

Development and Validation of an Immune-Related Gene-Pair Model of High-Grade Serous Ovarian Cancer After Platinum-Based Chemotherapy

Jiaying Lin

China Medical University First Hospital

Dan Sun

China Medical University First Hospital

Tianren Li (✉ trli@cmu.edu.cn)

China Medical University First Hospital <https://orcid.org/0000-0002-6739-5591>

Primary research

Keywords: HGSOC, Cox model, prognostic marker, immune gene

Posted Date: September 22nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-77353/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Frontiers in Oncology on February 10th, 2021. See the published version at <https://doi.org/10.3389/fonc.2020.626555>.

Abstract

Background:

High-grade serous ovarian cancer (HGSOC) is a common cause of death from gynecological cancer, with an overall survival rate that has not significantly improved in decades. Reliable bio-markers are needed to identify high-risk HGSOC to assist in the selection and development of treatment options.

Method:

The study included ten HGSOC cohorts, which were merged into four separate cohorts including a total of 1526 samples. We used the relative expression of immune genes to construct the gene-pair matrix, and the Least absolute shrinkage and selection operator regression was performed to build the prognosis model using the training set. The prognosis of the model was verified in the training set (363 cases) and three validation sets (of 251, 354, and 558 cases). Finally, the differences in immune cell infiltration and gene enrichment pathways between high and low score groups were identified.

Results:

A prognosis model of HGSOC overall survival rate was constructed in the training set, and included data for 35 immune gene-related gene pairs and the regression coefficients. The risk stratification of HGSOC patients was successfully performed using the training set, with a p-value of Kaplan-Meier of < 0.001 . A score from this model is an independent prognostic factor of HGSOC, and prognosis was evaluated in different clinical subgroups. This model was also successful for the other three validation sets, and the results of Kaplan-Meier analysis were statistically significant. The model can also predict patient progression-free survival with HGSOC to reflect tumor growth status. There were differences in some immune cells between the high-risk and low-risk groups as defined by the model. There was a lower infiltration level of M1 macrophages in the high-risk group compared to that in the low-risk group ($p < 0.001$). Finally, many of the immune-related pathways were enriched in the low-risk group, with antigen processing and presentation identified as the most enriched pathways.

Conclusion:

The prognostic model based on immune-related gene pairs developed is a potential prognostic marker for high-grade serous ovarian cancer treated with platinum. The model has robust prognostic ability and wide applicability. More prospective studies will be needed to assess the practical application of this model for precision therapy.

Background

Ovarian cancer is the second most common cause of gynecological cancer death worldwide [1]. In 2018, approximately 22240 new ovarian cancer cases were diagnosed in the United States, and 41070 patients with ovarian cancer died [2]. High-grade serous ovarian cancer (HGSOC) accounts for about 70–80% of

ovarian cancer deaths, and there has been little improvement in overall survival over the decades [3]. Many patients suffer relapse and chemotherapy resistance [4] after the standard treatment of HGSOC of optimal debonding surgery and systemic cytotoxic platinum-based chemotherapy. Screening and monitoring of ovarian cancer typically focuses on detection of cancer antigen 125 (CA125) and use of the ovarian cancer risk algorithm (ROMA), but these methods are limited and have poor predictive ability [5, 6]. The Federation International of Gynecology and Obstetrics (FIGO) staging system is used in clinical application, and the correlation between ovarian cancer prognosis and FIGO staging is well confirmed [7]. However, the FIGO staging system focuses mainly on clinical features, while ovarian cancer is a disease with high histological and genetic heterogeneity [8]. Therefore, there is an urgent need to develop new molecular markers to identify HGSOC subsets with poor survival and high heterogeneity for better selection of clinical treatment options.

Due to advances in bioinformatics, many studies have proposed gene expression signatures for risk stratification in HGSOC patients [9]. However, many of these studies utilized excessive fitting or lack sufficient verification, so these models remain far from clinical application [10, 11]. There is a significant amount of HGSOC transcriptome data in the public database. However, the heterogeneity between cohorts and the diversity of sequencing platforms have limited use of traditional gene expression data for cross-cohort comparison. Some studies have gene pairs based on ranking of the relative expression levels of two genes, and then construct a prognostic model based on gene pairs. This method eliminates the heterogeneity of different data, and this approach has been successfully verified with reliable results for analysis of many cancers, such as non-small cell lung cancer [12] and colorectal cancer [13]. Still, it has not been reported in ovarian cancer.

Ovarian cancer is an immunogenic tumor. The spontaneous anti-tumor immune response of some patients can prolong survival time, while the immune escape of other patients can shorten survival time [14]. The main obstacle to implementing immunotherapy for ovarian cancer is an immunosuppressive tumor microenvironment [15]. The immune-related biomarkers can predict the response to different types of immunotherapy and promote our understanding of the interactions between molecules and tumor cells in the microenvironment.

In this study, we used multiple transcriptome data sets of HGSOC to develop and verify an individualized prognostic model of HGSOC based on immune-related gene pairs. We used this model to stratify the risk of patients and explore differences in potential immune infiltration and enrichment pathways in high and low-risk patients.

Materials And Methods

Study population and eligibility criteria

The search was performed using keywords of "ovarian cancer" and "survival" in the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) and The Cancer Genome Atlas (TCGA,

<https://portal.gdc.cancer.gov/>) databases. We found a total of 28 cohorts (N = 4115). First, study cohort screening was carried out. Eighteen cohorts were removed (N = 2215, Additional file 1: Table S2) that did not meet the inclusion criteria (treated with platinum; contained HGSOc; contained overall survival data). Ten research cohorts (N = 1900, Additional file 1: Table S1) were left for further analysis: TCGA-OV [16] from TCGA database, and GSE53963 [17], GSE63885 [18], GSE26712 [19], GSE17260 [20], GSE32062 [21], GSE32063 [21], GSE14764 [22], GSE30161 [23], and GSE9891 [24] from GEO database. Through sample screening, 1526 HGSOc samples with complete survival data and transcriptional groups were found (Additional file 1: Table S3).

For the specific data processing and research flow, see Fig. 1.

Cohort Merging

The research cohorts were merged according to the same microarray platform. There are smaller deviations from cohorts of identical microarray platforms. Merge cohorts can reduce data redundancy and increase data size. We used the `combat` function of R package "sva" to remove the batch effect. A total of three combined cohorts were merged out of the nine GEO cohorts (Table 1). The above three GEO combined cohorts were used as the validation cohort and the TCGA-OV cohort was used as the training cohort.

