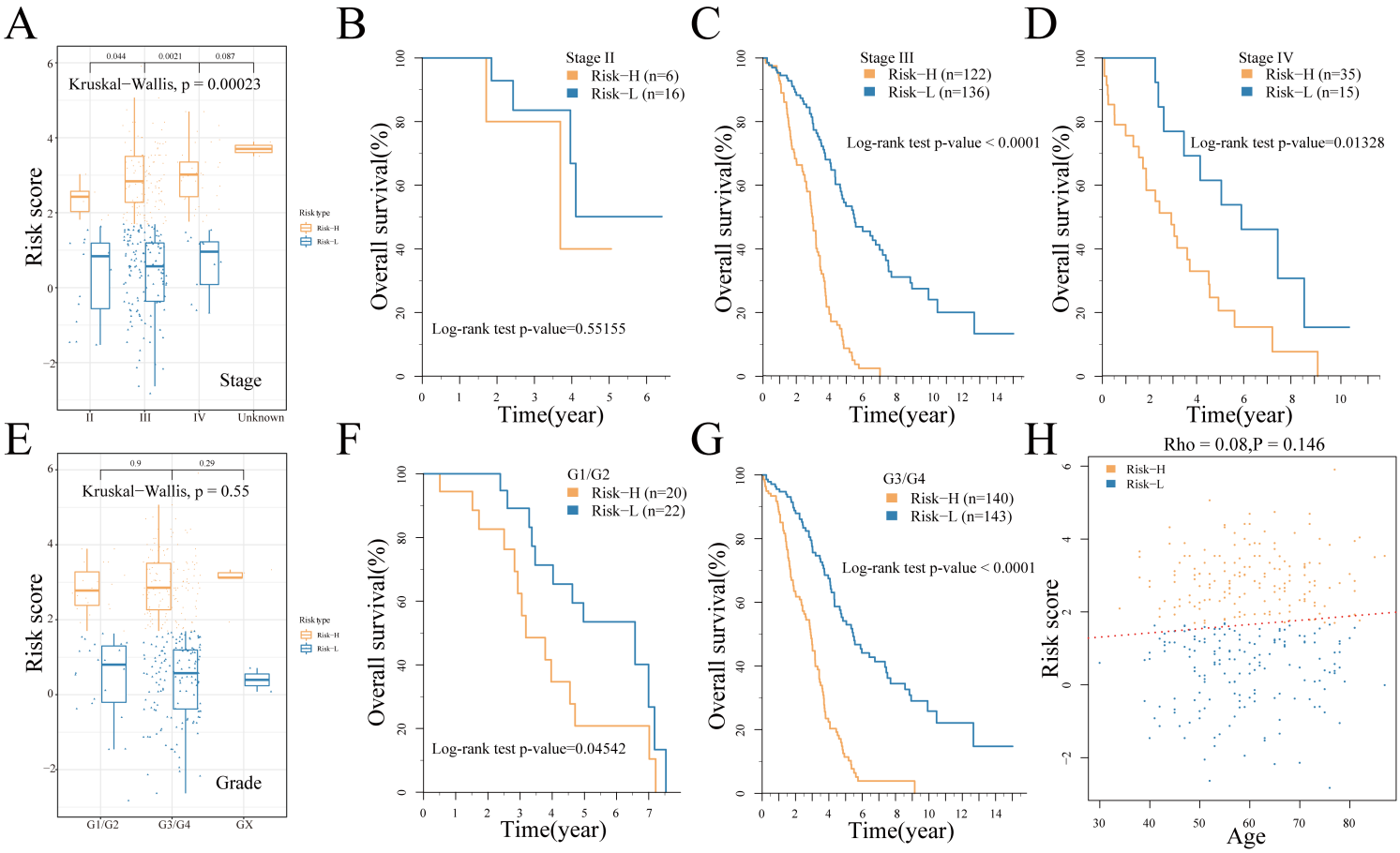

Figure 5

Relationship between the 17-IRGP risk model and clinical characteristics. (A) Stage risk distribution. (B) The K-M curve of stage II of high- and low-risk samples. (C) The K-M curve of stage III high- and low-risk samples. (D) The K-M curve of stage IV of high- and low-risk samples. (E) The grade risk distribution. (F) The K-M curve of G1/G2 of high- and low-risk samples. (G) The K-M curve of G3/G4 of high- and low-risk samples. (H) The correlation between age and risk scores.

