

## Article

# Identification of an Individualized Prognostic Biomarker for Serous Ovarian Cancer: A Qualitative Model

Fengyuan Luo <sup>1,†,‡</sup>, Na Li <sup>1,†,‡</sup>, Qi Zhang <sup>2,†,‡</sup>, Liyuan Ma <sup>1</sup>, Xinqiao Li <sup>1</sup>, Tao Hu <sup>1</sup>, Haijian Zhong <sup>1</sup>, Hongdong Li <sup>1,\*</sup> and Guini Hong <sup>1,\*</sup>

<sup>1</sup> School of Medical Information Engineering, Gannan Medical University, Ganzhou 341000, China

<sup>2</sup> Affiliated Hospital of Jiangxi, University of Chinese Medicine, Nanchang 330006, China

\* Correspondence: biomantis\_lhd@163.com (H.L.); hongguini08@gmail.com (G.H.)

† These authors have contributed equally to this work.

‡ Fengyuan Luo, Na Li and Qi Zhang are co-first authors.

**Abstract:** Serous ovarian cancer is the most common type of ovarian epithelial cancer and usually has a poor prognosis. The objective of this study was to construct an individualized prognostic model for predicting overall survival in serous ovarian cancer. Based on the relative expression orderings ( $E_a > E_b / E_a \leq E_b$ ) of gene pairs closely associated with serous ovarian prognosis, we tried constructing a potential individualized qualitative biomarker by the greedy algorithm and evaluated the performance in independent validation datasets. We constructed a prognostic biomarker consisting of 20 gene pairs (SOV-P20). The overall survival between high- and low-risk groups stratified by SOV-P20 was statistically significantly different in the training and independent validation datasets from other platforms ( $p < 0.05$ , Wilcoxon test). The average area under the curve (AUC) values of the training and three validation datasets were 0.756, 0.590, 0.630, and 0.680, respectively. The distribution of most immune cells between high- and low-risk groups was quite different ( $p < 0.001$ , Wilcoxon test). The low-risk patients tended to show significantly better tumor response to chemotherapy than the high-risk patients ( $p < 0.05$ , Fisher's exact test). SOV-P20 achieved the highest mean index of concordance (C-index) (0.624) compared with the other seven existing prognostic signatures (ranging from 0.511 to 0.619). SOV-P20 is a promising prognostic biomarker for serous ovarian cancer, which will be applicable for clinical predictive risk assessment.

**Keywords:** serous ovarian cancer; relative expression orderings; prognostic biomarker



**Citation:** Luo, F.; Li, N.; Zhang, Q.; Ma, L.; Li, X.; Hu, T.; Zhong, H.; Li, H.; Hong, G. Identification of an Individualized Prognostic Biomarker for Serous Ovarian Cancer: A Qualitative Model. *Diagnostics* **2022**, *12*, 3128. <https://doi.org/10.3390/diagnostics12123128>

Academic Editor: Jure Knez

Received: 3 November 2022

Accepted: 9 December 2022

Published: 12 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Ovarian cancer is one of the three major malignant tumors of the female reproductive system, and its incidence ranks third among gynecological malignancies [1,2]. However, because the ovaries are deep in the pelvis, the onset is insidious, and the early symptoms are not apparent [3]. As the most common ovarian epithelial tumor subtype, serous ovarian cancer accounts for about 40–50% of all malignant ovarian tumors, and its survival rate within five years is only about 30% [4]. Clinically, physicians treat serous ovarian cancer patients with the same clinical and pathological stages in the same way. However, due to the heterogeneity of the tumor, patients usually have different survival outcomes. Studies have shown clinical prognostic factors such as the stage of The Federation of Gynecology and Obstetrics (FIGO), the size of residual after debulking surgery, BRCA1/2 mutations, and the degree of lymphocyte infiltration could cause gene expression changes in serous ovarian cancer [5,6]. These influencing factors are related to the high recurrence rate and poor prognosis of serous ovarian cancer through the activation of oncogenes and the inactivation of tumor suppressor genes [7]. Therefore, there is an urgent need to explore gene expression biomarkers that can predict the prognostic effect of serous ovarian cancer.

In recent years, many studies have used gene expression information to develop biomarkers for serous ovarian cancer [8]. The level of patient risk was determined by

a predetermined threshold of the risk score in the models [9–11]. However, such a risk-score-based prognostic model has some limitations. Firstly, gene expression levels are sensitive to systematic bias in high-throughput data [12,13]. The risk thresholds generated by the training dataset cannot be directly applied to independent datasets. Therefore, the training and validation data need to be pre-normalized to make the trained thresholds applicable to new data [14]. This means that the risk of samples in the training data can significantly influence the risk prediction outcome of an unknown sample. In addition, data normalization requires prior sample collection, which may be clinically impractical. However, without data normalization, the risk scores of independent samples are not comparable [15,16]. Therefore, finding stable prognostic risk biomarkers with clinical translational value in serous ovarian cancer samples has become an urgent issue.

Several studies have developed disease-related biomarkers based on the within-sample relative expression orderings (REOs) of genes [17–19]. Studies have claimed that REOs are insensitive to systematic biases [20,21]. They are also robust against interindividual biological variation of gene expression levels. Moreover, REOs use qualitative information about gene expression. They relate only to relative changes in gene expression within individual samples. Thus, they are independent of other samples, indicating that REO-based biomarkers can be naturally applied to individual samples [22]. Thus, our objective in this study was to develop an individualized prognostic model for serous ovarian patients based on the within-sample REOs of genes.