Table 1
The information of combined cohorts.

Dataset-combide	GEO accession	Platform	Sample size
TCGA-OV	—	Affymetrix Human Genome U133A Array	363
GPL96-combined	GSE14764; GSE26712	Affymetrix Human Genome U133A Array/GPL96	251
GPL570-combined	GSE9891; GSE30161; GSE63885	Affymetrix Human Genome U133 Plus 2.0 Array/GPL570	354
GPL6480-combined	GSE17260; GSE32062; GSE32063; GSE53963	Agilent-014850 Whole Human Genome Microarray 4 × 44K G4112F (Probe Name version)/GPL6480	558

Construction Of Immune Gene Pair Matrix

An immune gene set was downloaded from Immport (<https://www.immport.org/>). The immune genes expressed in the TCGA-OV, GPL96-combined, GPL570-combined, and GPL6480-combined cohorts were identified, and the expression value of the genes should be more than 0.5. We defined the combination of two immune-related genes (IRG-1 and IRG-2) as an immune-related gene pair (IRGP). The IRGP score compares the expression levels of IRG-1 and IRG-2 for a given sample. When the expression of IRG-1 is greater than that of IRG-2, the IRGP score is 1; otherwise, the IRGP score is 0. Thus, the immune gene pair

matrix is made of 0's and 1's. Each IRGP must meet the cohort's standard that the proportions of 0's and 1's must be greater than 20% so that the IRGP is meaningful for subsequent analysis.

Cox Model Construction

Lasso (Least absolute shrinkage and selection operator) is a statistical method to reduce data dimensionality. Lasso selects the variables of the sample data based on a penalty method. By compressing the original coefficients, insignificant variables whose coefficients become zero are discarded, and any collinear variables are removed. Finally, a simplified model is obtained. We used the R package "glmnet" and "survival" to perform the lasso regression operation [25] and construct the IRGP index (IRGPI) model. First, we used the function "glmnet" for 1000 random simulations to build the model and obtain the correlation between the coef (regression coefficient) and lambda (punishment coefficient). Then the function "cv.glmnet" is used for random simulation 10-fold cross-validation (CV) 1000 times. 10-fold CV divides data into ten equal parts. One takes one as a training set and uses other parts to validate models. Deviations obtained by CV can be used to evaluate models, and smaller deviations indicate better models. The model can be expressed as: $IRGPI = \sum n_i (IRGP_i \cdot coef_i)$ (n is the number of IRGP, $IRGP_i$ is the score of the i th IRGP, and $coef_i$ is the regression coefficient of the i th IRGP).

Model Verification

We used the Kaplan-Meier method to describe the occurrence of survival outcomes. To do this, we used the R packages "survminer" and "survival" to compare the prognosis for different groups. The function "res.cat" was performed to obtain the best truncation value in continuous variables and then group the samples, requiring each group size to be less than 20% of the total group. Using this method for Kaplan-Meier analysis, the p-value is the smallest. The R package "survival" was then used to perform univariate and multivariate Cox regression analysis to observe the model's predictive ability and clinical factors. The R package "survivalROC" was used to draw receiver operating characteristic (ROC) curves and calculate the area under the curve (AUC) values. AUC values greater than 0.5, indicated that the factor can be used as an indicator of prognosis, with values closer to 1 indicating higher accuracy of prognosis.

Calculation of the infiltration level of 22 kinds of immune cells

CIBERSORT ("Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts") is a complex tissue deconvolution method based on linear support vector regression of gene expression profiles [26]. We used the R package "CIBERSORT" and the "Leukocyte signature matrix" to obtain the composition of 22 kinds of immune cells in each sample, and only the samples with $p < 0.05$ were used for the analysis.

Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) is a computational method used to determine if a predefined set of genes show differences between two biological states [27]. According to the risk scores of samples, the samples can be divided into two groups, and the enrichment difference of the two groups can be explored using "c2.cp.kegg.v7.0. Symbols". P -value < 0.05 and q -value < 0.05 indicate that the enriched items are statistically significant. We used the R package "ggplot2" to show the enrichment items in a bubble chart. R packages "ggplot2" and "clusterProfiler" were used to enhance the appearance of the enrichment plot.

Statistical analysis

All statistical analysis used in this study is based on R Programming Language software (Rx64 3.5.1). We used the R package "pheatmap" to visualize score distribution, the R package "fmsb" to draw radar maps, and the online website "Sangerbox" (<http://sangerbox.com/>) to draw Venn diagrams. Wilcoxon test was used to compare between the two groups, and a value of $P < 0.05$ was considered statistically significant. The online website "Metascape" was used to determine the correlated gene set pathways and perform enrichment analysis [28].

Results

Construction of Cox model based on IRGP

Data for a total of 1526 HGSOC samples that received platinum-based treatment were obtained from the TCGA-OV, GPL96-combined, GPL570-combined, and GPL6480-combined cohorts. TCGA-OV was used as the training set and GPL96-combined, GPL570-combined, and GPL6480-combined cohorts were used as the validation set. We obtained 238 common immune genes from the intersection of the four cohort immune genes (Fig. 2a). An IRGP analysis was performed by pairwise construction of 238 genes in each cohort. The IRGP that meets the requirements of the four cohorts was then intersected to obtain 2672 common IRGPs (Fig. 2b). The relationship between the 2672 IRGPs and overall survival rates was evaluated using the TCGA-OV cohort, and the results indicated that 141 IRGPs were associated with the prognosis of HGSOC ($p < 0.01$). The lambda and coef diagrams of IRGPs (Fig. 2c) were constructed using the lasso algorithm. As the lambda value increases, the coefficients of some IRGPs become zero, which means that the scores of these IRGPs do not affect the model. We then used 10-fold CV to calculate the partial likelihood deviance of the model (Fig. 2d). The deviance was smallest when 35 genes were used. The coef of each IRGP was obtained according to the corresponding lambda value (Fig. 2e). The 35 IRGPs and the corresponding coef values constitute the IRGPI prognostic model (Additional file 1: Table S4). The IRGPI values were calculated for each sample in the four cohorts. These IRGPs include 52 genes. Through pathway and process enrichment analysis, we found that these genes are mainly concentrated in "cell chemotaxis," "cytokine-mediated signaling pathway," and "positive regulation of cell migration" (Additional file 1: Table S5).

