



Comprehensive pathway-related genes signature for prognosis and recurrence of ovarian cancer

Xinnan Zhao¹ and Miao He²

¹Department of Rheumatology and Immunology, The First Affiliated Hospital of China Medical University, Shenyang, China

²Department of Pharmacology, China Medical University, Shenyang, China

ABSTRACT

Background. Ovarian cancer (OC) is a highly malignant disease with a poor prognosis and high recurrence rate. At present, there is no accurate strategy to predict the prognosis and recurrence of OC. The aim of this study was to identify gene-based signatures to predict OC prognosis and recurrence.

Methods. mRNA expression profiles and corresponding clinical information regarding OC were collected from The Cancer Genome Atlas (TCGA) database. Gene set enrichment analysis (GSEA) and LASSO analysis were performed, and Kaplan–Meier curves, time-dependent ROC curves, and nomograms were constructed using R software and GraphPad Prism7.

Results. We first identified several key signalling pathways that affected ovarian tumorigenesis by GSEA. We then established a nine-gene-based signature for overall survival (OS) and a five-gene-based-signature for relapse-free survival (RFS) using LASSO Cox regression analysis of the TCGA dataset and validated the prognostic value of these signatures in independent GEO datasets. We also confirmed that these signatures were independent risk factors for OS and RFS by multivariate Cox analysis. Time-dependent ROC analysis showed that the AUC values for OS and RFS were 0.640, 0.663, 0.758, and 0.891, and 0.638, 0.722, 0.813, and 0.972 at 1, 3, 5, and 10 years, respectively. The results of the nomogram analysis demonstrated that combining two signatures with the TNM staging system and tumour status yielded better predictive ability.

Conclusion. In conclusion, the two-gene-based signatures established in this study may serve as novel and independent prognostic indicators for OS and RFS.

Subjects Bioinformatics, Gynecology and Obstetrics, Oncology, Women's Health, Medical Genetics

Keywords Ovarian cancer, GSEA, LASSO, Overall survival, Recurrence-free survival

INTRODUCTION

Ovarian cancer (OC), one of the most common female malignant tumours, is the fifth-most common cause of mortality in women with severe gynaecological issues, with an estimated 13,940 deaths and 21,750 new cases reported in the 2020 United States cancer statistics (Siegel & Miller, 2020). Because OC is usually diagnosed at a late stage and currently lacks effective treatment options, patients with OC have extremely poor prognosis and only a

Submitted 13 July 2020
Accepted 6 November 2020
Published 1 December 2020

Corresponding author
Miao He, hemiao_cmu@163.com

Academic editor
Vladimir Uversky

Additional Information and
Declarations can be found on
page 14

DOI 10.7717/peerj.10437

© Copyright
2020 Zhao and He

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

30–40% 5-year survival rate (Enroth *et al.*, 2019). Despite the moderate improvements that have been made in the diagnosis and treatment of OC over the past 30 years, OC is still a highly malignant disease that is life-threatening (Li *et al.*, 2019). In addition, OC patients have an average relapse-free survival (RFS) of two to three years after first-line therapy and die due to chemotherapy resistance (Dinh *et al.*, 2008). To date, several common biomarkers for diagnosing OC remain ineffective due to lack of sensitivity and specificity; these include carbohydrate antigen 125 (CA125, AUC = 0.630) (Mansha, Gill & Thomson, 2019) and human epididymis protein 4 (HE4, AUC = 0.719) (Scaletta *et al.*, 2017). At the same time, a number of studies have identified certain genes that are significantly associated with OC patient prognosis, such as *TRIM44* (Wang *et al.*, 2018) and *CENPK* (Lee *et al.*, 2015). However, there is very limited prognostic value in a single candidate biomarker, which is attributed to inconsistent sample collection, detection methods, and small sample sizes. A number of studies have shown that combined biomarkers could improve prognostic accuracy compared to single biomarkers (Barata & Rini, 2017). Thus, extensive studies have attempted to establish molecular signatures based on gene expression to predict survival of patients, including mRNA- (Zhou *et al.*, 2019), microRNA- (Li *et al.*, 2019), and long non-coding RNA-based signatures (Xu *et al.*, 2019). However, only a small number of prognostic signatures have been developed, and none have been directly applied in clinical practice. Hence, there is an urgent need to identify and validate novel, highly sensitive, and specific molecular biomarkers for prognosis, monitoring, and therapy improvement.

In this study, we sought to explore signalling pathways related to ovarian tumorigenesis and attempted to identify the potential predictive abilities of genes participating in these pathways for OC prognosis and recurrence. We established a nine-gene signature panel for overall survival (OS) and a five-gene signature group for RFS using LASSO Cox regression analysis. Both signatures have powerful prognostic predictive abilities and high sensitivity and specificity. Our results revealed that both signatures were independent risk factors for OC survival and recurrence. In addition, we constructed two nomogram models encompassing signature, TNM stage, and tumour status for OS and RFS, respectively, and found that both nomogram models exhibited high potential clinical utility by decision curve analysis (DCA). In general, we constructed two reliable and superior gene signature panels as prognostic predictors of OC patient survival and recurrence.

MATERIALS AND METHODS

Data collection

The Cancer Genome Atlas (TCGA) –OC expression profiles (HTSeq-Counts) of genes and corresponding clinical information were downloaded from the TCGA database (<http://cancergenome.nih.gov>) (Tomczak, Czerwinska & Wiznerowicz, 2015). Detailed information is presented in Table S1. Other datasets, including GSE40595, GSE12470, GSE10971, GSE27651, GSE38666, and GSE17260, were downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). Detailed information is shown in Tables S2–S3.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) (<http://www.broadinstitute.org/gsea/index.jsp>) was implemented using the JAVA.8 program using the Molecular Signatures Database (*Subramanian et al., 2005*). The hallmark gene sets were considered as a reference gene set and $P < 0.05$ was considered statistically significant.

Establishment of prognostic gene signatures for overall survival and relapse-free survival with LASSO Cox regression model

For gene sets in activated signalling pathways in OC, we first conducted univariate Cox regression analysis to identify genes significantly associated with OS and RFS in patients. Furthermore, LASSO Cox regression analysis was performed to construct optimal prognostic-predicting signatures for OS and RFS in patients by using the above candidate genes based on the “glmnet” R package. The LASSO penalty model successfully achieves shrinkage and selects variables simultaneously. The optimal values of the penalty parameter λ are determined through 10-fold cross-validations. Here, gene-based signatures for OS and RFS were established based on the expression of each prognostic gene and its coefficient, respectively. The risk score formula = $\beta_1 \text{ gene}_1 * \text{expression of gene}_1 + \beta_2 \text{ gene}_2 * \text{expression of gene}_2 + \dots + \beta_n \text{ gene}_n * \text{expression of gene}_n$, where β indicates the LASSO coefficients of these genes.

Gene signatures independent of other clinical characteristics

Univariate and multivariate Cox regression analyses were implemented to appraise the predictive abilities of gene signatures and clinical characteristics using SPSS v.22.0. The clinical characteristics in TCGA-OC included age, TNM stage, race, tumour status, and so on.

Clinical application of gene signatures

Nomograms can be used to predict the survival rates in diseases with multiple indicators (*Iasonos et al., 2008*). To determine the clinical application of gene signatures, we generated an optimal nomogram model using the “rms”, “nomogramEx” and “regplot” R packages to evaluate the 1-, 3- and 5-year OS and RFS of OC patients. DCA was performed to assess the clinical usefulness of the risk prediction nomogram (*Vickers & Elkin, 2006*). The x -axis represented the percentage of threshold probability, and the y -axis represented the net benefit.