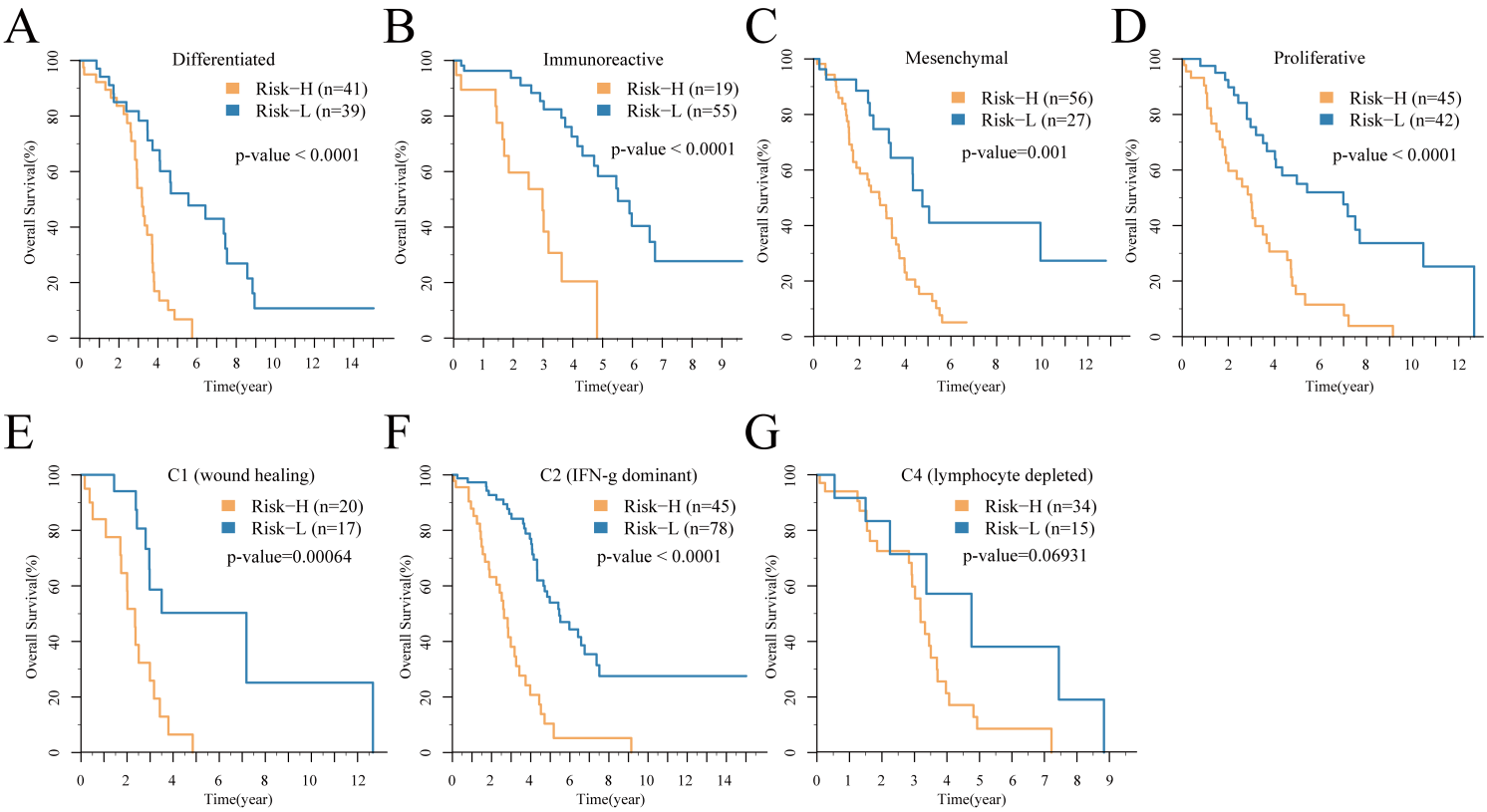

Figure 6

The performance of prognostic risk signature in OV subtypes. (A-D) The K-M curves of the 17-IRGP risk model on differentiated, immunopositive, mesenchymal and proliferative subtypes of the TCGA database. (E-G) The K-M curves of the 17-IRGP risk model on immune subtypes C1, C2 and C4.