## 2. Materials and Methods

### 2.1. Study Population

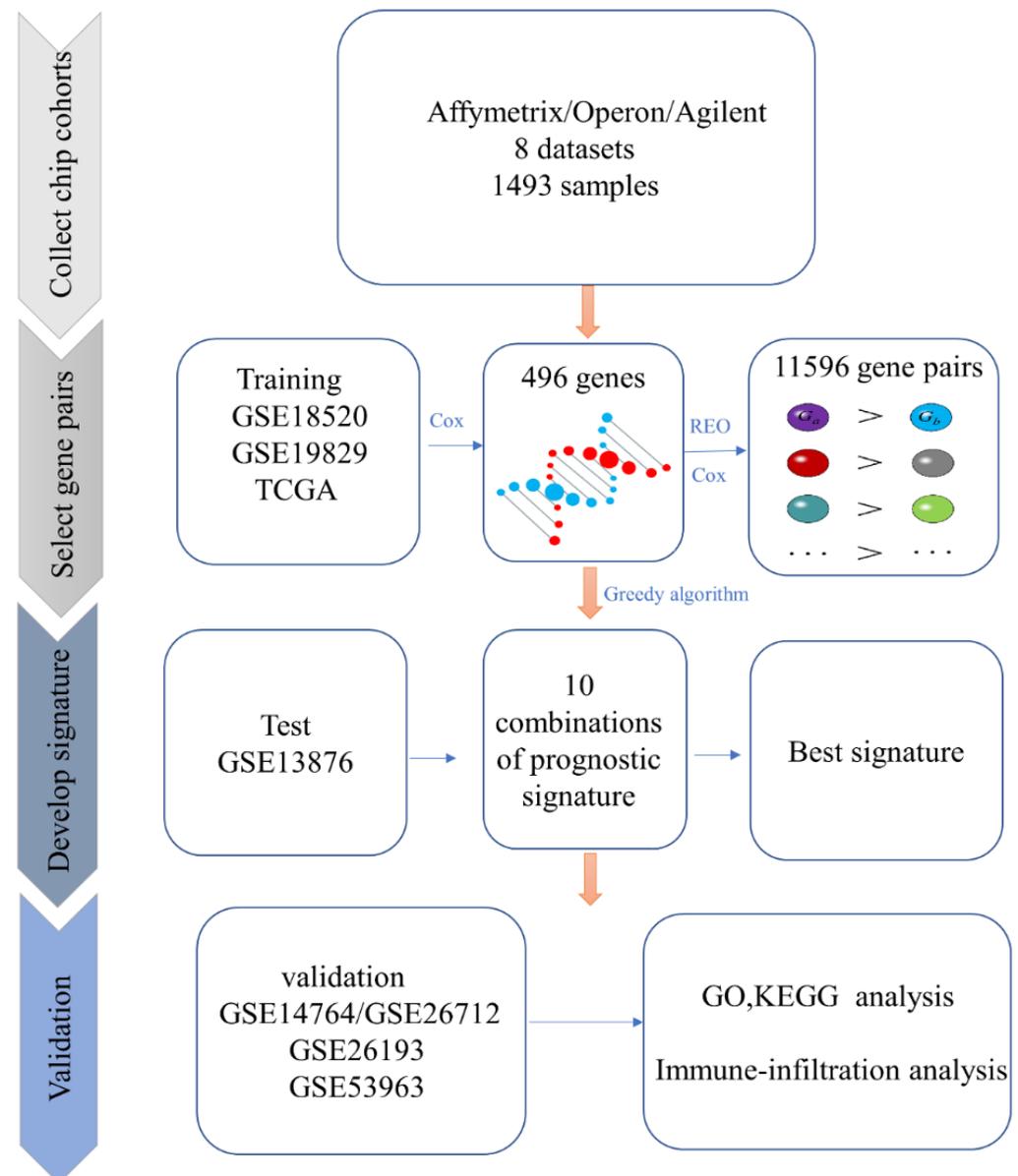
Gene expression profiles from eight independent array datasets containing 1493 samples were collected from Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>, accessed on 11 October 2020) and the UC Santa Cruz Xena platform (UCSC Xena, <https://xenabrowser.net/datapages/>, accessed on 1 November 2020). The division of data into training, test, and validation sets is shown in Figure 1. Training datasets included 561 samples divided into low-grade (G1,  $n = 6$ ), high-grade (G2–G4,  $n = 544$ ), borderline ( $n = 9$ ), and ungraded ( $n = 2$ ). Clinical information, including age, grade, and stage, was available for most datasets. The detailed clinical information for each dataset can be found in Table 1. Of all the data used, chemotherapy information was only available for the 466 samples from TCGA. Detailed tumor response information for TCGA samples was provided in Supplementary Table S1.

The RMA algorithm pre-processes the raw microarray data downloaded from GEO. The annotation files in each platform were used to map probe IDs to gene IDs. Probes that were not mapped to genes were deleted. For different probes mapped to the same gene, the average was used as the final expression value of the gene. Normalized gene expression profiling data and clinical data from 630 TCGA ovarian cancer patients were downloaded from the UCSC Xena platform. The data were pre-processed in the following steps before constructing a prognostic biomarker: (1) remove normal samples; (2) remove tumor samples without clinical information; (3) remove samples with overall survival of 0 days; and (4) remove low-expression genes (more than half of the samples with missing or 0 gene expression).

### 2.2. Identification of Candidate Prognosis-Related Gene Pairs

Univariate Cox regression analysis was used to evaluate the predictive value of genes and gene pairs. When Cox regression analysis was applied to single genes, a  $p$ -value less than 0.05 was considered significantly associated with patients' overall survival. Pairing every two significant prognosis-related genes yields candidate prognosis-related gene pairs. Then, the REO matrix  $X$  of the candidate prognosis-related gene pairs was constructed.  $X$  is a 0–1 matrix, wherein  $x_{ij} = 1$  indicates that the REO of the gene pair  $i$  ( $G_a, G_b$ ) in the serous ovarian cancer sample  $j$  is  $G_a > G_b$ , and  $x_{ij} = 0$  indicates  $G_a \leq G_b$ . When Cox

regression analysis was applied to gene pairs, an adjusted  $p$ -value less than 0.05 after the Benjamini–Hochberg correction was considered significant [23].



**Figure 1.** Flow chart of the study.

### 2.3. Identification of The Prognostic Biomarker

Given the concordance index (C-index) values for all gene pairs in the significant prognosis-related gene pairs, we selected the prognostic biomarker comprising top-ranked gene pairs in a greedy manner. First, the first  $N$  gene pairs with the highest C-index were selected. Then, for each top  $N$  gene pair, the remaining significant prognosis-related gene pairs were added to form the combination if the added gene pairs could increase the C-index values of the combination. For each of the  $N$  combinations, the test set GSE13876 was used to find the combination significantly correlated with the prognosis of serous ovarian cancer with a higher C-index value as the final prognostic-related biomarker gene pairs.

### 2.4. Performance Evaluation of The Prognostic Biomarker

A sample was classified into the high- or low-risk group based on the majority voting of the gene pairs in the prognostic biomarker by their REOs. Gene pairs not detected in

cross-platform validation data were removed, and the remaining gene pairs were used to determine the risk labels.

**Table 1.** Clinical characteristics of serous ovarian cancer samples used in this study.

	Training			Test	Validation			
	GSE18520	GSE19829	TCGA		Dataset 1	Dataset 2	Dataset 3	
GEO accession	GSE18520	GSE19829	TCGA	GSE13876	GSE14764	GSE26712	GSE26193	GSE53963
Microarray platform	GPL570	GPL8300	GPL96	GPL7759	GPL96	GPL96	GPL570	GPL6480
Sample No.	53	42	466	415	80	185	78	174
				Stage				
I	-	0	15	-	8	-	12	0
II	-	1	28	-	1	-	20	8
III	-	35	369	-	69	-	53	125
IV	-	6	51	-	2	-	14	41
Late	53 (III–IV)	-	-	415 (III–IV)	-	-	-	-
Unstaged	-	-	3	-	-	-	-	-
Age, median (range), y	-	58.3 (39–80)	59.9 (26–89)	57.9 (21–84)	-	-	-	63 (24–89)
				Grade				
G1	-	1	5	-	3	-	7	0
G2	-	9	60	-	23	-	33	4
G3	-	32	389	-	54	-	67	90
G4	-	0	1	-	-	-	-	80
High-grade (G2/3/4)	53	-	-	-	-	-	-	-
Borderline	-	-	9	-	-	-	-	-
Ungraded	-	-	2	-	-	-	-	-
Survival, median (range), m	40.4 (5–150)	38.5 (1–68)	59.5 (0.27–182.70)	45.6 (1–234)	35 (7–73)	38.3 (0.72–163.80)	60.2 (0.1–133.71)	55.7 (0.3–201.61)

Multivariate Cox risk regression analyses were conducted to evaluate the joint prognostic significance of the developed prognostic biomarker and clinical factors, including age, stage, and grade. As clinical information was unavailable in all datasets (see Materials and methods, Table 1), we only included the existing clinical elements in each dataset. The overall survival was calculated using the Kaplan–Meier, and the log-rank test was used to compare the two survival curves. Forest plots were used to display the results of multiple Cox regression analyses.

### 2.5. Functional Enrichment Analysis

Gene annotation and pathway analysis were performed in the database of Metascape (<http://www.metascape.org/>) to search for functions associated with genes in the prognostic biomarker. We considered the following ontology sources in Metascape: KEGG Pathways, GO Biological Processes, Reactome Gene Sets, Canonical Pathways, CORUM, TRRUST, DisGeNET, PaGenBase, Transcription Factor Targets, Wiki Pathways, and COVID. To further understand the biological function of the differentially expressed genes between the high- and low-risk groups predicted by the prognostic biomarker, functional enrichment analysis was carried out based on the Kyoto Encyclopedia of Genes and Genomes (KEGG). The relevance was considered significant if the Wilcoxon *p*-value was less than 0.05.