Transcription factor analysis and GO enrichment analysis

Transcription factor analysis was produced using a website (<https://www.gcbi.com.cn/gcanalyze/html/generadar/index>). The clusterProfiler package was implemented to further explore the biological function of the transcription factor. $P < 0.05$ was considered a significant enrichment.

Statistical analysis

OS and RFS differences between patients in high-risk and low-risk groups were estimated by Kaplan–Meier survival curve analysis and calculated using log-rank tests. Time-dependent receiver operating characteristic (time-dependent ROC) curves were constructed to

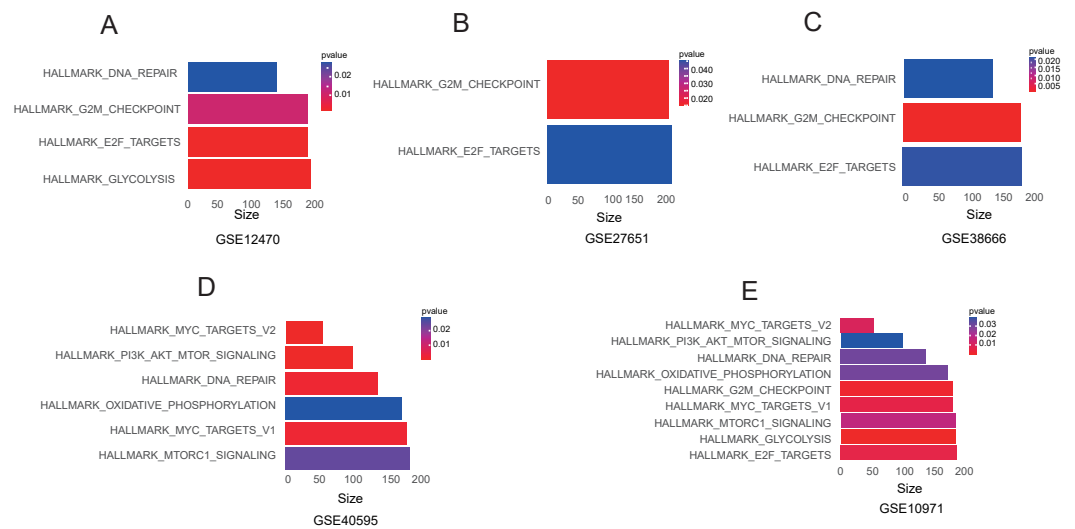


Figure 1 The GSEA enrichment analysis of different GEO datasets. (A) GSE12470. (B) GSE27651. (C) GSE38666. (D) GSE40595. (E) GSE10971.

Full-size DOI: 10.7717/peerj.10437/fig-1

determine the sensitivity and specificity of gene signatures by calculating the area under curve (AUC) using the “survivalROC” R package (Do, Le & Le, 2020; Le, 2019; Le et al., 2019). We compared two groups using Student’s *t*-test for numerical variables and Chi-squared tests for categorical variables. All statistical graphs were produced using R (<https://www.r-project.org/>, v3.5.1), SPSS v.22.0 (SPSS Inc., Chicago, IL) and GraphPad Prism7 (GraphPad Software Inc., La Jolla, CA).

RESULTS

Identification of signalling pathways correlated with OC tumorigenesis

To determine related signalling pathways involved in ovarian tumorigenesis, we performed GSEA on Hallmark gene sets for five independent GEO datasets, GSE40595, GSE12470, GSE10971, GSE27651, and GSE38666. The results showed that there are multiple signalling pathways significantly enriched in OC including “MYC_TARGETS_V1”, “MTORC1_SIGNALING”, “GLYCOLYSIS”, “E2F_TARGETS”, “G2M_CHECKPOINT”, “OXIDATIVE_PHOSPHORYLATION”, “DNA_REPAIR” and “PI3K_AKT_MTOR_SIGNALING” (Figs. 1A–1E and Table S4), suggesting that these signalling pathways might play critical roles in the development of OC.

Establishment of gene signatures for overall survival and relapse-free survival

To evaluate the effect of genes in these pathways on OS and RFS, we performed univariate Cox regression analysis using TCGA data. The results showed that 48 and 44 genes were correlated with patient OS and RFS, respectively. Furthermore, the LASSO Cox selection method was implemented to establish a prognosis-predicting model (OS: Figs. 2A, 2B; RFS:

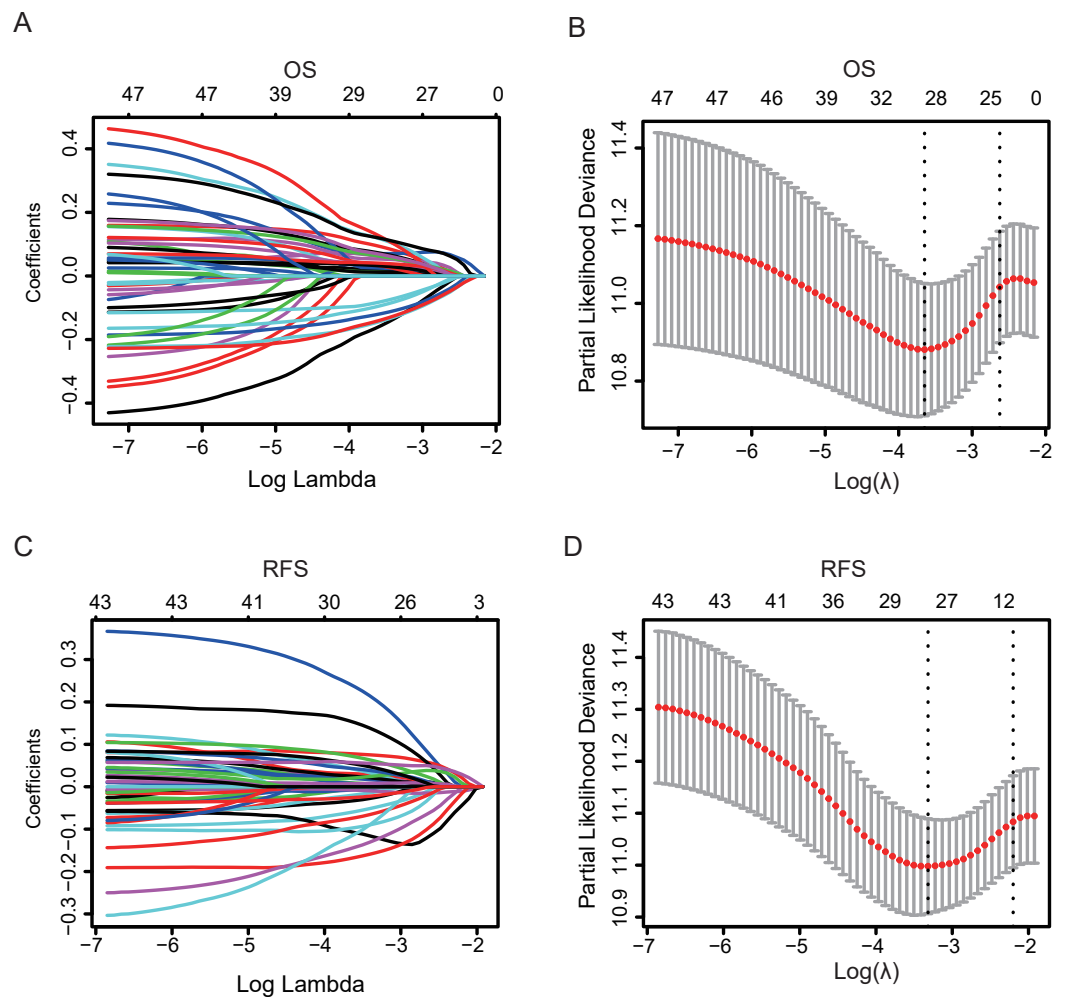


Figure 2 The LASSO Cox regression model for OS and RFS of patients in the TCGA dataset. The LASSO coefficients profiles of prognosis-related genes for OS (A) and RFS (C). Tuning parameter (λ) selection cross-validation error curve for OS (B) and RFS (D). The number of lines above the graph represent the number of genes involved in the LASSO analysis, with the change of the Log lambda, the most representative genes were selected.