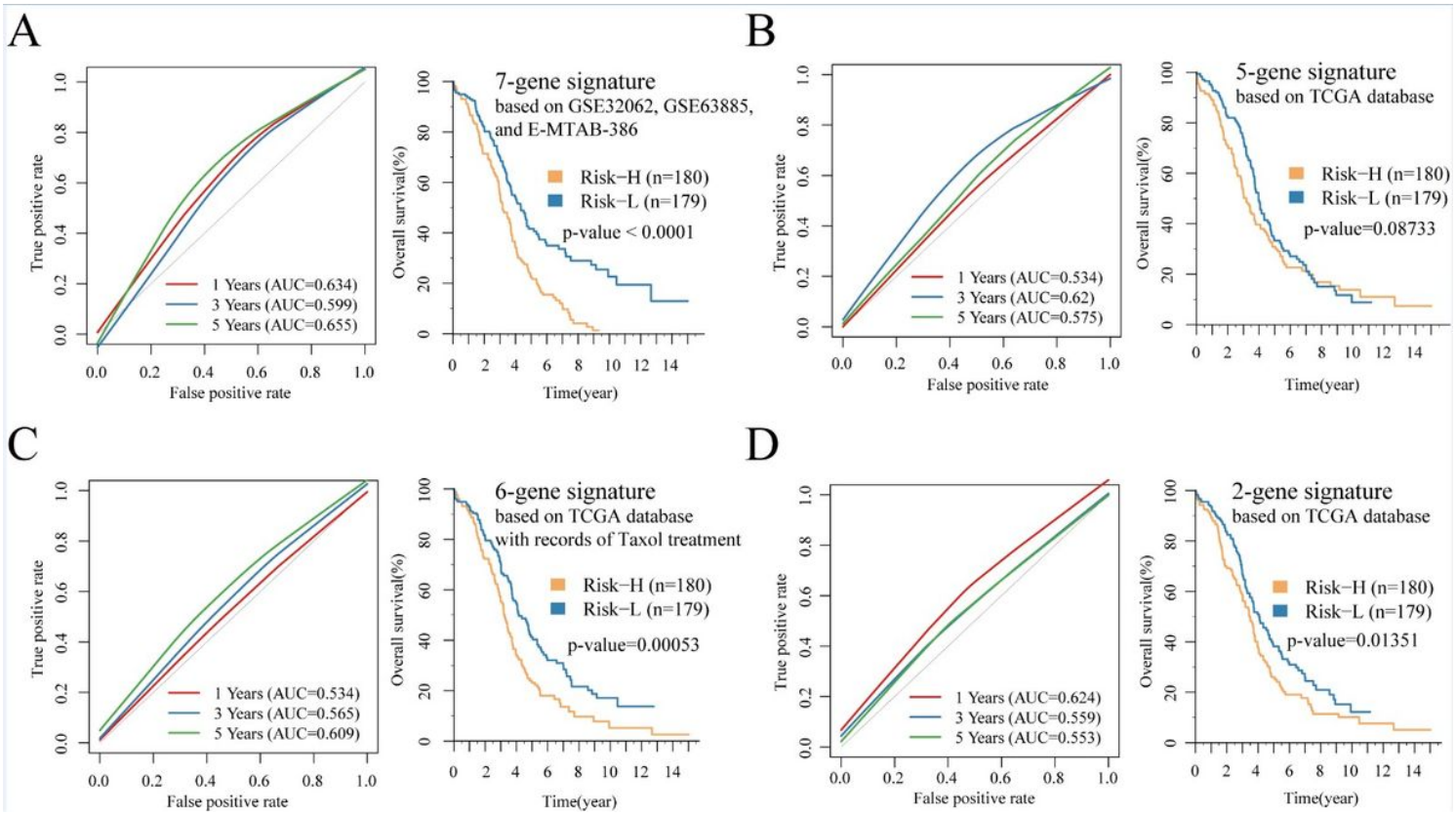

Figure 7

The ROC and K-M curves of the OV prognostic risk model. (A) The ROC and K-M curve of Risk-H/Risk-L samples of the 7-gene signature risk model. (B) The ROC and K-M curve of Risk-H/Risk-L samples of the 5-gene signature risk model. (C) The ROC and K-M curve of Risk-H/Risk-L samples of the 6-gene signature risk model. (D) The ROC and K-M curve of Risk-H/Risk-L samples of the 2-gene signature risk model.