Verification of the prognostic ability of IRGPI of overall survival rate in the training set

IRGPI significantly stratified HGSOC patients in TCGA-OV, with poor prognosis and high risk for high IRGPI ($p < 0.001$, Fig. 2f). The AUC values were 0.637, 0.711, 0.766, 0.826, and 0.897 for 1, 3, 5, 7, and 9 years, respectively (Fig. 2g). Univariate Cox analysis showed that age, stage, and IRGPI had predictive ability ($p < 0.05$). Multivariate Cox analysis showed that age and IRGPI could be used as independent prognostic factors ($p < 0.001$). IRGPI also effectively stratified the risk of HGSOC patients under different clinical types (Age ≤ 60 ; age > 60 ; grade 2; grade 3; stage 3; stage 4; $p < 0.001$, Fig. 2j). Overall, IRGPI can be used to independently evaluate the overall survival rate of patients with HGSOC.

Prognostic ability of the overall survival rate of IRGPI in the validation sets

IRGPI can be used as a prognostic factor for the GPL96-combined, GPL570-combined, and GPL6480-combined cohorts. In the GPL96-combined cohort, the results of Kaplan-Meier analysis showed worse prognosis of patients with the high-IRGPI group compared to the low-IRGPI group ($p = 0.001$, Fig. 3a); The IRGPI distribution and patient survival status plots are shown in Fig. 3b; The AUC values for 3, 5, and 7 years were 0.651, 0.637, and 0.569, for the GPL96-combined, GPL570-combined, and GPL6480-combined cohorts, respectively (Fig. 3c). Similarly, in the GPL570-combined cohort, Kaplan-Meier analysis showed worse prognosis of patients for the high-IRGPI group compared to that of the low-IRGPI group ($p = 0.002$, Fig. 3d); Fig. 3e shows that IRGPI distribution and patient survival status plots; and the AUC values for 3, 5, and 7 years were 0.574, 0.560, and 0.641, respectively (Fig. 3f). Finally, in the GPL6480-combined cohort, Kaplan-Meier analysis showed worse prognosis of patients for the high-IRGPI group compared to the low-IRGPI group ($p < 0.001$, Fig. 3g). Figure 3h shows the IRGPI distribution and patient survival status plots, and the AUC values for the three, five, and seven years were 0.609, 0.620, and 0.630, respectively (Fig. 3i). Overall, IRGPI successfully stratified the risk of HGSOC patients for the three validation sets.

IRGPI can also predict progression-free survival (PFS) probability in patients with HGSOC

Including only samples with PFS information, 1011 cases of HGSOC were further analyzed, including samples from the TCGA-OV ($N = 363$), GPL570-combined ($N = 277$), and GPL6480-combined cohorts ($N = 371$). Kaplan-Meier analysis showed that IRGPI allowed risk stratification of HGSOC in TCGA-OV ($p < 0.001$, Fig. 4a), GPL570-combined ($p = 0.007$, Fig. 4b), and GPL6480-combined cohorts ($p < 0.001$, Fig. 4c). Low PFS correlated with high IRGPI.

High-irgpi Group Showed Low Macrophages M1 Infiltration

We analyzed the differences in the infiltration levels of immune cells between the high-IRGPI and low-IRGPI groups. The radar map presented in Fig. 5a shows a significant difference in the level of infiltration in the TCGA-OV cohort for M1 macrophages, gamma delta T cells, and follicular helper T cells ($p < 0.001$). In the GPL96-combined cohort, significant differences were seen for M1 macrophages, CD4 memory resting T cells, monocytes, and resting dendritic cells ($p < 0.001$, Fig. 5b); in the GPL570-combined cohort, differences were seen for M1 macrophages and CD4 memory resting T cells ($p < 0.001$, Fig. 5c); in the

GPL6480-combined cohort, differences were seen for M1 macrophages, activated mast cells, plasma cells, monocytes, CD4 memory activated T cells, CD4 memory resting T cells, and follicular helper T cells ($p < 0.001$, Fig. 5d). Among these immune cells, only M1 macrophages showed significant differences in high-IRGPI and low-IRGPI groups for all four cohorts. M1 macrophages exhibited higher infiltration level in the low-IRGPI group than that in the high-IRGPI group ($p < 0.001$, Fig. 5e).

Gene Set Enrichment Analysis Based On Irgpi

IRGPI was applied to divide the samples into high-IRGPI and low-IRGPI groups to analyze differences in the enriched GSEA KEGG pathways between the two groups. The results showed few enrichment pathways in the high-IRGPI group, and no pathway is enriched in four cohorts at the same time (Fig. 6a). In contrast, in the low-IRGPI group, many pathways were enriched, and ten pathways were enriched in all four cohorts (Fig. 6b). The bubble charts presented in Fig. 6c-f show many of these enriched pathways are immune-related signal pathways. Among the four cohorts, the “Antigen processing and presentation” was the most highly enriched (Fig. 6g-j). This result suggests that the better prognosis of the low-IRGPI patients is related to the activity of this pathway.

Discussion

Epithelial ovarian cancer is usually detected in advanced stages and has a poor prognosis, making it the second common cause of death from gynecological cancer [29]. HGSOC is complex, with varying biological and molecular characteristics, so the realization of personalized precision medicine is the biggest challenge for treatment of HGSOC [1]. Reliable prognostic bio-markers are needed to stratify the risk of HGSOC patients to facilitate decisions about treatment options such as chemotherapy, targeted drugs, or immunotherapy. In this study, we screened data from HGSOC patients treated with platinum therapy. We developed an overall survival and prognostic model of HGSOC based on 35 immune-related gene pairs. The model exhibited good predictive ability for four integrated data sets ($n = 1526$). We also detected an obvious activity of M1 macrophages and enrichment of the “antigen processing and presentation” KEGG pathway in the low-risk group.

The IRGPI model was constructed using 35 IRGPs, a model that can stratify HGSOC patient risk based on platinum therapy. We did a comprehensive search of HGSOC transcriptome data from the TCGA and GEO databases. We included a total of 10 cohorts that met the requirements ($n = 1526$), thus this analysis of HGSOC patients is the largest (highest number of samples) to date, increasing the reliability of the conclusions. To prevent data redundancy, we merged the cohorts into four datasets. Most traditional prognostic models are based on gene expression, but this method is unstable for application to a cross-cohort platform. To prevent the technical deviation of genes due to different experimental platforms, we used the relative expression of genes as a unit. This method compares different genes in a single sample, so there is no need to standardize the data. This characteristic would facilitate the application of IRGPI to practical clinical use. The TCGA-OV cohort was used to construct IRGPI, and the OS of patients was

estimated successfully in four cohorts. The results show that the IRGPI model has robust prediction ability and wide adaptability. Additionally, the IRGPI model can estimate the PFS of patients with HGSOC (n = 1011) and individually predict disease progression.