Full-size DOI: [10.7717/peerj.10437/fig-2](https://doi.org/10.7717/peerj.10437/fig-2)

[Figs. 2C, 2D](#)). Finally, a nine-gene signature panel for OS and a five-gene signature group for RFS were established by LASSO Cox regression analysis. The formula of the nine-gene signature panel for OS = $(-0.0027) * \text{expression of } HMGB3 + 0.0122 * \text{expression of } PYGB + 0.0094 * \text{expression of } IDUA + 0.0092 * \text{expression of } LHX9 + (-0.0254) * \text{expression of } SLC7A11 + (-0.0037) * \text{expression of } DPYSL4 + 0.0017 * \text{expression of } ANGPTL4 + (-0.0357) * \text{expression of } ISG20 + (-0.0353) * \text{expression of } GRM8$; and the formula of the five-gene signature group for RFS = $(-0.0114) * \text{expression of } SNRPA1 + (-0.0346) * \text{expression of } WARS + 0.0234 * \text{expression of } CITED2 + 0.002 * \text{expression of } ANGPTL4 + 0.0357 * \text{expression of } EFNA5$ ([Table S5](#)). Next, we calculated risk scores for the OS and RFS of each patient based on the above formula. The distribution of risk scores, survival times, patient status, and expression heat maps of genes were obtained. The results showed that

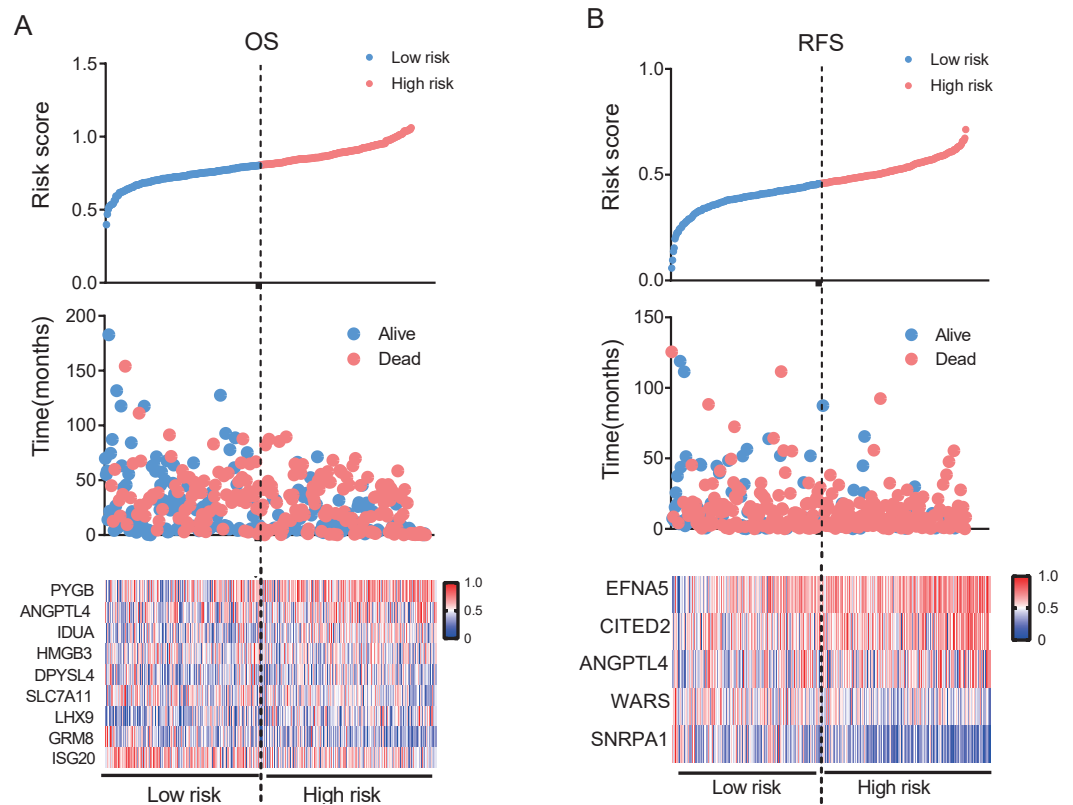


Figure 3 Establishment of gene signature for OS and RFS in the TCGA dataset. Risk score distribution, patients' survival and status and expression heatmap of genes for OS (A) and RFS (B).

Full-size [DOI: 10.7717/peerj.10437/fig-3](https://doi.org/10.7717/peerj.10437/fig-3)

patients in the high-risk group had shorter OS and RFS and higher mortality than those in the low-risk group (Figs. 3A, 3B and Figs. S1A, S1B). According to these two signatures, patients were divided into high-risk or low-risk groups using the median risk score as the cut-off point (OS: median = 0.863; RFS: median = 0.457). We performed Kaplan–Meier analysis and found that patients with high risk scores had significantly shorter OS and RFS than those with low risk scores (OS: $P < 0.0001$, HR = 2.289, 95% CI [1.724–3.04]; RFS: $P < 0.0001$, HR = 1.783, 95% CI [1.465–2.264]; Figs. 4A, 4B). Furthermore, to explore the sensitivity and specificity of the risk score, we performed time-dependent ROC analysis. The results showed that for OS, the prognostic accuracy of the nine-gene signature panel was 0.640, 0.663, 0.758, and 0.891 at 1, 3, 5, and 10 years, respectively. The AUCs for RFS were 0.638, 0.722, 0.813, and 0.972 at 1, 3, 5, and 10 years, respectively (Figs. 4C, 4D). These results demonstrate that the nine-gene signature panel is a helpful indicator of OS for OC patients and that the five-gene signature group is a powerful prognostic predictor of OC patient recurrence.