### 2.6. Immune Infiltration Analysis

TIMER2.0 (<http://timer.cistrome.org/>, accessed on 10 September 2021) was used to quantify the proportions of immune cell infiltration, and the differences between the high- and low-risk groups were compared. The content of six immune cells (B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells) was estimated, and the differences were compared using the Wilcoxon test method.

### 2.7. Drug Sensitivity Analysis

The chemotherapeutic response for each serous ovarian cancer sample in the training sets was predicted according to the *Genomics of Drug Sensitivity in Cancer* (GDSC, <https://www.cancerrxgene.org/>, accessed on 24 November 2022) database. The chemotherapy sensitivity of each tumor sample was evaluated by the half-maximal inhibitory concentration (IC50) value, which was obtained using regression analysis, and the prediction accuracy was measured by 10-fold cross-validation. The differences in IC50 values between high- and low-risk groups were compared using the Wilcoxon test method.

### 2.8. Performance Comparison with Other Prognostic Biomarkers

Seven published prognostic models based on gene expression data were selected [24–29]. Since most of the models are constructed based on TCGA-OV sequencing data, the RNA sequencing data of TCGA-OV in UCSC Xena are collated to ensure comparativeness. According to the median risk score of each model, the samples were also split into high- and low-risk groups, and the overall survivals were compared.

### 2.9. Statistical Analysis

Statistical analyses were performed using R (version 4.2.0). We used the *survival* R package for univariate and multivariate risk regression analysis. The *timeROC* and *survival* packages were used to evaluate the receiver operating characteristic curve (ROC curve) and calculate area under curve (AUC) values. The *survminer* and *survival* R packages were used to show the K-M curves and Forest plots of grouped samples. The *ggplot2* and *clusterProfiler* packages were used to enhance the appearance of the enrichment plot, and *ggplot2* was also used to draw a violin boxplot diagram. The *oncoPredict* was used to determine the IC50 for evaluating drug sensitivity. All parameters are default, including “*combat*”, to eliminate the batch effect and the average of repeated gene expression. The *rms* R package was used for calculating the C-index.

## 3. Results

### 3.1. Identification of the Prognosis-Related Biomarker

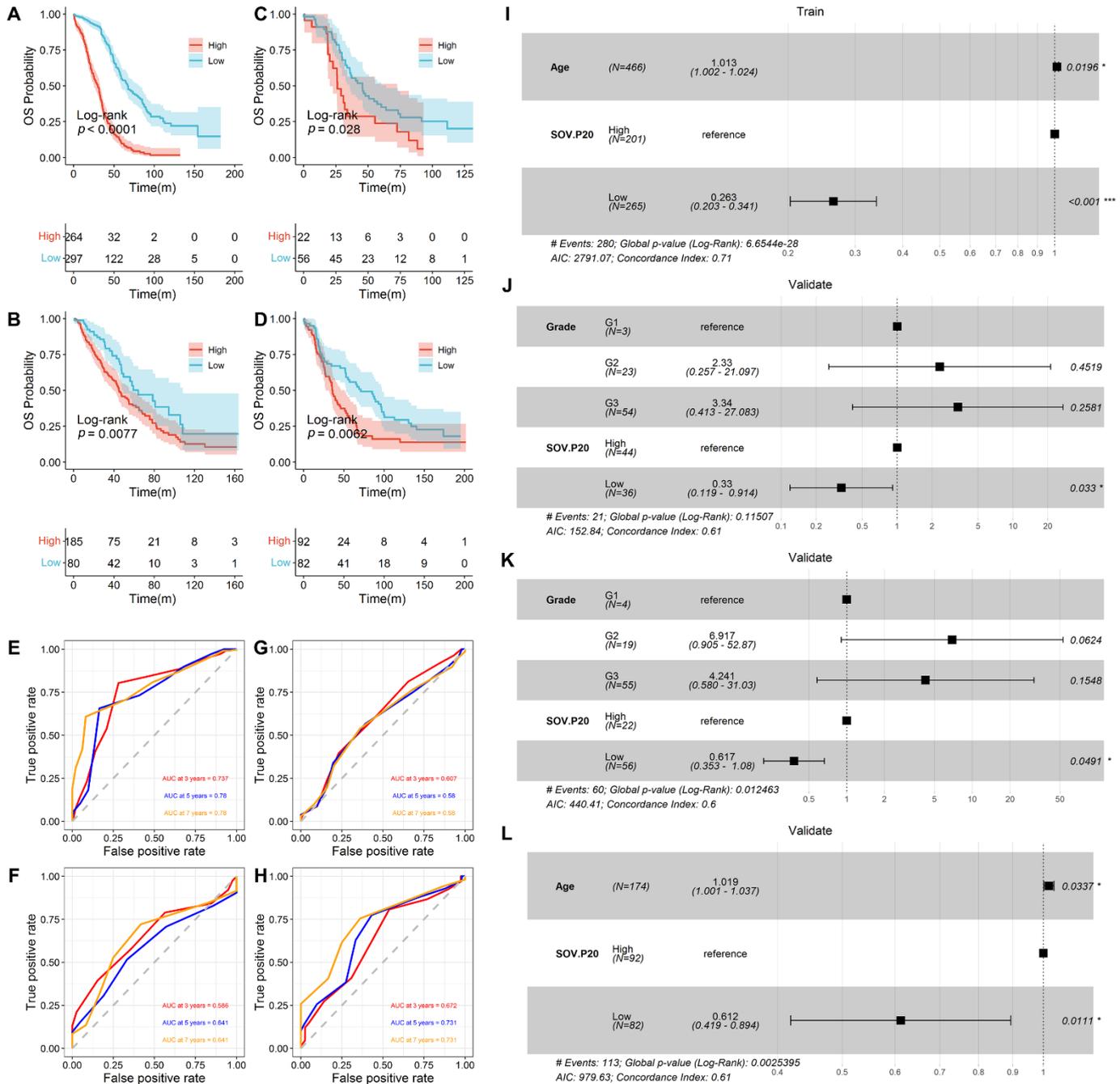
To detect the prognosis-related gene pairs, we first performed the Cox regression to find genes that are significant predictors of survival. Only genes common to all platforms in the training set were considered. There were 6934 genes in common, of which 82,550, and 796 were significantly correlated with overall survival in GSE18520, GSE19829, and TCGA-OV data, respectively ( $p < 0.05$ , Wilcoxon test, Supplementary Table S2). Among them, 1347 were significant in at least one of the three datasets, and 486 genes had consistent directions of hazard coefficients (positive or negative). These 486 genes were selected for subsequent analysis. After pairing the 486 genes, we obtained 117,855 prognosis-related gene pairs, and 11,569 were significantly associated with overall survival ( $p < 0.05$ , Wilcoxon, BH adjusted for multiple testing).

Then, the prognosis-related biomarker was developed by applying the greedy method to the top gene pairs with the largest C-index values (see Materials and Methods). Here, we used the top 10 gene pairs (C-index ranging from 0.68 to 0.72) to form the final combinations of selected gene pairs. The test dataset was used to obtain the best combination, and univariate Cox regression analysis was performed for each of the ten combinations. The combination significantly correlated with overall survival with the largest C-index value ( $p = 0.00046$ , Wilcoxon test; C-index = 0.69) was selected as the final prognostic biomarker, which contains 20 gene pairs (Supplementary Table S3), and referred to as SOV-P20.