Many tumor immune-related biomarkers have been developed, which are potential immune targets that can help guide the proper treatment of patients. This study is the first comprehensive analysis of the immune gene prognosis of HGSOC. Most genes used to construct gene pairs are cytokines and cytokine receptors, which act in cell chemotaxis, angiogenesis, and tumor escape [30, 31]. We compared the infiltration level of immune cells between the high-IRGPI and low-IRGPI groups. M1 macrophages showed the most obvious difference, with obvious infiltration inhibition in the high-IRGPI group. Macrophages are important immune cells involved in inflammation and tumorigenesis [32]. Macrophages that infiltrate around tumor cells are called tumor-associated macrophages (TAM). TAM include anti-tumor M1 macrophages and tumor-promoting M2 macrophages [33]. Studies have shown that a high M1 / M2 ratio in ovarian tumor tissue is associated with prolonged survival [34]. M2 macrophages can release immunosuppressive factors to support immune escape of ovarian cancer [35]. Paclitaxel is used to treat ovarian cancer by polarizing M2 into M1 macrophages in a TLR4-dependent manner [36]. Our study shows significant inhibition of M1 macrophages in high-risk patients, suggesting that paclitaxel treatment for HGSOC in high-IRGPI may improve patient survival. In the enrichment analysis, the low-IRGPI group showed enrichment of immune-related pathways, indicating that when the immune pathway is active, the risk of HGSOC is low. The high-IRGPI group may experience immunosuppression, so the prognosis is poor. To summarize, the imbalance of immune function may explain the difference in survival among the patient groups defined by IRGPI.

Although we have used multiple HGSOC cohorts to verify our model with good results, there are limitations of our conclusions. This study is based on large-scale cohorts in a network database and the findings were not verified by additional data. We used the relative expression of genes in a single sample to build the model. Although this method can effectively reduce the sample's batch processing effect, there may be some complex internal and external interference factors that affect the results. The functions of the genes that make up the identified gene pair with high prognostic ability have not been investigated. Overall, more biological experiments are needed to verify the functional mechanism of ovarian cancer.

Conclusions

The immune gene pair-based model developed in this study is a good prognostic indicator for high-grade serous ovarian cancer. This model is suitable for individualized selection of medical treatment options to improve patient survival time and quality. Additionally, prospective studies are needed to verify the accuracy of the model and evaluate the clinical efficacy of the model for individualized treatment.

Abbreviations

HGSOC: High-grade serous ovarian cancer; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; AUC: Area Under the Curve; ROC: Receiver Operating Characteristic; IRGP: Immune-related gene pair; IRGPI: IRGP index; Lasso: Least absolute shrinkage and selection operator; coef: Regression coefficient; GSEA: Gene set enrichment analysis; PFS: Progression-free survival; OS: Overall survival.

Declarations

Acknowledgements

We thank TCGA and GEO for free use

Authors' contributions

TL provided ideas and solutions. JL for data analysis and drafted preparation. DS downloaded and preprocessed data. Both authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The TCGA-OV dataset can be gained from TCGA database (<https://cancergenome.nih.gov/>). Nine GEO datasets (GSE53963, GSE63885, GSE26712, GSE17260, GSE32062, GSE32063, GSE14764, GSE30161, and GSE9891) can be gained from GEO database (<https://www.ncbi.nlm.nih.gov/geo/>).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Urology, The First Hospital of China Medical University, Shenyang 110001, Liaoning, China

²Department of Gynaecology, The First Hospital of China Medical University, Shenyang 110001, Liaoning, China

References

1. Lheureux S, Braunstein M. Am Oza: Epithelial ovarian cancer: Evolution of management in the era of precision medicine. *CA Cancer J Clin.* 2019;69(4):280–304.
2. La Torre B, Trabert C, Desantis Kd, Miller G, Samimi Cd, Runowicz, et al. Ovarian cancer statistics, 2018. *CA Cancer J Clin.* 2018;68(4):284–96.
3. Dd Bowtell S, Böhm A, Ahmed P, Aspuria R, Bast VB, et al. Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat Rev Cancer.* 2015;15(11):668–79.
4. Liu J. Ua Matulonis: New strategies in ovarian cancer: translating the molecular complexity of ovarian cancer into treatment advances. *Clinical cancer research: an official journal of the American Association for Cancer Research.* 2014;20(20):5150–6.
5. Ak Karam B. Karlan: Ovarian cancer: the duplicity of CA125 measurement. *Nat Rev Clin Oncol.* 2010;7(6):335–9.
6. Ma Karlsen N, Sandhu C, Høgdall I, Christensen L, Nedergaard L, Lundvall et al. Evaluation of HE4, CA125, risk of ovarian malignancy algorithm (ROMA) and risk of malignancy index (RMI) as diagnostic tools of epithelial ovarian cancer in patients with a pelvic mass. *Gynecol Oncol.* 2012;127(2):379–83.
7. Narod S. Can advanced-stage ovarian cancer be cured? *Nat Rev Clin Oncol.* 2016;13(4):255–61.
8. An Karnezis K, Cho C, Gilks. Cl Pearce, Dg Huntsman: **The disparate origins of ovarian cancers: pathogenesis and prevention strategies.** *Nat Rev Cancer.* 2017;17(1):65–74.
9. Wang J, Dean D, Hornicek F, Shi H, Duan Z. RNA sequencing (RNA-Seq) and its application in ovarian cancer. *Gynecol Oncol.* 2019;152(1):194–201.
10. Liu J, Xu F, Cheng W, Gao L. Identification and verification of a ten-gene signature predicting overall survival for ovarian cancer. *Exp Cell Res.* 2020;395(2):112235.
11. Zheng M, Li X, Hu Y, Dong H, Gou R, X Nie et al. Identification of molecular marker associated with ovarian cancer prognosis using bioinformatics analysis and experiments. *J Cell Physiol.* 2019;234(7):11023–36.
12. 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 oncology.* 2017;3(11):1529–37.
13. Wang K, Song K, Ma Z, Yao Y, Liu C, J Yang et al. Identification of EMT-related high-risk stage II colorectal cancer and characterisation of metastasis-related genes. *Br J Cancer.* 2020;123(3):410–7.
14. Le Kandalaf D, Powell N, Singh G, Coukos. Immunotherapy for ovarian cancer: what's next? *Journal of clinical oncology: official journal of the American Society of Clinical Oncology.* 2011;29(7):925–33.
15. Odunsi K. Immunotherapy in ovarian cancer. *Annals of oncology: official journal of the European Society for Medical Oncology.* 2017;28:viii1–7.
16. **Integrated genomic analyses of ovarian carcinoma.** *Nature* 2011, 474(7353):609–615.