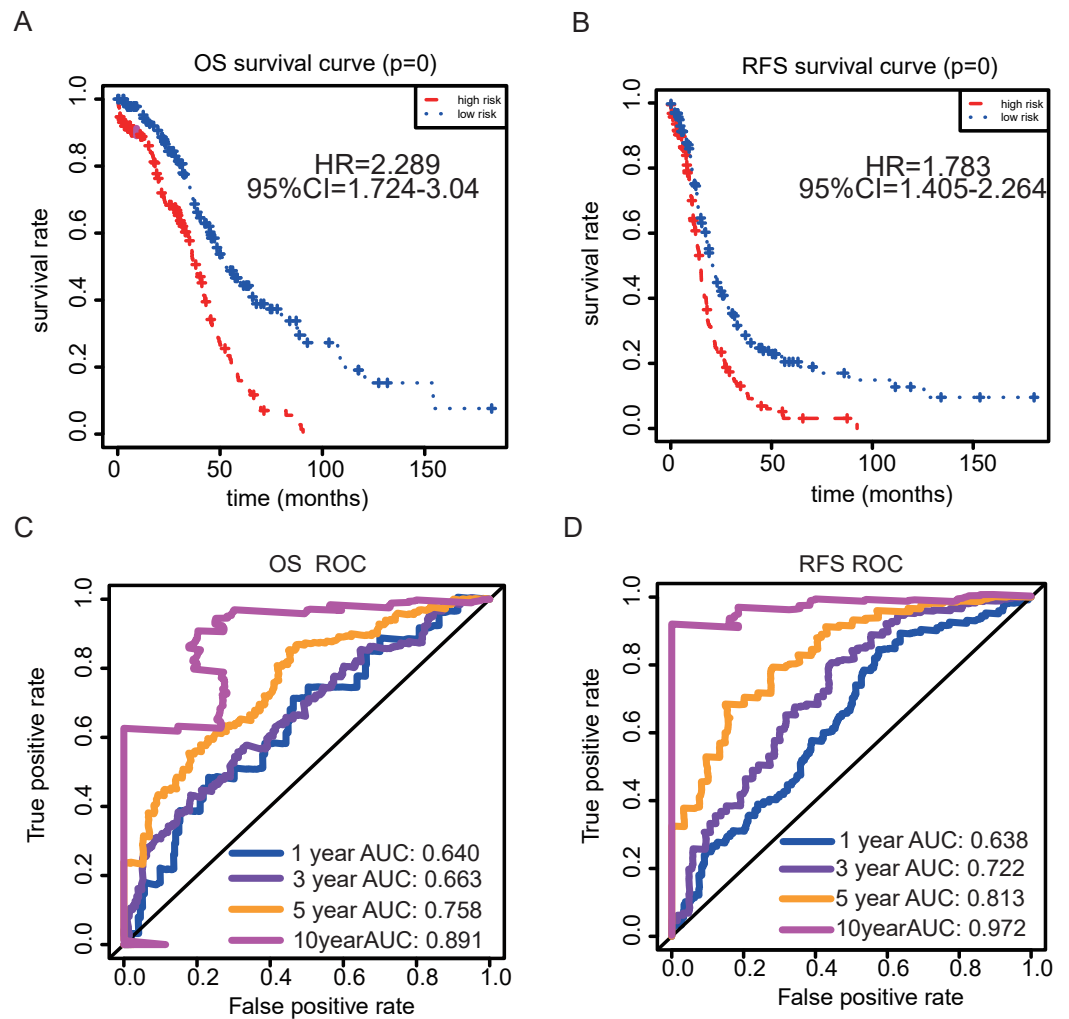


Figure 4 Kaplan–Meier curve analysis and Time-dependent ROC curve analysis. Kaplan–Meier curve analysis for OS (A) and RFS (B). Time-dependent ROC curve analysis of the sensitivity and specificity for OS (C) and RFS (D).

Full-size DOI: [10.7717/peerj.10437/fig-4](https://doi.org/10.7717/peerj.10437/fig-4)

Validation of the nine-gene signature panel for overall survival and five-gene signature group for relapse-free survival

To verify the prognostic value of our signatures, we investigated the independent dataset [GSE17260](#) for validation of OS and RFS. We also performed Kaplan–Meier analysis and observed the same results as above, which showed that patients in the high-risk group had shorter OS and RFS and higher mortality than those in the low-risk group based on our nine- and five-gene-based signatures, respectively ([Figs. 5A, 5B](#)). In addition, we performed cross validation to prove the consistency of these signatures. We found the same gene sets in [GSE17260](#) dataset ([Figs. S2A, S2B](#)), the only difference is the coefficients of each gene have changed a little. Furthermore, we also performed Kaplan–Meier analysis and observed the same results in TCGA dataset ([Figs. S2C, S2D](#)).

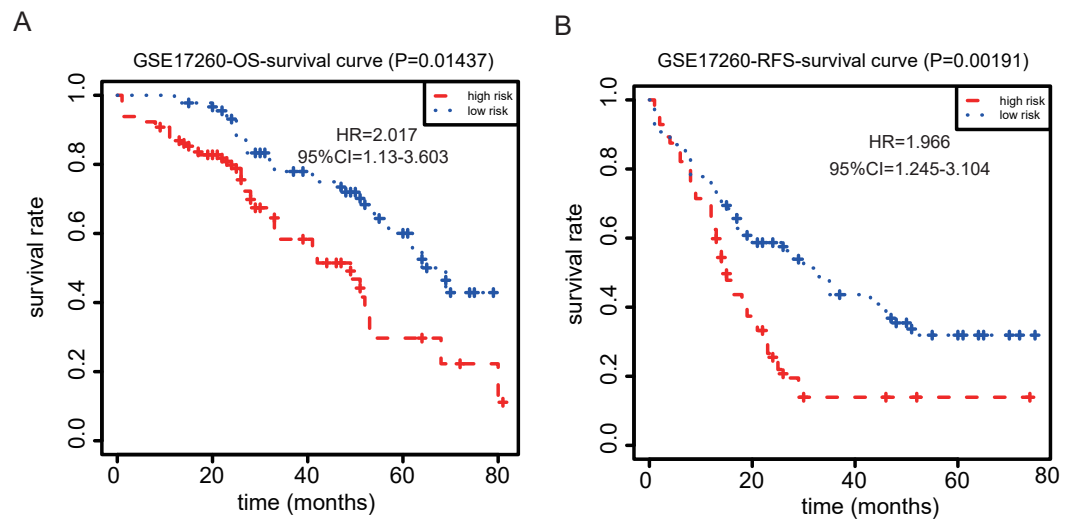


Figure 5 Kaplan–Meier curve analysis. Kaplan-Meier curve analysis for OS (A) and RFS (B) in the GSE17260 dataset.

Full-size DOI: 10.7717/peerj.10437/fig-5

The signature perform better in survival prediction than other signature

Then, the ROC curves were implemented to contrast prognostic ability between our signature and other signature (risk score = $0.38 \times \text{XPC} - 0.24 \times \text{PALB2} + 0.29 \times \text{RECQL} - 0.18 \times \text{XRCC2} + 0.32 \times \text{GTF2H5} - 0.19 \times \text{GTF2H4} - 0.22 \times \text{SSBP1} + 0.24 \times \text{RAD54L} - 0.25 \times \text{MUTYH} - 0.3 \times \text{SMUG1} - 0.16 \times \text{TDP1} - 0.24 \times \text{DDB2} + 0.26 \times \text{RNH1} + 0.18 \times \text{TP53BP1}$) (Sun *et al.*, 2019). The result of ROC analysis illustrated that our signature has significantly higher ROC (AUC = 0.891) than other signature (AUC = 0.751) (Fig. S3). This result demonstrated that predictive power of our signature is greater than other signature. In general, our signature could be used as a superior indicator to predict the prognosis of OV patients.

Relationships involving gene signatures and clinical characteristics

To further investigate potential correlations between risk scores and patient clinical characteristics, we conducted Chi-squared tests. We found that for OS, the risk score was significantly associated with race and tumour status. Similarly, for RFS, the risk score was significantly correlated with TNM stage and tumour status. However, for both OS and RFS, there was no significant correlation between risk score and other clinical characteristics, including age and tumour grade (Tables 1 and 2).

Prognostic value of gene signatures is independent of other clinical characteristics

To further evaluate the predictive ability of gene signatures and other clinical characteristics in terms of OS and RFS, univariate and multivariate Cox analyses were performed. As shown in Figs. 6A and 6B, the nine-gene signatures, stage, race, and tumour status were significantly correlated with OS, as seen in the univariate Cox regression analysis.