### 3.2. Prediction of Overall Survival by SOV-P20

In the training set, 298 and 452 samples were categorized into high- and low-risk groups, respectively. The high-risk group had a significantly decreased overall survival compared to the low-risk group (Figure 2A,  $p < 0.0001$ , HR = 0.23, 95% CI: 0.18–0.29). In the three validation datasets, 292 and 136, 44 and 145, and 129 and 150 samples were divided

into high- and low-risk groups, respectively. Kaplan–Meier survival analyses showed that in all datasets, the high-risk group had a significantly lower overall survival rate. In contrast, the low-risk patients usually showed a longer survival time, with a significant difference between the two groups ( $p = 0.0077$ , HR = 0.60, 95% CI: 0.41–0.88, Figure 2B;  $p = 0.028$ , HR = 0.54, 95% CI: 0.31–0.94, Figure 2C;  $p = 0.0062$ , HR = 0.59, 95% CI: 0.41–0.87, Figure 2D).



**Figure 2.** SOV-P20 predicts the high- and low-risk stratification of overall survival. (A–D) Kaplan–Meier overall survival curves of serous ovarian cancer patients in the training and three validation datasets; (E–H) the time-dependent ROC curves of SOV-P20 predict the 3-, 5-, and 7-year overall survival in the training and three validation datasets; (I–L) multivariate Cox regression analysis forest plots in the training and three validation datasets, \*  $p < 0.05$ ; \*\*\*  $p < 0.001$ .

Since the overall survival time of patients exceeded five years, the AUCs of SOV-P20 for 3, 5, and 7 years were calculated. The mean AUC was 0.756 in the training set, while in the three validation datasets, the mean AUCs were 0.590, 0.630, and 0.680 (Figure 2E–H), respectively, indicating a potential predictive value of the SOV-P20 biomarker.

Multivariate Cox analyses were conducted to confirm whether the SOV-P20 was an independent prognostic factor. The results showed that age and SOV-P20 were significantly associated with overall survival ( $p < 0.05$ , log-rank test, Figure 2I–L), and SOV-P20 was an independent risk factor for patients with serous ovarian cancer.

### 3.3. Functional Enrichment Analysis

Functional enrichment analysis was performed on the 37 genes comprising SOV-P20. As shown in Figure 3, eight GO terms related to cytokine production were significantly enriched (Figure 3A). Cytokine refers to small molecule polypeptides secreted primarily by immune cells that regulate cell function [30]. Cytokines and cytokine receptor processes are mainly related to regulating the body's immune response, hematopoietic function, and inflammatory response [31]. They can inhibit tumorigenesis and progression and have been shown to be effective in cancer treatment [32]. Three genes are involved in the virus's entry into the host cell biology pathway. Pathway analysis showed that five genes were associated with tumor recurrence (Figure 3B), confirming that SOV-P20 plays a role in cancer onset and development.

Further enrichment analysis for differentially expressed genes between the high- and low-risk groups identified three significant KEGG pathways (BH adjusted  $p < 0.05$ ), including PI3K-Akt signaling pathway, proteoglycan pathway, and AGE-RAGE signaling pathway in cancer (Figure 3C). These pathways have been reported to play an essential role in the progression and prognosis of serous ovarian cancer [33–35].

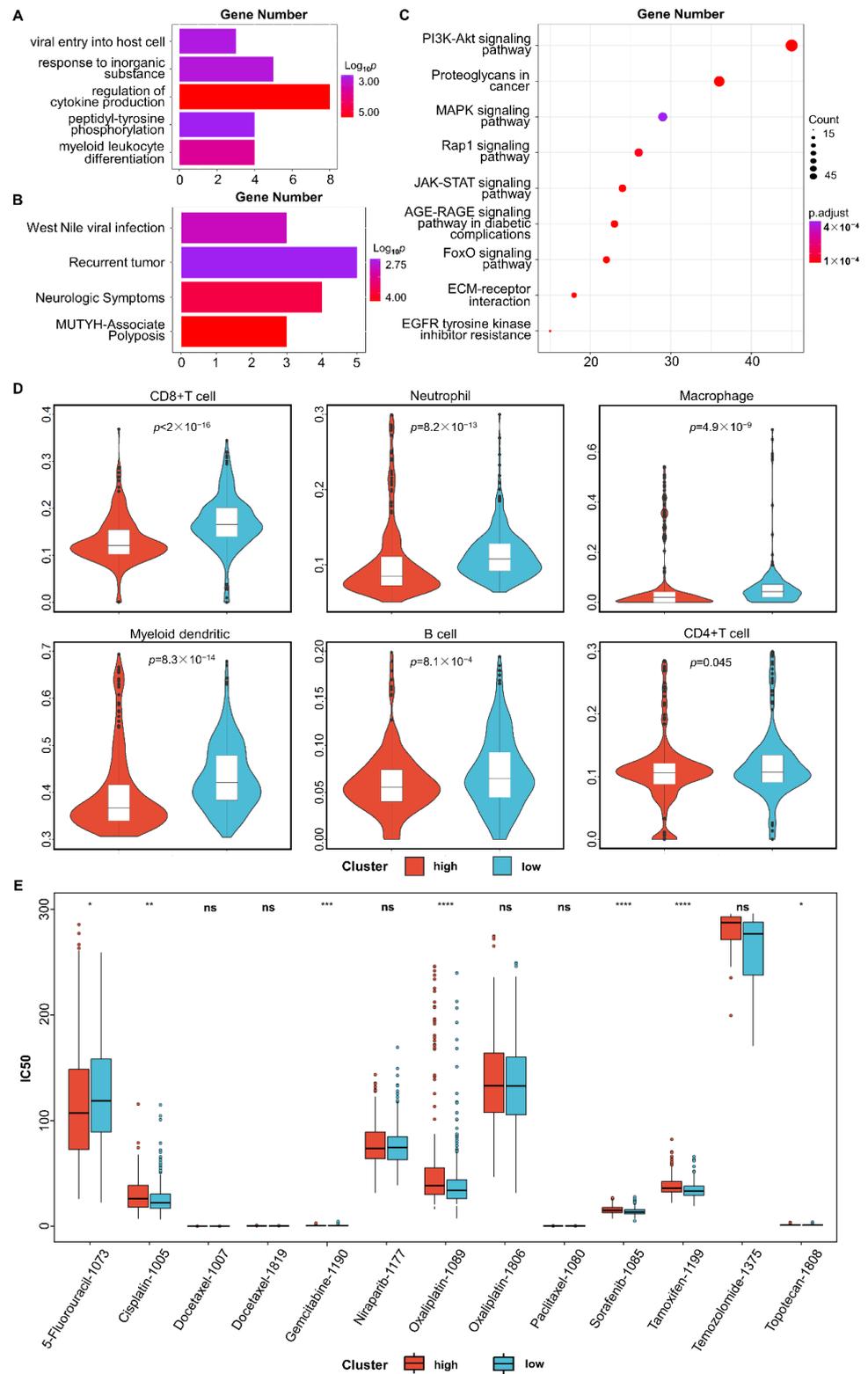
### 3.4. Immune Infiltration Analysis

The distribution of tumor-infiltrating immune cells is an essential indicator of immune invasion and prognosis in patients [36]. Results from TIMER showed significant differences in the infiltration levels of immune cells between the high- and low-risk groups ( $p < 0.05$ , Wilcoxon test, Figure 3D), and the low-risk group had higher infiltration levels.

### 3.5. Tumor Response to Drug Treatment

For all training samples, only TCGA data provide information about an individual patient's response to chemotherapy drugs (Supplementary Table S1). The RECIST was used to evaluate tumor response to chemotherapy drugs, categorized as complete response (CR), partial response (PR), stable disease (SD), or progressive disease (PD). Only 338 patients with RECIST measurements were analyzed. Among them, 143 were high-risk, and 195 were low-risk. They were further classified as chemo-sensitive (CR and PR) or chemo-resistant (SD and PD). Of the 143 high-risk patients, 107 were chemo-sensitive, and 36 were chemo-resistant. In contrast, of the 195 low-risk patients, 171 were chemo-sensitive, and 24 were chemo-resistant, significantly different from the high-risk groups (Fisher's exact test,  $p = 2.50 \times 10^{-3}$ ). In addition, compared to high-risk patients, low-risk cases tended to be chemo-sensitive (odds ratio = 2.397) and were less likely to be chemo-resistant (odds ratio = 0.417).