17. Ge Konecny C, Wang H, Hamidi B, Winterhoff K, Kalli J, Dering et al: **Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer.** *J Natl Cancer Inst* 2014, 106(10).
18. Km Lisowska M, Olbryt V, Dudaladava J, Pamuła-Piłat K, Kujawa E, Grzybowska, et al. Gene expression analysis in ovarian cancer - faults and hints from DNA microarray study. *Frontiers in oncology.* 2014;4:6.
19. Bonome T, Levine D, Shih J, Randonovich M, Pise-Masison C, Bogomolnii F, et al. A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. *Cancer Res.* 2008;68(13):5478–86.
20. Yoshihara K, Tajima A, Yahata T, Kodama S, Fujiwara H, Suzuki M, et al. Gene expression profile for predicting survival in advanced-stage serous ovarian cancer across two independent datasets. *PLoS ONE.* 2010;5(3):e9615.
21. Yoshihara K, Tsunoda T, Shigemizu D, Fujiwara H, Hatae M, H Fujiwara et al. High-risk ovarian cancer based on 126-gene expression signature is uniquely characterized by downregulation of antigen presentation pathway. *Clinical cancer research: an official journal of the American Association for Cancer Research.* 2012;18(5):1374–85.
22. Denkert C, Budczies J, Darb-Esfahani S, Györfy B, Sehouli J, D Könsgen et al. A prognostic gene expression index in ovarian cancer - validation across different independent data sets. *J Pathol.* 2009;218(2):273–80.
23. Dugalić D, Colović R. M Savić: [Duodenal obstruction caused by gallstones (Bouveret syndrome)]. *Acta Chir Iugosl.* 1990;37(1):75–82.
24. Tothill R, Tinker Av, George J, Brown R, Fox S, S Lade et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clinical cancer research: an official journal of the American Association for Cancer Research.* 2008;14(16):5198–208.
25. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software.* 2010;33(1):1–22.
26. Am Newman C, Liu M, Green A, Gentles W, Feng Y, Xu, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Meth.* 2015;12(5):453–7.
27. Subramanian A, Tamayo P, Mootha V, Mukherjee S, Ebert B, Gillette M, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–50.
28. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi Ah, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* 2019;10(1):1523.
29. Lheureux S, Gourley C, Vergote I. Am Oza: Epithelial ovarian cancer. *Lancet.* 2019;393(10177):1240–53.
30. Ww Lin M, Karin. A cytokine-mediated link between innate immunity, inflammation, and cancer. *J Clin Investig.* 2007;117(5):1175–83.

31. Ps Hegde D, Chen. Top 10 Challenges in Cancer Immunotherapy. *Immunity*. 2020;52(1):17–35.
32. Si Grivennikov F, Greten MK. Immunity, inflammation, and cancer. *Cell*. 2010;140(6):883–99.
33. Sica A. Role of tumour-associated macrophages in cancer-related inflammation. *Exp Oncol*. 2010;32(3):153–8.
34. Zhang M, He Y, Sun X, Li Q, Wang W, A Zhao et al. A high M1/M2 ratio of tumor-associated macrophages is associated with extended survival in ovarian cancer patients. *J Ovarian Res*. 2014;7:19.
35. Cheng H, Wang Z, Fu L, Xu T. Macrophage Polarization in the Development and Progression of Ovarian Cancers: An Overview. *Frontiers in oncology*. 2019;9:421.
36. Rolny C, Mazzone M, Tugues S, Laoui D, Johansson I, C Coulon et al. HRG inhibits tumor growth and metastasis by inducing macrophage polarization and vessel normalization through downregulation of PIGF. *Cancer Cell*. 2011;19(1):31–44.

Figures

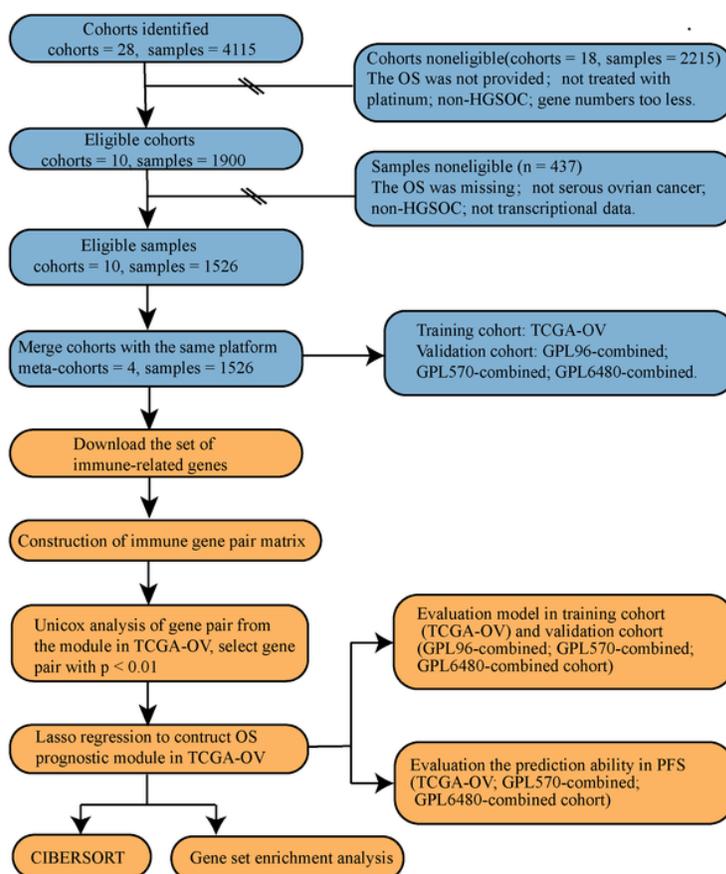


Figure 1

Data processing and research process.