Table 1 Correlation between risk score and clinicopathological features of OV patients for OS.

Characteristics	N	Risk score level		P value ^a
		Low	High	
Age(years)				0.518
>65	114	60	54	
≤65	239	117	122	
TNM stage				0.534
I and II	21	12	9	
III and IV	329	165	164	
Race				0.033
Non-white	36	12	24	
White	307	160	147	
Tumor status				0.029
Tumor free	79	48	31	
With tumor	228	106	122	
Grade				0.092
G1+G2	42	16	26	
G3+G4	302	157	145	
Lymphatic invasion				0.721
NO	43	22	21	
YES	94	45	49	

Notes.

Bold, significant values < 0.05.

^aP values were calculated by X² test.

Furthermore, multivariate Cox regression analysis results showed that the nine-gene signatures and tumour status were independent prognostic factors for OS. Meanwhile, univariate Cox regression analysis revealed that the five-gene signatures, staging, and tumour status were significantly associated with RFS. The five-gene signatures and tumour status were independent prognostic factors for RFS after multivariate Cox regression analysis.

Clinical application of gene signatures

We further constructed a nomogram predictive model combining gene signatures, race, stage, and tumour status to predict the probability of 1-, 3-, and 5-year OS and RFS in OC patients (Figs. 7A and 7C). The results showed that the nomogram model combined gene signatures, stage, and tumour status (OS: C-index: 0.743; RFS: C-index: 0.712). These results demonstrate that our gene signature-based nomogram approach exhibits superior performance in clinical settings. Furthermore, DCA was conducted to evaluate the clinical usefulness of the risk prediction nomogram for OS and RFS (Figs. 7B and 7D). The DCA results showed that for OS and RFS, the complex factor had a broader range of threshold probabilities and higher net benefits than single clinical characteristics or risk scores. Taken together, the DCA results demonstrate that our nomogram for OS and RFS has high potential clinical utility.

Table 2 Correlation between risk score and clinicopathological features of OV patients for RFS.

Characteristics	N	Risk score level		P value ^a
		Low	High	
Age(years)				0.623
>65	114	55	59	
≤65	239	122	117	
TNM stage				0.046
I and II	21	15	6	
III and IV	329	161	168	
Race				0.738
Non-white	36	19	17	
White	307	153	154	
Tumor status				0.002
Tumor free	79	51	28	
With tumor	228	102	126	
Grade				0.175
G1+G2	42	17	25	
G3+G4	302	156	146	
Lymphatic invasion				0.171
NO	43	26	17	
YES	94	45	49	

Notes.

Bold, significant values < 0.05.

^aP values were calculated by X² test.

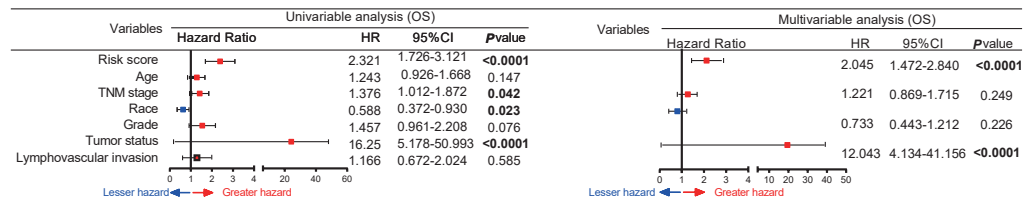
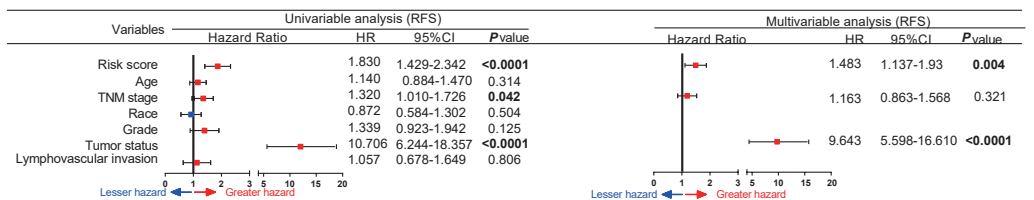
A**B**

Figure 6 Prognostic value of gene signature is independent of other clinical characteristics. Forest plot summary of univariate and multivariate Cox regression analysis of gene signature and clinical characteristics on OS (A) and RFS (B).

Full-size DOI: 10.7717/peerj.10437/fig-6

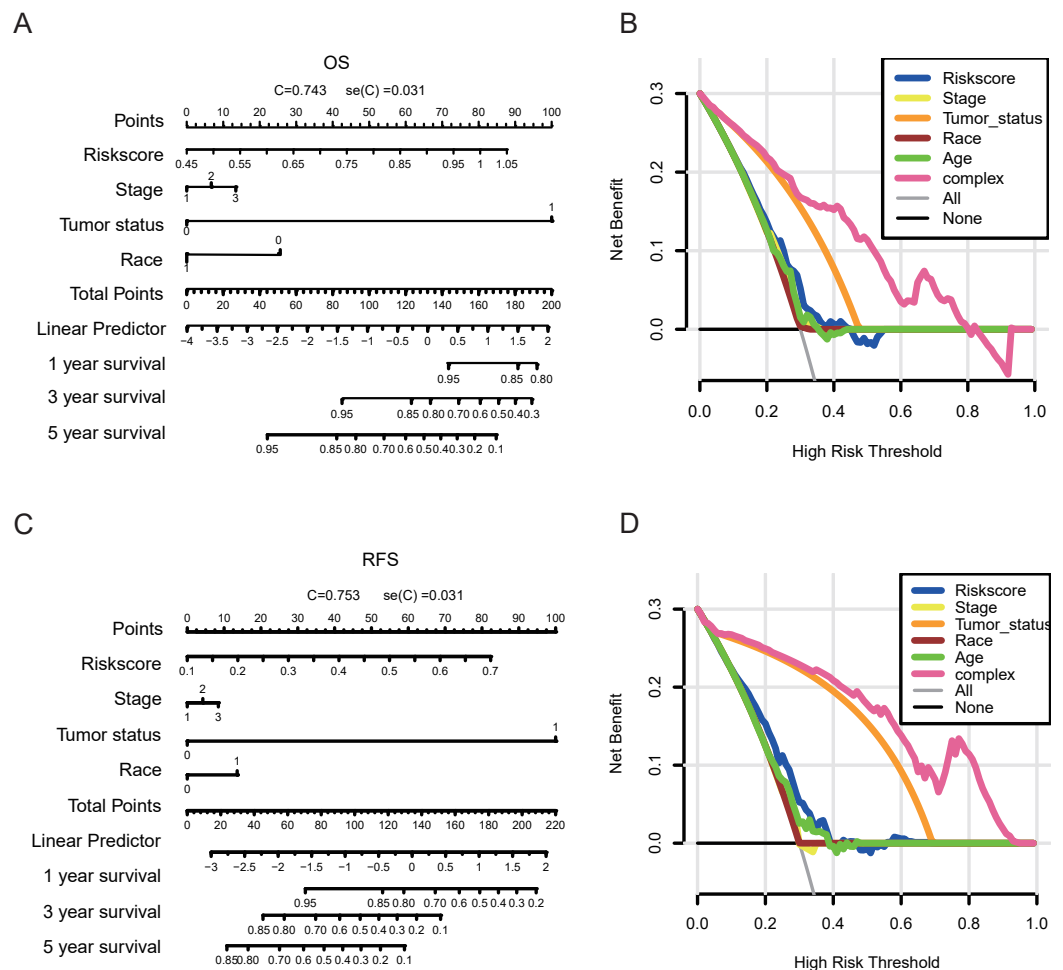


Figure 7 Nomogram to predict 1-, 3- and 5-year OS and RFS for ovarian cancer patients. The nomogram for predicting survival rate of patients with 1-, 3- and 5-year OS (A) and RFS (C) (stage: 0 = stage I, 1 = stage II, 2 = stage III, 3 = stage IV; tumor status: 0 = tumor free, 1 = with tumor). Decision analysis curve for OS (B) and RFS (D).