Similar analyses were performed for 310 TCGA patients who received platinum-based chemotherapy. There was a significant difference in sensitivity and resistance to platinum-based chemotherapy between the high- and low-risk groups (Fisher's exact,  $p = 6.59 \times 10^{-4}$ ), as the number of chemo-sensitive and chemo-resistant patients were 100 and 33 for high-risk groups, respectively, while in low-risk groups, the numbers were 157 and 20, respectively. Low-risk cases were more likely to be chemo-sensitive to platinum (odds ratio = 2.591) and less inclined to develop chemoresistance to platinum (odds ratio = 0.386) relative to the high-risk group.



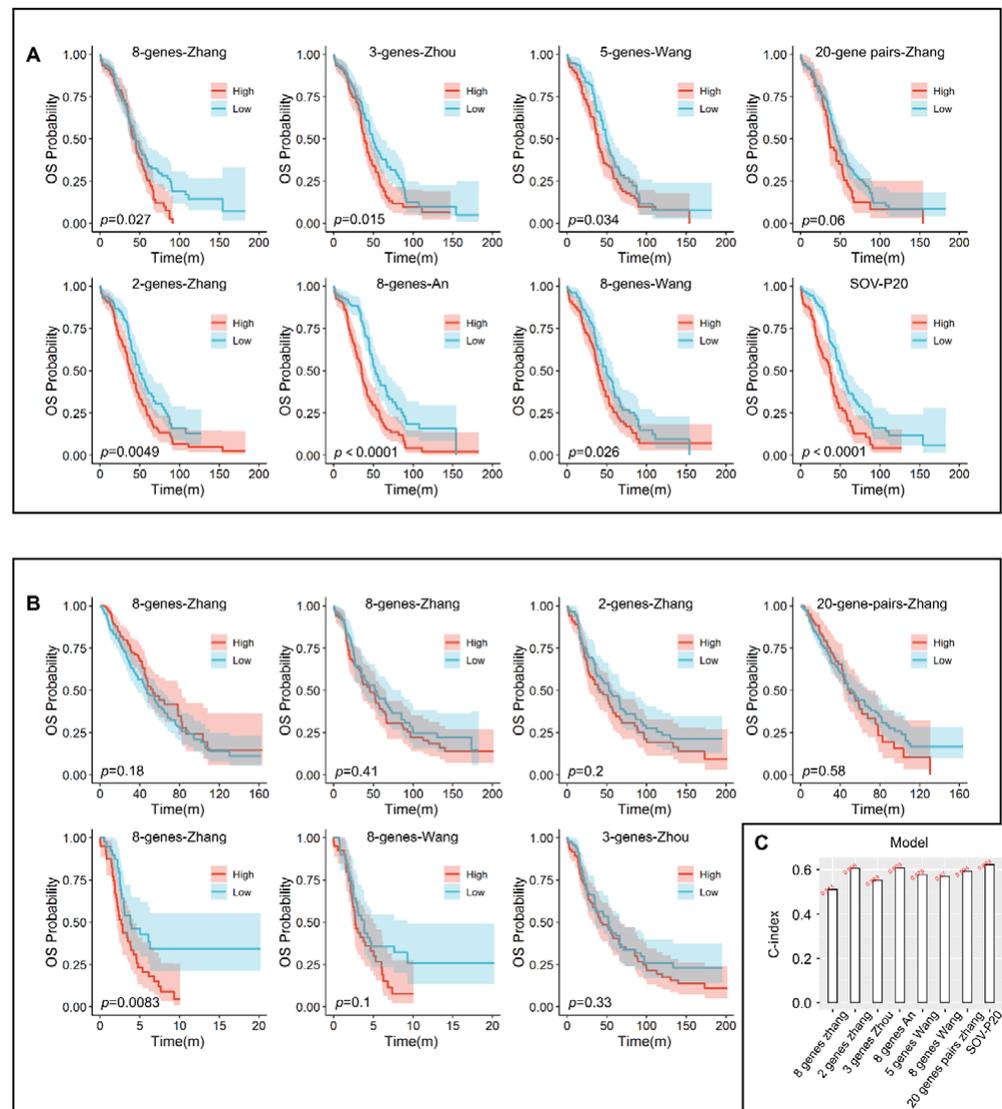
**Figure 3.** Functional enrichment and immune infiltration analysis. (A) Significant GO biological processes enriched by SOV-P20. (B) Significant DisGeNet pathways enriched by SOV-P20. (C) Significant KEGG pathways enriched by differentially expressed genes between the high- and low-risk groups divided by SOV-P20. (D) Vioplot of the difference in the infiltration scores of CD8+ T cells, neutrophils, macrophages, myeloid dendritic, B cells, and CD4+ T cells. (E) The difference in drug response between the high- and low-risk groups divided by SOV-P20. The legend is as following: ns, non-significant; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ; \*\*\*\*  $p < 0.001$ .

Since only TCGA samples had drug-response information, we used the GDSC database to predict the response to commonly used chemotherapeutic drugs for each training sample. As shown in Figure 3E, patients in the low-risk groups were more sensitive to Cisplatin, Gemcitabine, Oxaliplatin, Sorafenib, Tamoxifen, and Topotecan. In contrast, patients in the high-risk groups were more sensitive to 5-Fluorouracil, and there were significant differences between the high- and low-risk groups ( $p < 0.05$ , Wilcox test).

The above results revealed a significant difference in tumor response to chemotherapy between the high- and low-risk groups predicted by SOV-P20.

### 3.6. Comparison with Other Models

To further validate the predictive value of SOV-P20, we compared its performance with seven different prognostic models previously reported using TCGA-OV RNA sequencing data. Survival was significantly different between high- and low-risk groups in TCGA-OV samples assessed by the seven models, and the high-risk group had a poor prognosis (Figure 4A, all  $p < 0.05$ ). The C-index of the above seven prognostic models was smaller than the C-index of SOV-P20 (Figure 4A), indicating that the overall predictive performance of the prognostic model in this paper was better than that of the other seven models in other studies.



**Figure 4.** (A) Kaplan–Meier plots of the seven published models and SOV-P20 in TCGA-OV RNA sequencing data; (B) Kaplan–Meier plot of the seven published models in validation set; (C) bar plot of C-index values for the seven published models and SOV-P20.

The performance comparison of SOV-P20 with the seven predictive models was also performed using the validation set collected in this study. In the seven models, only the genes detected in each dataset were analyzed. Results showed that, except for the 8-gene model by Zhang et al. [37] with a significant Kaplan–Meier curve in the validation set 2 ( $p = 0.0083$ ), all Kaplan–Meier curves of the remaining prognostic models in each validation dataset showed no statistically significant difference between high- and low-risk groups (Figure 4B). SOV-P20 achieved the highest mean C-index (0.624) compared with the other six existing prognostic signatures (ranging from 0.511 to 0.619, Figure 4C). The above results further suggested that the prognostic biomarker developed in this study had a robust predictive ability.

