

A novel six-gene signature for prognosis prediction in ovarian cancer

Xin Pan

Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang, 110021, PR China

Xiao-xin Ma (✉ maxx@sj-hospital.org)

Shengjing Hospital of China Medical University

Research

Keywords: ovarian cancer, EMT, prognostic, mRNAs, survival

Posted Date: February 19th, 2020

DOI: <https://doi.org/10.21203/rs.2.23976/v1>

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

[Read Full License](#)

Abstract

Ovarian cancer (OC) is the most malignant tumor in the female reproductive tract. Although abundant molecular biomarkers have been identified, a robust and accurate gene expression signature is still essential to assist oncologists in evaluating the prognosis of ovarian cancer patients. In the current study, 367 patients were adopted through The Cancer Genome Atlas (TCGA) database, and mRNA expression profiling was performed. Then, we used a gene set enrichment analysis (GSEA) to screen genes correlated with the epithelial-to-mesenchymal transition (EMT) process and identify these genes with the Cox proportional regression model. Six genes (TGFB1, SFRP1, COL16A1, THY1, PPIB, BGN) associated with overall survival (OS) were used to construct a risk assessment model, by which the patients were divided into high-risk and low-risk groups. The six-gene signature was identified as an independent prognostic biomarker of the OS of ovarian cancer patients via multivariate Cox regression analysis. Besides, the six-gene model was also validated significantly by Gene Expression Omnibus (GEO) database. In summary, we established a six-gene signature relevant to the prognosis of ovarian cancer, which might become a therapeutic target with clinical usefulness in the future.

1. Background

OC has become the fifth leading cause of death and is becoming a major threat to the reproductive health of women. According to a previous report, 80% of OC patients are initially diagnosed with advanced OC[1, 2], and 28% of these patients show distant metastasis[3]. Although standard therapies aiming at complete resection combined with chemotherapy have been widely used in clinical practice and various neoadjuvant therapies are emerging, the prognosis of OC is poor, and the overall five-year survival rate is only 30–50%[4]. The traditional FIGO (2014) staging system is the dominant assessment criterion for estimating the treatment and outcome of OC patients. However, individuals in the same stage and receiving the same therapy, meet different prognoses[5, 6]. Thus, the current staging system is inadequate, and it is urgent to seek more significant indicators to identify the high-risk population. There were significant differences in five-year survival rates among ovarian cancer patients at different stages, so we grouped the samples based on different stages.

With the rapid development of high-throughput gene microarrays and whole-genome-sequencing, numerous molecular landscapes and gene variants have been discovered in many tumors[7, 8]. However, it is time-consuming and laborious to find key genome-wide mutations by experiments. Therefore, developing a convenient and effective programmatic screening method for high-risk populations is of vital importance[9].

Currently, a large number of databases have been built by researchers to provide access to explore genomic alterations. Increasing evidence demonstrates that molecular biomarkers contribute to promoting the prognosis evaluation of tumors. For example, upregulated delta-like protein 3 (DLL3) overexpression is an independent prognostic predictor for endometrial cancer and could be a potential and novel tumor marker for the early diagnosis of endometrial cancer[10]. Plasma Tie2 is a tumor

vascular response biomarker for vascular endothelial growth factor (VEGF) inhibitors in metastatic colorectal cancer[11]. Nevertheless, researchers found that rather than single gene biomarkers, gene signatures that include several genes can be good evidence for the prognosis and survival of tumors. For example, a 5-gene signature (ADM, ASPM, DCBLD2, E2F7, and KRT6A) demonstrates significant power to predict patient survival in two distinct patient cohorts and is independent of American Joint Committee on Cancer (AJCC) TNM staging in pancreatic ductal adenocarcinoma[12]. Cross-validation of this gene signature reported a better area under curve (AUC) of receiver operating characteristic (ROC) (≥ 0.8) when compared to existing pancreatic ductal adenocarcinoma (PDAC) survival signatures. Gene signatures may provide insights into the mechanism of carcinogenesis to facilitate a preferable treatment schedule.