Full-size [DOI: 10.7717/peerj.10437/fig-7](https://doi.org/10.7717/peerj.10437/fig-7)

Transcription factor analysis and GO enrichment analysis

Transcription factors are a family of DNA-binding proteins whose gene regulatory capabilities are of vital importance in defining the molecular state of a cell. Therefore, we performed Transcription factor analysis of genes above using website (<https://www.gcbi.com.cn/gcanalyze/html/generadar/index>). To explore the potential function of these transcription factor, we performed GO enrichment analysis by clusterprofiler package. The GO analysis results showed that these transcription factor analysis were mainly enriched in "cell fate commitment", "pattern specification process", "embryonic organ development" and so on (Fig. S4).

DISCUSSION

To a certain extent, indolent and aggressive cancers cannot be accurately stratified by the TNM staging system, which mostly relies on anatomical information without any molecular biological information or features (*Choudhary et al., 2019; McIntyre et al., 2017*). Therefore, it is indispensable to identify novel and effective predictive biomarkers for OC prognosis. In this study, we established and validated novel prognostic gene-based signatures for OS and RFS to improve the prediction of survival and tumour recurrence in OC patients.

Increasing number of studies are gradually clarifying the key signalling pathways involved in cancer initiation, progression, and metastasis, including the PI3K-Akt signalling pathway (*Mabuchi et al., 2015*), MYC signalling pathway (*Dang, 2012*) and so on. Moreover, the genes participating in these key signalling pathways might play critical roles in tumorigenesis. Therefore, it is necessary to explore the functions and roles of these key genes. Consistently, extensive studies have confirmed several signalling pathway-related genes that are significantly associated with the prognosis of cancer patients. *Cao et al. (2019)* confirmed that EMT-related genes promote bladder cancer subtype transition from non-muscle invasive bladder cancer (NMIBC) to muscle invasive bladder cancer (MIBC) and affect the prognosis of bladder cancer patients. In this study, our candidate genes for OS and RFS were screened for signalling pathways affecting the development of OC, including "MYC_TARGETS_V1", "MTORC1_SIGNALING", "GLYCOLYSIS", "E2F_TARGETS" and "G2M_CHECKPOINT". MYC is one of the most highly amplified oncogenes in many human cancers and plays a pivotal role in tumorigenesis (*Dang, 2012*). Mammalian target of rapamycin complex 1 (mTORC1) signalling is found to be elevated in many human cancers, including OC, and promotes cell growth, proliferation, and metabolism (*Menon & Manning, 2008*). Increasing evidence suggests that aerobic glycolysis is related to tumour growth and chemotherapy resistance (*Chakraborty et al., 2017*). The E2F family plays a significant role in cancer initiation, progression, and resistance to therapy (*Kent & Leone, 2019*) and E2Fs can mediate fundamental cell-cycle deregulation in high-grade serous OC (*DeMeyer et al., 2009*). Hence, it is a plausible inference for these genes to be associated with the survival of patients with clear cell renal cell carcinoma (ccRCC).

High-dimensional data survival models have a high risk of overfitting, which is characterised by reduced importance when applying predictors to independent datasets. To solve this problem, we applied a Cox proportional hazard model using the LASSO penalty method to optimise gene selection. This is useful for selecting genes with strong prognostic value, high expression differences, and low correlation among each other (*Tibshirani, 1997*). The potential of LASSO has been previously implemented in colon cancer (*Dai et al., 2018*), bladder cancer (*He et al., 2019*) and other studies. Therefore, in this study, we successfully applied this method and established a nine-gene-based signature panel for OS and a five-gene-based signature group for RFS to optimise the predictive ability of prognosis in OC patients.

Although extensive studies have advocated several molecular biomarkers for risk stratification in OC, along with traditional clinicopathologic parameters, there are still some points to be improved as follows: (1) small sample sizes need to be enlarged; (2)

independent validation datasets are needed to further confirm the stability and accuracy of the established signatures; (3) it is indispensable to compare the predictive power between signatures and the existing TNM staging system; and (4) it is more valuable to explore prognostic biomarkers for OC recurrence. In our study, we established novel superior prognostic signatures for OS and RFS in the TCGA dataset and validated the prognostic value of these signatures using an independent GEO dataset, suggesting the repeatability and practicability of our two signatures. Then, we confirmed that these signatures were independent risk factors for OS and RFS by univariate and multivariate Cox analyses. In addition, we constructed a nomogram model to better predict survival of OC patients and validated its promising clinical application by DCA. In general, our findings provided a superior and useful biomarker for OS and RFS in OC patients, based on the results of the LASSO regression analysis.

Our study demonstrated that nine-gene- and five-gene-based signatures for OS and RFS, respectively, were both closely associated with the prognosis of OC patients. However, the functions of most of the genes in our signatures have not been entirely elucidated. For genes associated with OS, the chromatin-associated high-mobility group box 3 (HMGB3) protein is upregulated and has poor prognosis, and its targeted depletion can attenuate cisplatin resistance in human OC cells ([Mukherjee et al., 2019](#)). *PYGB* has been shown to be upregulated in OC tissue and markedly associated with poor prognosis in OC patients ([Zhou, Jin & Wang, 2019](#)). *LHX9* is a biomarker in ovarian stem cells ([Auersperg, 2013](#)). *SLC7A11* is an independent prognostic risk factor for OS in OC ([Yin et al., 2019](#)). Low levels of *DPYSL4* are significantly correlated with poor prognosis in OC ([Nagano et al., 2018](#)). In addition, [Jiang et al. \(2019\)](#) suggested that *DPYSL4* is significantly associated with OS in hepatocellular carcinoma. High expression of *ISG20* is related to poor clinical outcomes in glioblastoma ([Gao et al., 2019](#)). *GRM8* is a tumour-suppressor gene in endometrial cancer cell lines ([Liang et al., 2012](#)). For genes associated with RFS, *SNRPA1* is an oncogene in colorectal cancer ([Zeng et al., 2019](#)). [Yanagie et al. \(2009\)](#) demonstrated that CBP/p300-interacting transactivator (*CITED2*) is potentially a multidrug-resistance gene, which is consistent with our results. *ANGPTL4* is a novel biomarker of hypoxia in OC, and is upregulated in partial responders to chemotherapy compared to responders ([McEvoy et al., 2015](#)). [Yang et al. \(2019\)](#) found that *EFNA5* can be used as a novel prognostic biomarker and potential therapeutic target for OC patients.

Inevitably, there are certain innate limitations to our study that need to be addressed. First, screening for these signatures was based on data from TCGA and GEO datasets without validating their prognostic value in clinical practice. Second, although these candidate genes are involved in key signalling pathways that influence prognosis in OC, the mechanisms underlying some genes in our signature panels remains unclear. Undoubtedly, further investigation of the roles of these genes in OC progression is needed. It is reported that blacks vs whites were over two-fold less likely to receive a surgery-chemotherapy, suggesting that race may affect survival status of ovarian cancer patients greatly, which was consistent with our results. In addition, higher socioeconomic status is associated with a greater probability of undergoing surgical resection and with improved survival in patients with ovarian cancer, demonstrating socioeconomic status might also affect

prognosis of patients. Unfortunately, several important clinical characteristics, such as treatment modality and socioeconomic status are not available in the TCGA cohort. Finally, prospective, large-scale, multicentre studies are indispensable to verify the prognostic ability of our signatures before applying them in clinical settings.