## 4. Discussion

### 4.1. Main Findings

In this study, we constructed the gene-pair-based biomarker, SOV-P20, to predict the prognosis of serous ovarian cancer and further validated the predictive value in independent datasets. Our results suggest that SOV-P20 has strong robustness and stable predictive performance in datasets from different platforms. Many genes involved in SOV-P20 have been reported as prognostic biomarkers or potential therapeutic targets for serous ovarian cancer. The literature supporting the association between genes in SOV-P20 and serous ovarian cancer is shown in Supplementary Table S3. Genes in SOV-P20 were significantly enriched in the biological pathways related to regulating cytokine production, virus entry into host cells, and tumor recurrence. The high- and low-risk groups predicted by SOV-P20 showed a significant difference in overall survival time and tumor-infiltrating immune cells. Differentially expressed genes between the high- and low-risk groups were significantly enriched in the PI3K-Akt signaling pathway, the proteoglycan pathway, and the AGE-RAGE signaling pathway in cancer. Compared with the published prognostic models, SOV-P20 stood out in terms of C-index, demonstrating its good predictive ability for ovarian cancer prognosis.

### 4.2. Strengths and Limitations

Many studies have developed genetic biomarkers for stratifying patient survival on different datasets. However, these biomarkers have not been applied in the clinical setting due to various limitations. There are many reasons. Firstly, most researchers agree that developing a robust prognostic biomarker requires a large sample size. Most studies have a relatively small sample size, which may suffer from model overfitting problems. As shown in this study, by applying Cox regression to the three datasets in the training set, the prognosis-related genes were detected and intersected, and few intersection genes were found (Supplementary Table S2). Therefore, when developing the biomarker, we combined 561 samples into the training set to meet the requirement of large sample size while trying to include data from different platforms. Second, most current predictive biomarkers for patient stratification are based on risk scores, which are sensitive to system deviations and poorly robust. To address this issue, we developed the predictive model based on the REOs within each sample. As a qualitative model, it is highly robust and insensitive to systematic bias and laboratory batch effects. The prognostic value of the developed biomarker was confirmed in the test set of 415 samples and the validation set of 517 samples from different platforms.

However, there are some limitations to this study. First, the method of identifying disease biomarkers based on within-sample REOs is a qualitative metric. Considering only REOs of genes may lose some subtle quantitative information about gene expression. Second, the imperfect clinical characteristics of the dataset may reduce the accuracy of multivariate survival analysis and comprehensive prediction model when performing multivariate Cox regression analysis. Third, genes in the prognostic biomarker were not subjected to a wet biological experiment, which deserves continued study in detail in future work.

### 4.3. Interpretation of Findings

Many studies have used gene expression information to develop predictive biomarkers or diagnostic models for serous ovarian cancer (Supplementary Table S4). We divided these studies into two main categories according to the number of feature genes.

In the first category, researchers assessed the impact of individual genetic biomarkers on the prognosis of ovarian cancer. For example, Buttarelli et al. identified the long-chain noncoding RNA-MEG3 gene as a tumor suppressor for serous ovarian cancer through *in vitro* and *in vivo* experiments, demonstrating that its high expression predicts better progression-free survival and overall survival [8]. A study by Zhang et al. showed that miR-212-3p might suppress serous ovarian cancer by directly targeting MAP3K3 [38]. However, such single-gene-based studies do not consider the effects of gene-to-gene interactions and might not accurately predict the prognosis of patients with serous ovarian cancer.

The second category evaluates the predictive value of a set of genes based on a specific biological function. For example, Yang et al. constructed a risk scoring model containing seven genes to predict the prognosis of serous ovarian cancer based on stromal and immune infiltrates in the immune tumor microenvironment [39]. Wang et al. constructed an 8-gene predictive risk assessment model for serous ovarian cancer from energy metabolism-related genes [28]. Zheng et al. built a prognostic risk scoring model for serous ovarian cancer using 11 lipid metabolism-related genes [11]. However, such biological function-based predictive biomarkers may ignore the heterogeneity of individual patients and the complexity of influencing factors and are prone to overfitting.

We found more than 20 published prognostic models for serous ovarian cancer (see Supplementary Table S4). However, many models' genes could appear undetected when applied to datasets from different platforms. In this study, we did not compare the single-gene-based biomarkers and only selected those models whose comprising genes were detected across platforms for comparison. Finally, only seven models were included in the comparison. We found that the Kaplan–Meier curve results for six models were not statistically significant enough to predict the prognosis in independent datasets other than their own. According to our study, when the median risk score of the training set was taken as the threshold, the test set could not make prognostic predictions. In contrast, SOV-P20 was developed from samples from different platforms, and it can still predict the prognosis on the principle of half voting when the platform does not detect all genes. Actually, SOV-P20 can be applied to any data set where no thresholds are needed. Moreover, it is REO-based and thus robust to noise in the data and offers a natural way to overcome the heterogeneity across datasets, especially across different platforms. Therefore, SOV-P20 can achieve the goal of personalized precision medicine.

Most serous ovarian cancer samples analyzed in this study were high-grade. Currently, the standard first-line care for high-grade serous ovarian cancer remains surgical resection and platinum-based chemotherapy. We analyzed the tumor response to platinum-based chemotherapy. The results showed a significant difference in sensitivity to platinum-based chemotherapy between the high- and low-risk groups. Notably, resistance to platinum-based chemotherapy was present in both high- and low-risk groups, consistent with the assumption that a small proportion of ovarian cancer cells are chemo-resistant from the beginning [40]. Such inherent resistance in ovarian cancer may occur due to reduced immunosurveillance and drug-resistant cells that belong to cancer stem cells (CSCs) [7,40]. CSCs from the ovaries have a particular genetic signature that can engineer the original tumor mass, develop drug resistance, and promote recurrence. In addition to this possibility, platinum resistance may also emerge due to reduced intracellular drug accumulation, intracellular inactivation of the agent, increased DNA repair, or impaired apoptotic signaling pathways [41]. Therefore, the prediction of a tumor to platinum chemotherapy in ovarian cancer will provide further research directions in the future.

The study only used the retrospective data generated on a genome-wide scale. In the future, we will use real-time PCR (RT-PCR) techniques to validate SOV-P20 on a separate ovarian cancer patient cohort prospectively. We will recruit patients with ovarian cancer

in the First Affiliated Hospital of Gannan Medical University. After obtaining informed consent from each patient according to the protocol approved by the Ethics Committee of the hospital, the Department of Obstetrics and Gynecology will admit patients diagnosed with ovarian cancer between September 2021 and October 2023. Two hundred patients will be enrolled. Patient inclusion criteria include patients with a diagnosis of serous or endometrioid tumor. Specific guidelines are as follows: age 18 years or older; histologically confirmed epithelial ovarian, fallopian tube, or primary peritoneal cancer; FIGO stage I–IV; and planned to receive platinum-based combination chemotherapy. Patients will be followed up until 31 October 2028. After obtaining tissue samples, the RT-PCR techniques will determine the expression of individual mRNAs of SOV-P20 in patients to validate the predictive efficacy of SOV-P20.

## 5. Conclusions

Biomarker discovery based on within-sample REOs is a promising approach that can address the critical limitations of the risk-scoring-based model. The developed ovarian cancer prognostic model can easily predict patients with different survival prognoses. It is worth testing in prospective clinical trials to achieve its individualized management clinical effect in patients with serous ovarian cancer.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/diagnostics12123128/s1>, Table S1: Tumor response to chemotherapy drugs in TCGA serous ovarian cancer samples, Table S2: The prognosis-related genes in the training set, Table S3: SOV-P20 and its correlation with serous ovarian cancer, Table S4: Existing models for predicting serous ovarian cancer prognosis.