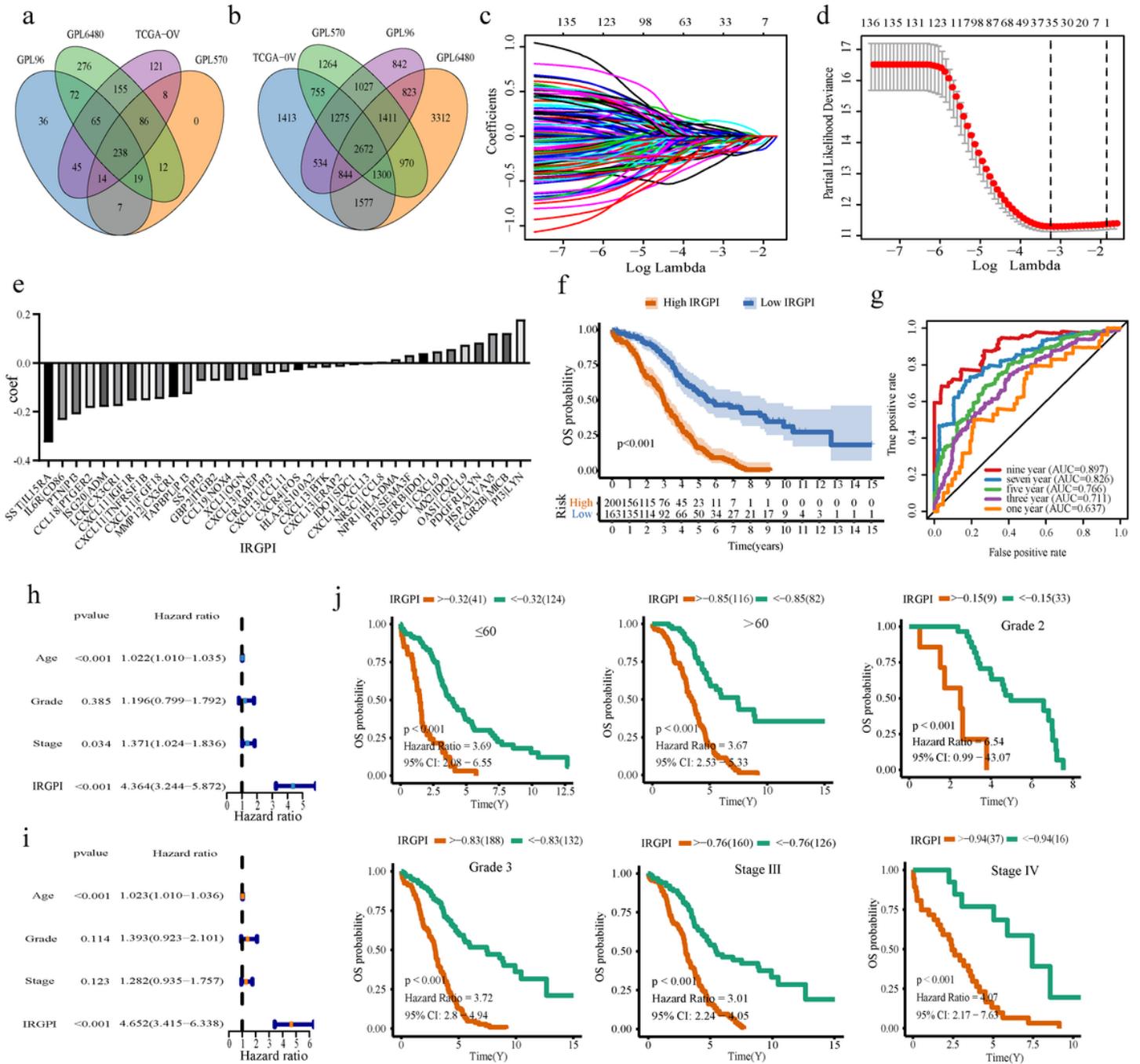


Figure 2

Build and verify IRGPI in the training set. a Venn map of immune genes. b Venn diagrams of immune gene pairs. c The diagram of lambda and coef. d Partial likelihood deviance corresponding to different models. e The coef value corresponding to each gene. f IRGPI's Kaplan-Meier curve of overall survival in the TCGA-OV cohort. g The ROC curve of IRGPI in TCGA-OV cohort. h Forest map based on single factor cox analysis. i Forest map based on multivariate cox analysis. j Kaplan-Meier analyzed IRGPI under different clinical types. $P < 0.05$ indicates that it is statistically significant.