CONCLUSIONS

In conclusion, we established a nine-gene-based signature panel for OS, and a five-gene-based signature group for RFS using a LASSO Cox regression model. These signatures could be potentially novel and independent prognostic indicators for predicting survival and disease recurrence in OC patients. Moreover, combining these signatures with current clinical indicators may serve to optimise their predictive ability and guide future clinical planning for patient monitoring.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Xinnan Zhao conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Miao He conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data are available from TCGA (Ovarian cancer) and GEO ([GSE40595](#), [GSE12470](#), [GSE10971](#), [GSE27651](#), [GSE38666](#), and [GSE17260](#)).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.10437#supplemental-information>.

REFERENCES

- Auersperg N. 2013.** The stem-cell profile of ovarian surface epithelium is reproduced in the oviductal fimbriae-with increased stem-cell marker density in distal parts of the fimbriae. *International Journal of Gynecological Pathology* 32:444–453
[DOI 10.1097/PGP.0b013e3182800ad5](https://doi.org/10.1097/PGP.0b013e3182800ad5).

- Barata PC, Rini BI. 2017.** Treatment of renal cell carcinoma: current status and future directions. *CA: A Cancer Journal for Clinicians* **67**:507–524.
- Cao R, Yuan L, Ma B, Wang G, Qiu W, Tian Y. 2019.** An EMT-related gene signature for the prognosis of human bladder cancer. *Journal of Cellular and Molecular Medicine* **24**:605–617 DOI [10.1111/jcmm.14767](https://doi.org/10.1111/jcmm.14767).
- Chakraborty PK, Mustafi SB, Xiong X, Dwivedi SKD, Nesin V, Saha S, Zhang M, Dhanasekaran D, Jayaraman M. 2017.** MICU1 drives glycolysis and chemoresistance in ovarian cancer. *Nature Communications* **8**:14634 DOI [10.1038/ncomms14634](https://doi.org/10.1038/ncomms14634).
- Choudhary D, Sharma G, Singh T, Chishi N, Garg PK. 2019.** Isolated abdominal wall metastasis in early ovarian cancer: a true systemic metastasis or local tumour cells implantation. *Journal of Obstetrics and Gynaecology* **39**:724–726 DOI [10.1080/01443615.2019.1578734](https://doi.org/10.1080/01443615.2019.1578734).
- Dai W, Li Y, Mo S, Feng Y, Zhang L, Xu Y, Li Q, Cai G. 2018.** A robust gene signature for the prediction of early relapse in stage I-III colon cancer. *Molecular Oncology* **12**:463–475 DOI [10.1002/1878-0261.12175](https://doi.org/10.1002/1878-0261.12175).
- Dang CV. 2012.** MYC on the path to cancer. *Cell* **149**:22–35 DOI [10.1016/j.cell.2012.03.003](https://doi.org/10.1016/j.cell.2012.03.003).
- Dinh P, Harnett P, Piccart-Gebhart MJ, Awada A. 2008.** New therapies for ovarian cancer: cytotoxics and molecularly targeted agents. *Critical Reviews in Oncology/Hematology* **67**:103–112 DOI [10.1016/j.critrevonc.2008.01.012](https://doi.org/10.1016/j.critrevonc.2008.01.012).
- Do DT, Le TQT, Le NQK. 2020.** Using deep neural networks and biological subwords to detect protein S-sulfenylation sites. *Briefings in Bioinformatics* Epub ahead of print July 02 2020.
- Enroth S, Berggrund M, Lycke M, Broberg J, Lundberg M, Assarsson E, Olovsson M, Stalberg K, Sundfeldt K, Gyllensten U. 2019.** High throughput proteomics identifies a high-accuracy 11 plasma protein biomarker signature for ovarian cancer. *Communications Biology* **2**:221 DOI [10.1038/s42003-019-0464-9](https://doi.org/10.1038/s42003-019-0464-9).
- Gao Y, Wang Y, Wang X, Wang Y, Zhang X, Sun X. 2019.** TNF-like ligand 1A is associated with progression and prognosis of human gastric cancer. *OncoTargets and Therapy* **12**:7715–7723 DOI [10.2147/OTT.S210939](https://doi.org/10.2147/OTT.S210939).
- He A, He S, Peng D, Zhan Y, Li Y, Chen Z, Gong Y, Li X, Zhou L. 2019.** Prognostic value of long non-coding RNA signatures in bladder cancer. *Aging* **11**:6237–6251 DOI [10.18632/aging.102185](https://doi.org/10.18632/aging.102185).
- Iasonos A, Schrag D, Raj GV, Panageas KS. 2008.** How to build and interpret a nomogram for cancer prognosis. *Journal of Clinical Oncology* **26**:1364–1370 DOI [10.1200/JCO.2007.12.9791](https://doi.org/10.1200/JCO.2007.12.9791).
- Jiang L, Zhao L, Bi J, Guan Q, Qi A, Wei Q, He M, Wei M, Zhao L. 2019.** Glycolysis gene expression profilings screen for prognostic risk signature of hepatocellular carcinoma. *Aging* **11**:10861–10882 DOI [10.18632/aging.102489](https://doi.org/10.18632/aging.102489).
- Kent LN, Leone G. 2019.** The broken cycle: E2F dysfunction in cancer. *Nature Review Cancer* **19**:326–338.
- Le NQK. 2019.** Fertility-GRU: identifying fertility-related proteins by incorporating deep-gated recurrent units and original position-specific scoring matrix profiles. *Journal of Proteome Research* **18**:3503–3511 DOI [10.1021/acs.jproteome.9b00411](https://doi.org/10.1021/acs.jproteome.9b00411).