**Author Contributions:** G.H. and H.L. conceived the idea and conceptualized the study; F.L. and G.H. conducted the bioinformatics analysis and interpreted results; N.L., Q.Z., L.M., X.L. and T.H. collected and pre-processed data; F.L. and Q.Z. generated the figures and tables; F.L. and G.H. wrote the paper; H.L. supervised the whole study process; H.Z. helped to revise the manuscript; H.L. and G.H. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Natural Science Foundation of China (Grant No. 61961002) and the Thousand Talents Program of Jiangxi for High-level talents in innovation and entrepreneurship (No. Jxsq2020101096).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data were derived from the following resources available in GEO (<http://www.ncbi.nlm.nih.gov/geo/>, accessed on 11 October 2020) and the UC Santa Cruz Xena platform (UCSC Xena, <https://xenabrowser.net/datapages/>, accessed on 1 November 2020).

**Conflicts of Interest:** The authors declare no potential conflict of interest.

## Abbreviations

SOV-P20	Serous ovarian cancer prognostic biomarker consisting of 20 gene pairs
AUC	Area under the curve
C-index	Index of the concordance
REO	Relative expression ordering
OV	Ovarian cancer
GO	Gene ontology
GEO	Gene Expression Omnibus
TCGA	The Cancer Genome Atlas
KEGG	Kyoto Encyclopedia of Genes and Genomes
BH	Benjamini–Hochberg
PI3K-AKT	Phosphatidylinositol 3 kinase/protein kinase B
AGE-RAGE	Advanced glycation end products
TIMER	Tumor Immune Estimation Resource

UCSC	University of California Santa Cruz
RMA	Robust multi-array average
ROC	Receiver operating characteristic
GDSC	Genomics of Drug Sensitivity in Cancer
CR	Complete response
PR	Partial response
SD	Stable disease
PD	Progressive disease
CSC	Cancer stem cell
RT-PCR	Real-time PCR

## References

- Chen, H.; Zhou, S.; Sun, Z. Research status of the prevalence and burden of disease of three malignant tumors in common gynecology. *Chin. J. Mod. Med.* **2015**, *25*, 108–112.
- Zhang, S.; Xia, Q.; Zheng, R.; Chen, W. Analysis of ovarian cancer incidence and mortality in China in 2010. *China Oncol.* **2016**, *25*, 169–173.
- Colombo, N.; Sessa, C.; Du Bois, A.; Ledermann, J.; McCluggage, W.G.; McNeish, I.; Morice, P.; Pignata, S.; Ray-Coquard, I.; Vergote, I.; et al. ESMO–ESGO consensus conference recommendations on ovarian cancer: Pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease. *Ann. Oncol.* **2019**, *30*, 672–705. [[CrossRef](#)] [[PubMed](#)]
- Torre, L.A.; Trabert, B.; DeSantis, C.E.; Miller, K.D.; Samimi, G.; Runowicz, C.D.; Gaudet, M.M.; Jemal, A.; Siegel, R.L. Ovarian cancer statistics, 2018. *CA Cancer J. Clin.* **2018**, *68*, 284–296. [[CrossRef](#)]
- Wang, Y.K.; Bashashati, A.; Anglesio, M.S.; Cochrane, D.R.; Grewal, D.S.; Ha, G.; McPherson, A.; Horlings, H.M.; Senz, J.; Prentice, L.M.; et al. Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. *Nat. Genet.* **2017**, *49*, 856–865. [[CrossRef](#)]
- Wang, Z.; Zhang, Y.; Deng, Q.; Liang, H. Expression and mutations of BRCA in breast cancer and ovarian cancer: Evidence from bioinformatics analyses. *Int. J. Mol. Med.* **2018**, *42*, 3542–3550. [[CrossRef](#)]
- Laganà, A.S.; Colonese, F.; Colonese, E.; Sofo, V.; Salmeri, F.M.; Granese, R.; Chiofalo, B.; Ciancimino, L.; Triolo, O. Cytogenetic analysis of epithelial ovarian cancer’s stem cells: An overview on new diagnostic and therapeutic perspectives. *Eur. J. Gynaecol. Oncol.* **2015**, *36*, 495–505.
- Buttarelli, M.; De Donato, M.; Raspaglio, G.; Babini, G.; Ciucci, A.; Martinelli, E.; Baccaro, P.; Pasciuto, T.; Fagotti, A.; Scambia, G.; et al. Clinical Value of lncRNA MEG3 in High-Grade Serous Ovarian Cancer. *Cancers* **2020**, *12*, 966. [[CrossRef](#)]
- Cao, T.; Pan, W.; Sun, X.; Shen, H. Increased expression of TET3 predicts unfavorable prognosis in patients with ovarian cancer—a bioinformatics integrative analysis. *J. Ovarian Res.* **2019**, *12*, 101. [[CrossRef](#)]
- Feng, H.; Gu, Z.-Y.; Li, Q.; Liu, Q.-H.; Yang, X.-Y.; Zhang, J.-J. Identification of significant genes with poor prognosis in ovarian cancer via bioinformatical analysis. *J. Ovarian Res.* **2019**, *12*, 35. [[CrossRef](#)]
- Zheng, M.; Mullikin, H.; Hester, A.; Czogalla, B.; Heidegger, H.; Vilsmaier, T.; Vattai, A.; Chelariu-Raicu, A.; Jeschke, U.; Trillsch, F.; et al. Development and Validation of a Novel 11-Gene Prognostic Model for Serous Ovarian Carcinomas Based on Lipid Metabolism Expression Profile. *Int. J. Mol. Sci.* **2020**, *21*, 9169. [[CrossRef](#)] [[PubMed](#)]
- Qi, L.; Chen, L.; Li, Y.; Qin, Y.; Pan, R.; Zhao, W.; Gu, Y.; Wang, H.; Wang, R.; Chen, X.; et al. Critical limitations of prognostic signatures based on risk scores summarized from gene expression levels: A case study for resected stage I non-small-cell lung cancer. *Briefings Bioinform.* **2016**, *17*, 233–242. [[CrossRef](#)] [[PubMed](#)]
- Shi, P.; Ray, S.; Zhu, Q.; A Kon, M. Top scoring pairs for feature selection in machine learning and applications to cancer outcome prediction. *BMC Bioinform.* **2011**, *12*, 375. [[CrossRef](#)] [[PubMed](#)]
- Zhu, C.; Ding, K.; Strumpf, D.; Weir, B.A.; Meyerson, M.; Pennell, N.; Thomas, R.K.; Naoki, K.; Ladd-Acosta, C.; Liu, N.; et al. Prognostic and Predictive Gene Signature for Adjuvant Chemotherapy in Resected Non-Small-Cell Lung Cancer. *J. Clin. Oncol.* **2010**, *28*, 4417–4424. [[CrossRef](#)]
- Geman, D.; D’Avignon, C.; Naiman, D.Q.; Winslow, R.L. Classifying Gene Expression Profiles from Pairwise mRNA Comparisons. *Stat. Appl. Genet. Mol. Biol.* **2004**, *3*, 1–19. [[CrossRef](#)]
- Kadara, H.; Behrens, C.; Yuan, P.; Solis, L.; Liu, D.; Gu, X.; Minna, J.D.; Lee, J.J.; Kim, E.; Hong, W.-K.; et al. A Five-Gene and Corresponding Protein Signature for Stage-I Lung Adenocarcinoma Prognosis. *Clin. Cancer Res.* **2011**, *17*, 1490–1501. [[CrossRef](#)]
- Guo, Y.; Jiang, W.; Ao, L.; Song, K.; Chen, H.; Guan, Q.; Gao, Q.; Cheng, J.; Liu, H.; Wang, X.; et al. A qualitative signature for predicting pathological response to neoadjuvant chemoradiation in locally advanced rectal cancers. *Radiother. Oncol.* **2018**, *129*, 149–153. [[CrossRef](#)]
- Peng, F.; Wang, R.; Zhang, Y.; Zhao, Z.; Zhou, W.; Chang, Z.; Liang, H.; Zhao, W.; Qi, L.; Guo, Z.; et al. Differential expression analysis at the individual level reveals a lncRNA prognostic signature for lung adenocarcinoma. *Mol. Cancer* **2017**, *16*, 98. [[CrossRef](#)]
- Zhang, B.; Nie, X.; Miao, X.; Wang, S.; Li, J.; Wang, S. Development and verification of an immune-related gene pairs prognostic signature in ovarian cancer. *J. Cell. Mol. Med.* **2021**, *25*, 2918–2930. [[CrossRef](#)]