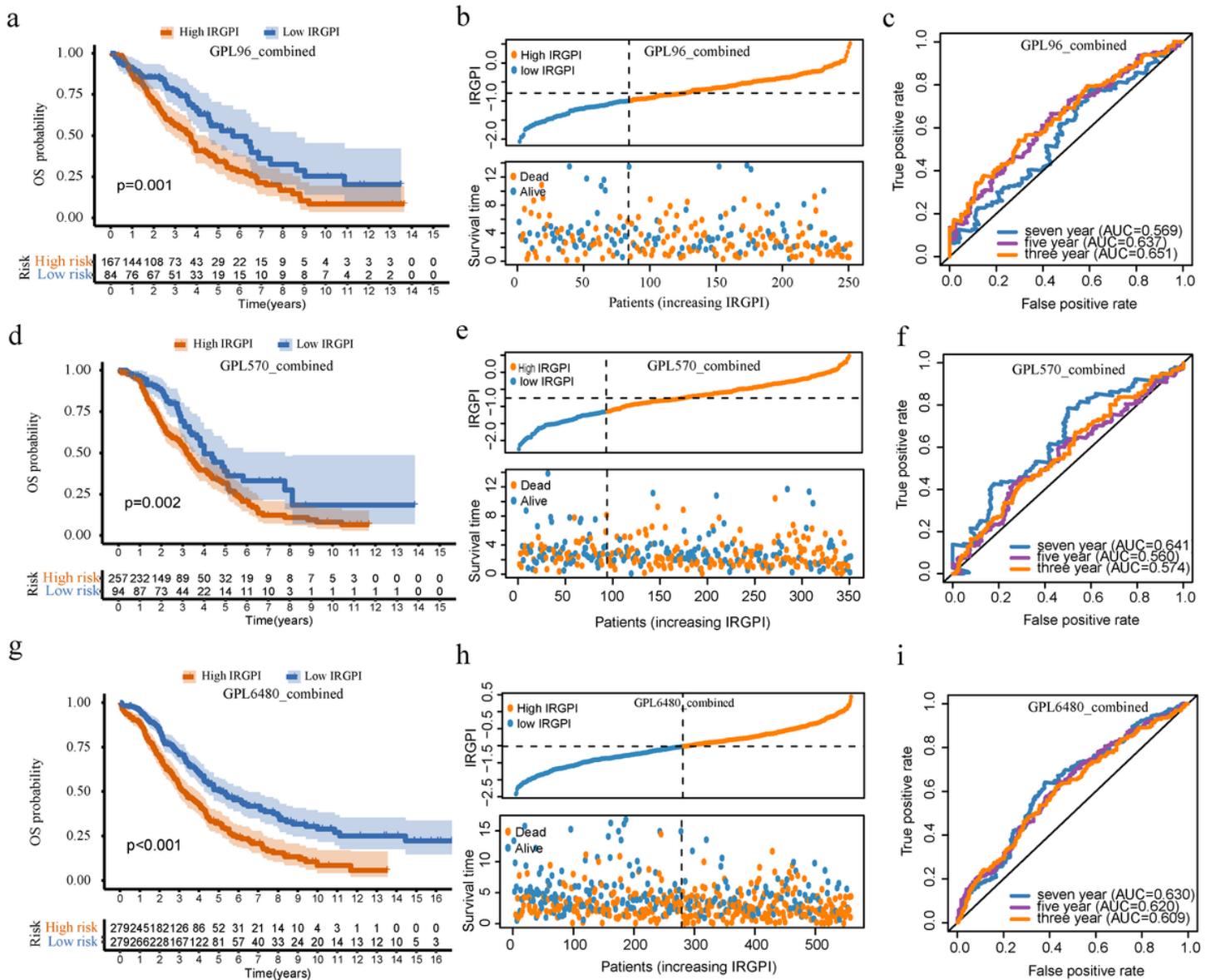


Figure 3

The prognostic ability of IRGPI was verified using the verification set. For the GPL96-combined cohort, the Kaplan-Meier curve (a), patient distribution (b), and ROC curve (c) of IRGPI. For the GPL570-combined cohort, the Kaplan-Meier curve (d), patient distribution (e), and ROC curve (f) of IRGPI. For the GPL6480-combined cohort, the Kaplan-Meier curve (g), patient distribution (h), and ROC curve (i) of IRGPI. $P < 0.05$ indicates that it is statistically significant.

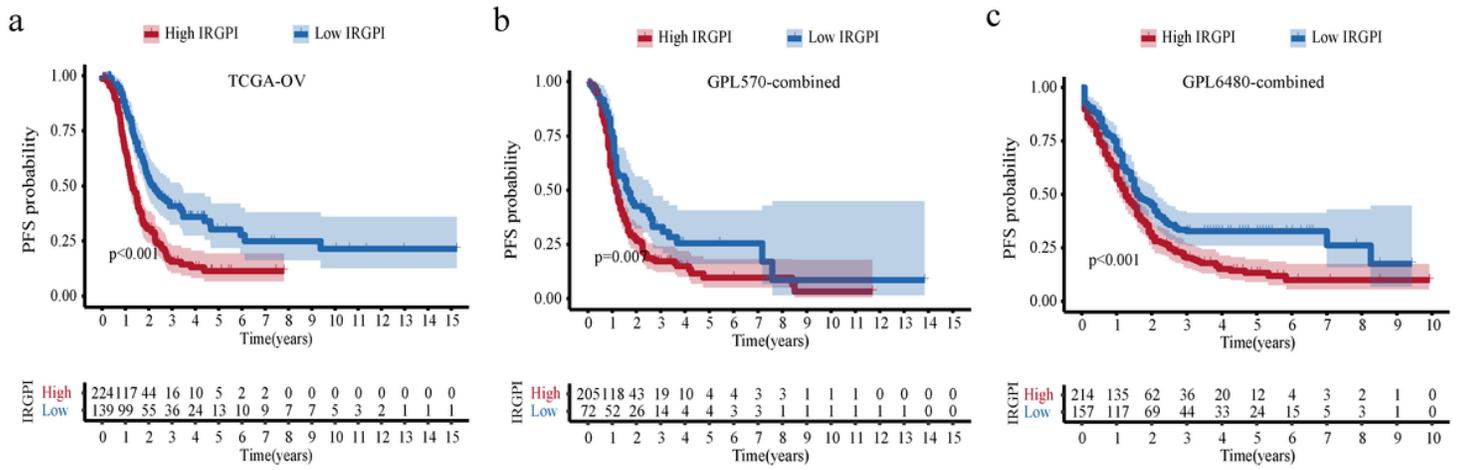


Figure 4

Verification of IRGPI's ability to predict PFS. Kaplan-Meier curve of PFS in TCGA-OV (a), GPL570-combined (b), and GPL6480-combined cohorts (c).