- Le NQK, Yapp EKY, Nagasundaram N, Chua MCH, Yeh HY. 2019. Computational identification of vesicular transport proteins from sequences using deep gated recurrent units architecture. *Computational and Structural Biotechnology Journal* 17:1245–1254 DOI [10.1016/j.csbj.2019.09.005](https://doi.org/10.1016/j.csbj.2019.09.005).
- Lee YC, Huang CC, Lin DY, Chang WC, Lee KH. 2015. Overexpression of centromere protein K (CENPK) in ovarian cancer is correlated with poor patient survival and associated with predictive and prognostic relevance. *PeerJ* 3:e1386 DOI [10.7717/peerj.1386](https://doi.org/10.7717/peerj.1386).
- Li L, Gu H, Chen L, Zhu P, Zhao L, Wang Y, Zhao X, Zhang X, Zhang Y, Shu P. 2019. Integrative network analysis reveals a microRNA-based signature for prognosis prediction of epithelial ovarian cancer. *BioMed Research International* 2019:1056431.
- Liang H, Cheung LW, Li J, Ju Z, Yu S, Stemke-Hale K, Dogruluk T, Lu Y, Liu X, Gu C, Guo W, Scherer SE, Carter H, Westin SN, Dyer MD, Verhaak RG, Zhang F, Karchin R, Liu CG, Lu KH, Russell R, Broaddus KL, Hennessy BT, Mills GB. 2012. Whole-exome sequencing combined with functional genomics reveals novel candidate driver cancer genes in endometrial cancer. *Genome Research* 22:2120–2129 DOI [10.1101/gr.137596.112](https://doi.org/10.1101/gr.137596.112).
- Mabuchi S, Kuroda H, Takahashi R, Sasano T. 2015. The PI3K/AKT/mTOR pathway as a therapeutic target in ovarian cancer. *Gynecologic Oncology* 137:173–179.
- Mansha M, Gill A, Thomson PC. 2019. Potential risk factors of ovarian cancer and analysis of CA125, a biomarker used for its monitoring and diagnosis. *Molecular Biology Reports* 46:3325–3332 DOI [10.1007/s11033-019-04794-8](https://doi.org/10.1007/s11033-019-04794-8).
- McEvoy LM, O’Toole SA, Spillane CD, Martin CM, Gallagher MF, Stordal B, Blackshields G, Sheils O, O’Leary JJ. 2015. Identifying novel hypoxia-associated markers of chemoresistance in ovarian cancer. *BMC Cancer* 15:547 DOI [10.1186/s12885-015-1539-8](https://doi.org/10.1186/s12885-015-1539-8).
- McIntyre JB, Rambau PF, Chan A, Yap S, Morris D, Nelson GS, Kobel M. 2017. Molecular alterations in indolent, aggressive and recurrent ovarian low-grade serous carcinoma. *Histopathology* 70:347–358 DOI [10.1111/his.13071](https://doi.org/10.1111/his.13071).
- Menon S, Manning BD. 2008. Common corruption of the mTOR signaling network in human tumors. *Oncogene* 27(Suppl 2):S43–S51 DOI [10.1038/onc.2009.352](https://doi.org/10.1038/onc.2009.352).
- De Meyer T, Bijsmans IT, Van de Vijver KK, Bekaert S, Oosting J, Van Criekinge W, van Engeland M, Sieben NL. 2009. E2Fs mediate a fundamental cell-cycle deregulation in high-grade serous ovarian carcinomas. *Journal of Pathology* 217:14–20 DOI [10.1002/path.2452](https://doi.org/10.1002/path.2452).
- Mukherjee A, Huynh V, Gaines K, Reh WA, Vasquez KM. 2019. Targeting the high-mobility group box 3 protein sensitizes chemoresistant ovarian cancer cells to cisplatin. 79:3185–3191.
- Nagano H, Hashimoto N, Nakayama A, Suzuki S, Miyabayashi Y, Yamato A, Higuchi S, Fujimoto M, Sakuma I, Beppu M, Yokoyama M, Suzuki Y, Sugano S, Ikeda K, Tatsuno I, Manabe I, Yokote K, Inoue S, Tanaka T. 2018. p53-inducible DPYSL4 associates with mitochondrial supercomplexes and regulates energy metabolism in adipocytes and cancer cells. 115:8370–8375.

- Scaletta G, Plotti F, Luvero D, Capriglione S, Montera R, Miranda A, Lopez S, Terranova C, De Cicco Nardone C, Angioli R. 2017.** The role of novel biomarker HE4 in the diagnosis, prognosis and follow-up of ovarian cancer: a systematic review. *Expert Review of Anticancer Therapy* **17**:827–839 DOI [10.1080/14737140.2017.1360138](https://doi.org/10.1080/14737140.2017.1360138).
- Siegel RL, Miller KD. 2020.** Cancer Statistics. *CA-A Cancer Journal for Clinicians* **70**:7–30 DOI [10.3322/caac.21590](https://doi.org/10.3322/caac.21590).
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005.** Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**:15545–15550 DOI [10.1073/pnas.0506580102](https://doi.org/10.1073/pnas.0506580102).
- Sun H, Cao D, Ma X, Yang J, Peng P, Yu M, Zhou H, Zhang Y, Li L, Huo X, Shen K. 2019.** Identification of a prognostic signature associated with DNA repair genes in ovarian cancer. *Frontiers in Genetics* **10**:839 DOI [10.3389/fgene.2019.00839](https://doi.org/10.3389/fgene.2019.00839).
- Tibshirani R. 1997.** The lasso method for variable selection in the Cox model. *Statistics in Medicine* **16**:385–395 DOI [10.1002/\(SICI\)1097-0258\(19970228\)16:4<385::AID-SIM380>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3).
- Tomczak K, Czerwinska P, Wiznerowicz M. 2015.** The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary Oncology [Wspolczesna Onkologia]* **19**:A68–A77.
- Vickers AJ, Elkin EB. 2006.** Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making* **26**:565–574 DOI [10.1177/0272989X06295361](https://doi.org/10.1177/0272989X06295361).
- Wang S, Jia M, He Z, Liu XS. 2018.** APOBEC3B and APOBEC mutational signature as potential predictive markers for immunotherapy response in non-small cell lung cancer. *Oncogene* **37**:3924–3936 DOI [10.1038/s41388-018-0245-9](https://doi.org/10.1038/s41388-018-0245-9).
- Xu L, Wu Y, Che X, Zhao J, Wang F, Wang P, Qu X, Liu Y, Li Z. 2019.** Cox-LASSO analysis reveals a ten-lnc RNA signature to predict outcomes in patients with high-grade serous ovarian cancer. *DNA and Cell Biology* **38**:1519–1528 DOI [10.1089/dna.2019.4826](https://doi.org/10.1089/dna.2019.4826).
- Yanagie H, Hisa T, Ogata A, Miyazaki A, Nonaka Y, Nishihira T, Osada I, Sairennji T, Sugiyama H, Furuya Y, Kidani Y, Takamoto S, Takahashi H, Eriguchi M. 2009.** Improvement of sensitivity to platinum compound with siRNA knockdown of upregulated genes in platinum complex-resistant ovarian cancer cells in vitro. *Biomedicine and Pharmacotherapy* **63**:553–560 DOI [10.1016/j.biopha.2008.04.006](https://doi.org/10.1016/j.biopha.2008.04.006).
- Yang L, Bie L, Sun L, Yue Y. 2019.** Neural activities are unfavorable for the prognosis of ovarian cancer through mRNA expression analysis. *Biomarkers in Medicine* **13**:663–673 DOI [10.2217/bmm-2019-0056](https://doi.org/10.2217/bmm-2019-0056).
- Yin F, Yi S, Wei L, Zhao B, Li J, Cai X, Dong C, Liu X. 2019.** Microarray-based identification of genes associated with prognosis and drug resistance in ovarian cancer. **120**:6057–6070.
- Zeng Q, Lei F, Chang Y, Gao Z, Wang Y, Gao Q, Niu P, Li Q. 2019.** An oncogenic gene, SNRPA1, regulates PIK3R1, VEGFC, MKI67, CDK1 and other

genes in colorectal cancer. *Biomedicine and Pharmacotherapy* **117**:109076
[DOI 10.1016/j.biopha.2019.109076](https://doi.org/10.1016/j.biopha.2019.109076).

Zhou J, Yi Y, Wang C, Su C, Luo Y. 2019. Identification of a 3-mRNA signature as a novel potential prognostic biomarker in patients with ovarian serous cystadenocarcinoma in G2 and G3. *Oncology Letters* **18**:3545–3552.

Zhou Y, Jin Z, Wang C. 2019. Glycogen phosphorylase B promotes ovarian cancer progression via Wnt/beta-catenin signaling and is regulated by miR-133a-3p. *Biomedicine and Pharmacotherapy* **120**:109449 [DOI 10.1016/j.biopha.2019.109449](https://doi.org/10.1016/j.biopha.2019.109449).