20. Guan, Q.; Zeng, Q.; Yan, H.; Xie, J.; Cheng, J.; Ao, L.; He, J.; Zhao, W.; Chen, K.; Guo, Y.; et al. A qualitative transcriptional signature for the early diagnosis of colorectal cancer. *Cancer Sci.* **2019**, *110*, 3225–3234. [[CrossRef](#)]
21. Song, K.; Guo, Y.; Wang, X.; Cai, H.; Zheng, W.; Li, N.; Song, X.; Ao, L.; Guo, Z.; Zhao, W. Transcriptional signatures for coupled predictions of stage II and III colorectal cancer metastasis and fluorouracil-based adjuvant chemotherapy benefit. *FASEB J.* **2019**, *33*, 151–162. [[CrossRef](#)] [[PubMed](#)]
22. Cirenajwis, H.; Lauss, M.; Planck, M.; Vallon-Christersson, J.; Staaf, J. Performance of gene expression-based single sample predictors for assessment of clinicopathological subgroups and molecular subtypes in cancers: A case comparison study in non-small cell lung cancer. *Brief. Bioinform.* **2020**, *21*, 729–740. [[CrossRef](#)] [[PubMed](#)]
23. Thissen, D.; Steinberg, L.; Kuang, D. Quick and Easy Implementation of the Benjamini-Hochberg Procedure for Controlling the False Positive Rate in Multiple Comparisons. *J. Educ. Behav. Stat.* **2002**, *27*, 77–83. [[CrossRef](#)]
24. Zhang, J.; Xu, M.; Gao, H.; Guo, J.; Guo, Y.; Zou, M.; Wu, X. Two protein-coding genes act as a novel clinical signature to predict prognosis in patients with ovarian serous cystadenocarcinoma. *Oncol. Lett.* **2018**, *15*, 3669–3675. [[CrossRef](#)] [[PubMed](#)]
25. Zhou, J.; Yi, Y.; Wang, C.; Su, C.; Luo, Y. Identification of a 3-mRNA signature as a novel potential prognostic biomarker in patients with ovarian serous cystadenocarcinoma in G2 and G3. *Oncol. Lett.* **2019**, *18*, 3545–3552. [[CrossRef](#)]
26. An, Y.; Bi, F.; You, Y.; Liu, X.; Yang, Q. Development of a Novel Autophagy-related Prognostic Signature for Serous Ovarian Cancer. *J. Cancer* **2018**, *9*, 4058–4071. [[CrossRef](#)]
27. Wang, R.; Ye, X.H.; Zhao, X.L.; Liu, J.L.; Zhang, C.Y. Development of a five-gene signature as a novel prognostic marker in ovarian cancer. *Neoplasma* **2019**, *66*, 343–349. [[CrossRef](#)]
28. Wang, L.; Li, X. Identification of an energy metabolism-related gene signature in ovarian cancer prognosis. *Oncol. Rep.* **2020**, *43*, 1755–1770. [[CrossRef](#)]
29. Zhang, L.; Zhu, P.; Tong, Y.; Wang, Y.; Ma, H.; Xia, X.; Zhou, Y.; Zhang, X.; Gao, F.; Shu, P. An immune-related gene pairs signature predicts overall survival in serous ovarian carcinoma. *OncoTargets Ther.* **2019**, *ume 12*, 7005–7014. [[CrossRef](#)]
30. Waldmann, T.A. Cytokines in Cancer Immunotherapy. *Cold Spring Harb. Perspect. Biol.* **2018**, *10*, a028472. [[CrossRef](#)]
31. Gonzalez, V.D.; Huang, Y.-W.; Delgado-Gonzalez, A.; Chen, S.-Y.; Donoso, K.; Sachs, K.; Gentles, A.J.; Allard, G.M.; Kolahi, K.S.; Howitt, B.E.; et al. High-grade serous ovarian tumor cells modulate NK cell function to create an immune-tolerant microenvironment. *Cell Rep.* **2021**, *36*, 109632. [[CrossRef](#)] [[PubMed](#)]
32. Lin, W.W.; Karin, M. A cytokine-mediated link between innate immunity, inflammation, and cancer. *J. Clin. Investig.* **2007**, *117*, 1175–1183. [[CrossRef](#)] [[PubMed](#)]
33. Hu, L.; Hofmann, J.; Lu, Y.; Mills, G.B.; Jaffe, R.B. Inhibition of phosphatidylinositol 3'-kinase increases efficacy of paclitaxel in vitro and in vivo ovarian cancer models. *Cancer Res.* **2002**, *62*, 1087–1092. [[PubMed](#)]
34. Biffo, S.; Manfrini, N.; Ricciardi, S. Crosstalks between translation and metabolism in cancer. *Curr. Opin. Genet. Dev.* **2018**, *48*, 75–81. [[CrossRef](#)] [[PubMed](#)]
35. Li, X.; Wang, Q.; Wu, Z.; Zheng, J.; Ji, L. Integrated Bioinformatics Analysis for Identification of the Hub Genes Linked with Prognosis of Ovarian Cancer Patients. *Comput. Math. Methods Med.* **2022**, *2022*, 5113447. [[CrossRef](#)]
36. Ko, Y.S.; Pyo, J.-S. Clinicopathological significance and prognostic role of tumor-infiltrating lymphocytes in colorectal cancer. *Int. J. Biol. Markers* **2019**, *34*, 132–138. [[CrossRef](#)]
37. Zhang, Q.-F.; Li, Y.-K.; Chen, C.-Y.; Zhang, X.-D.; Cao, L.; Quan, F.-F.; Zeng, X.; Wang, J.; Liu, J. Identification and validation of a prognostic index based on a metabolic-genomic landscape analysis of ovarian cancer. *Biosci. Rep.* **2020**, *40*, bsr20201937. [[CrossRef](#)]
38. Zhang, L.; Zhang, Y.; Wang, S.; Tao, L.; Pang, L.; Fu, R.; Fu, Y.; Liang, W.; Li, F.; Jia, W. MiR-212-3p suppresses high-grade serous ovarian cancer progression by directly targeting MAP3K3. *Am. J. Transl. Res.* **2020**, *12*, 875–888.
39. Yang, J.; Hong, S.; Zhang, X.; Liu, J.; Wang, Y.; Wang, Z.; Gao, L.; Hong, L. Tumor Immune Microenvironment Related Gene-Based Model to Predict Prognosis and Response to Compounds in Ovarian Cancer. *Front. Oncol.* **2021**, *11*, 807410. [[CrossRef](#)]
40. Laganà, A.S.; Sofo, V.; Vitale, S.G.; Triolo, O. Epithelial ovarian cancer inherent resistance: May the pleiotropic interaction between reduced immunosurveillance and drug-resistant cells play a key role? *Gynecol. Oncol. Rep.* **2016**, *18*, 57–58. [[CrossRef](#)]
41. Kim, H.; Xu, H.; George, E.; Hallberg, D.; Kumar, S.; Jagannathan, V.; Medvedev, S.; Kinose, Y.; Devins, K.; Verma, P.; et al. Combining PARP with ATR inhibition overcomes PARP inhibitor and platinum resistance in ovarian cancer models. *Nat. Commun.* **2020**, *11*, 3726. [[CrossRef](#)] [[PubMed](#)]