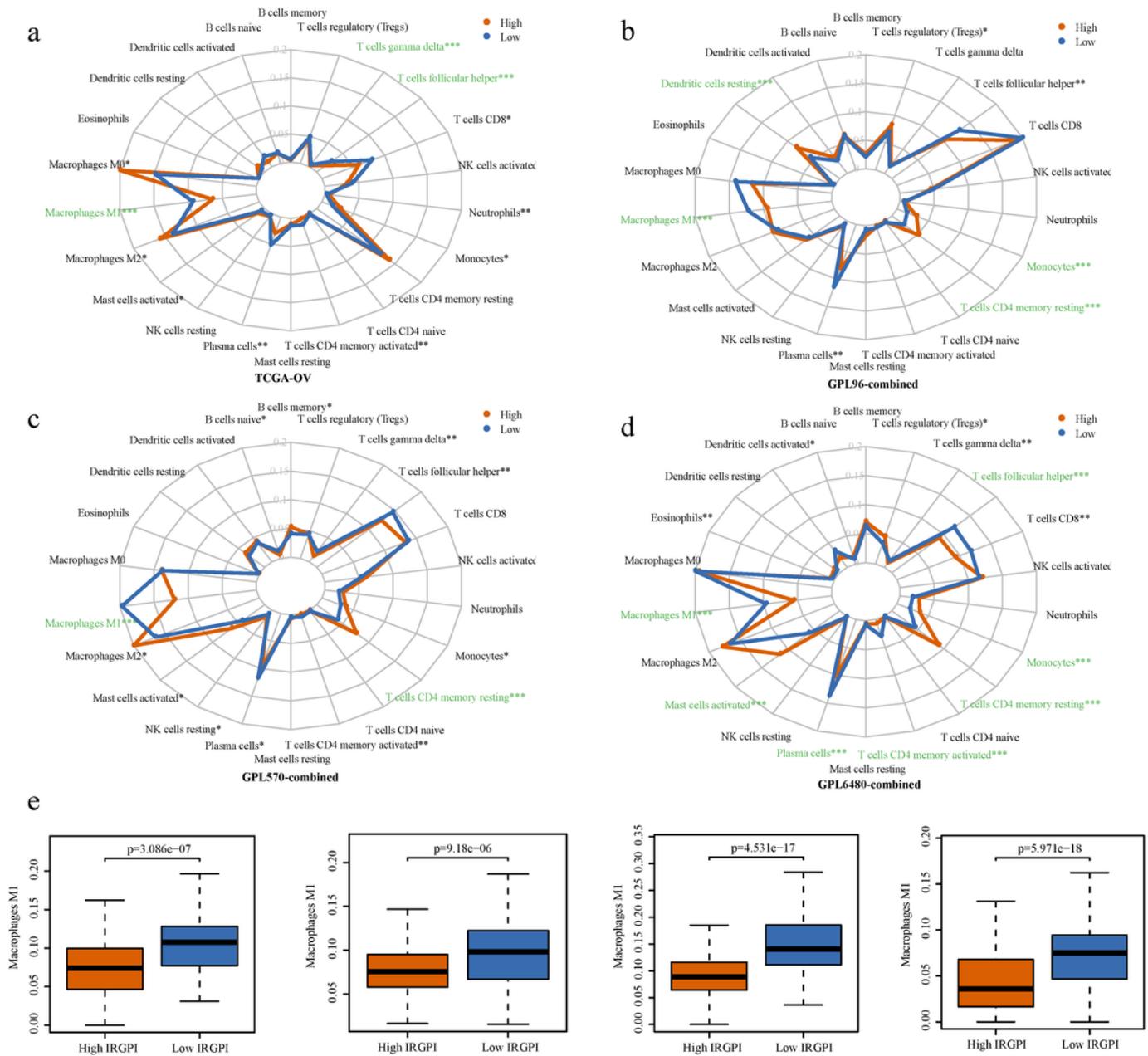


Figure 5

The infiltration levels of immune cells were compared for the high-IRGPI and the low-IRGPI groups. Infiltration level radar map of TCGA-OV (a), GPL96-combined (b), GPL570-combined (c), and GPL6480-combined cohorts (d). A difference in the infiltration of M1 macrophages M1 was observed between the two groups (e). $P < 0.05$ indicates that this difference is statistically significant.

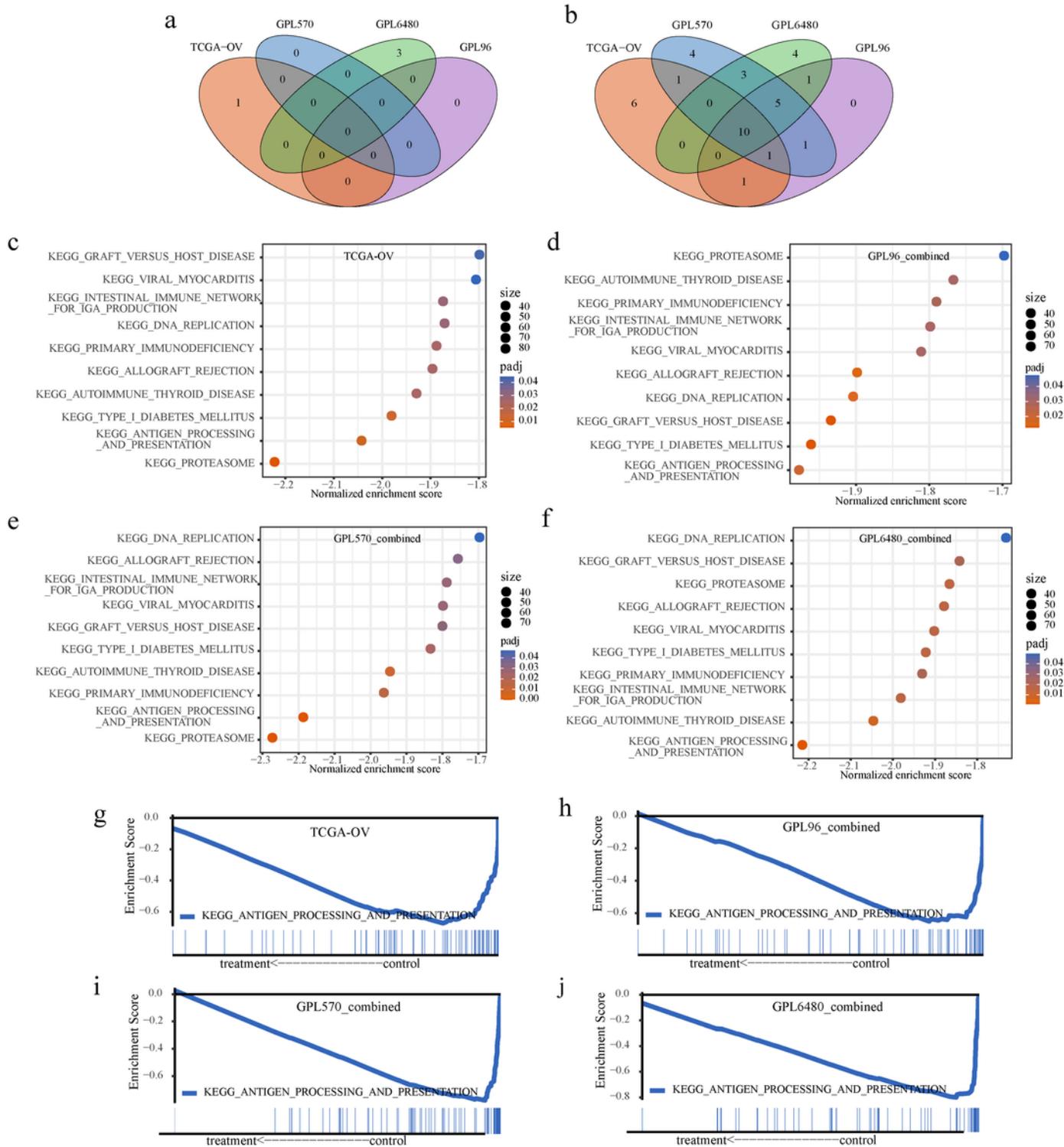


Figure 6

Gene set enrichment analysis. a Venn diagram of High-IRGPI group enrichment pathways. b Venn diagram of Low-IRGPI group enrichment pathways. c-f For the Low-IRGPI group, the bubble diagram of the common enrichment pathways of four cohorts. g-j GSEA line chart of the KEGG "Antigen processing and presentation" pathway.

Supplementary Files

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

- [Additionalfile1TableS1S5.xlsx](#)
- [Additionalfile1TableS1S5.xlsx](#)