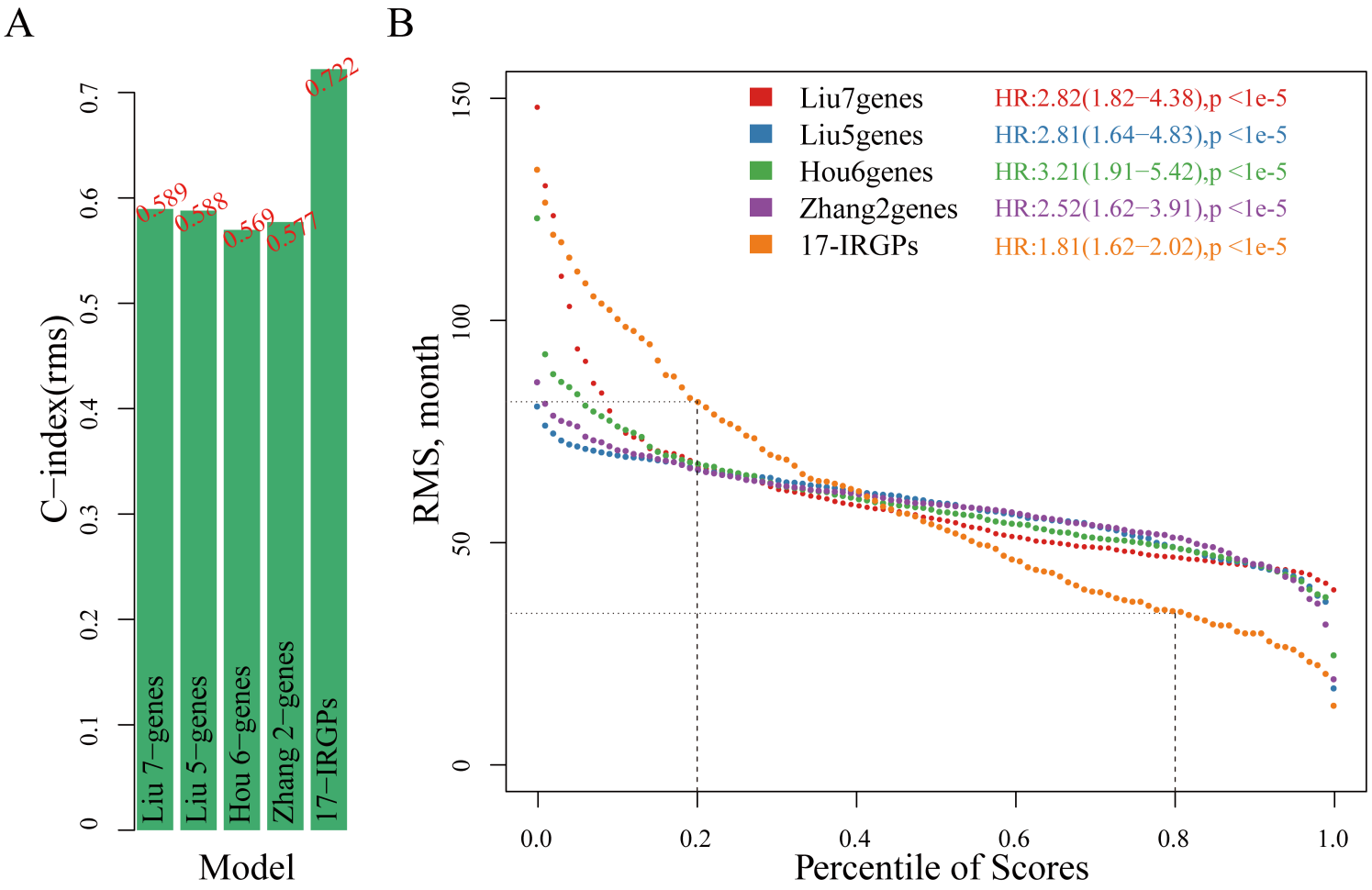

Figure 8

Comparison of OV prognostic risk models. (A) The C-index of 5 prognostic risk models. (B) The RMST curves of 5 prognostic risk models. The dashed line represents the RMS time (months) corresponding to the 20th and 80th percentile scores. Each point represents the corresponding model's RMS time corresponding to the sample risk score.

Supplementary Files

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

- SupplementaryTable3.xlsx
- SupplementaryFigure5.tif

- SupplementaryFigure4.tif
- SupplementaryFigure6.tif
- SupplementaryFigure1.tif
- SupplementaryFigure3.tif
- SupplementaryTable2.xlsx
- SupplementaryTable6.xlsx
- SupplementaryTable7.xlsx
- SupplementaryInformation.docx
- SupplementaryTable4.xlsx
- SupplementaryTable5.xlsx
- SupplementaryFigure2.tif
- SupplementaryTable1.xlsx