In this study, we used GSEA to screen target genes, and this technique is superior to traditional analysis. GSEA is not limited to focusing on gene expression differences in disease biomarkers and analyzing survival changes, that is, this technique only focuses on statistically significant genes with aberrant expression[13]. For Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, researchers need to provide a clear threshold for differentially expressed genes. However, the actual changes in RNA expression observed by Chromatin Immunoprecipitation (ChIP) are usually the result of multiple negative feedback regulation, and the sensitivity of expressive discrepancy is different from one tissue to another. GSEA can avoid ignoring those genes that have no significant expression difference but contribute to biological behavior, gene function and even related gene regulation networks. In addition, the method of gene set enrichment analysis monitors the expression changes of gene sets rather than individual genes, so these subtle expression changes can be included to obtain more ideal results[14, 15].

In our study, we attempted to produce enrichment gene sets based on molecular signatures with prognosis of OC in different stages and profiled specific gene sets in 367 OC patients with integrated mRNA expression datasets from the TCGA database. A total of 197 mRNAs were related to the epithelial mesenchymal transition, and a six-gene risk signature was established, which can predict patient outcomes and identify high-risk OC populations that indicate poor prognosis.

2. Materials And Methods

Patient Clinical Data Source and mRNA Expression Dataset

The mRNA expression and corresponding clinical information for OC patients were extracted from the TCGA data portal (<https://cancergenome.nih.gov/>), which was imputed on an Illumina HiSeq RNA-Seq platform, including 1 OC patient in stage I, 22 in stage II, 289 in stage III and 55 in stage IV with their matching stage, age, cancer status, grade, venous invasion and lymphatic invasion. The abovementioned data are displayed in Table 1. Both profiles and clinical characteristics can be obtained publicly, so there is no need to provide ethics committee approval. Furthermore, hierarchical cluster analysis was used to show the volcano plot. GSE9891 was download from the GEO database

(<http://www.ncbi.nlm.nih.gov/gds>) and was used to validate the prognostic prediction significance between different risk groups of six-gene model furtherly.

Table 1
Clinical pathological parameters of patients with ovary cancer in this study.

Clinical pathological parameters	N	%	Dead number
Stage			
Stage I	1	0.27	1
Stage II	22	5.99	5
Stage III	289	78.74	159
Stage IV	55	14.98	34
Cancer Status			
Tumor Free	85	23.22	3
With Tumor	281	76.78	196
Grade			
G1	1	0.27	0
G2	42	11.48	25
G3	314	85.79	167
G4	1	0.27	1
GX	8	2.186	5
Age			
≤58	183	50	92
>58	183	50	107
Venous invasion			
No	40	38.83	17
Yes	63	61.17	23
Lymphatic invasion			
No	47	41.23	19
Yes	97	58.77	45

Gene Set Enrichment Analysis

We used GSEA (<http://www.broadinstitute.org/gsea/index.jsp>) to determine whether the specific sets of genes showed different expression levels statistically. The expression of 24991 mRNAs among diversified stages of OC were analyzed, taking the standard of $p < 0.05$ to screen the biological process for further investigation.

Functional Enrichment Analysis

We used the GO and the KEGG to explore the pathways of biological processes that are enriched by prognostic-related molecules. The results of the two strategies were substantially a channel for the overall and deep understanding of biological systems for the GSEA analysis in our study. A protein-protein interaction (PPI) network was developed to explore the relationships between these genes using the online database STRING (<https://string-db.org/>). STRING was used to map all of the hub genes connected with each other.

RNA isolation and Quantitative Real-Time polymerase chain reaction (qRT-PCR)

In brief, total RNA was extracted by TRIzol reagent (Takara, Dalian, China), cDNA for mRNAs analysis was generated from total RNA using the PrimeScript RT-polymerase (Takara). qRT-PCR was performed using SYBR-Green Premix (Takara) with specific PCR primers (Sangon Biotech Co.,Ltd, Shanghai, China). Primers for TGFBI was F: ACTCAGCCAAGACACTATTTGA, R: CTTGTATGGGCATCAATTGGAG; Primers for SFRP1 was F: GCTCAACAAGAACTGCCAC, R: CTTGTCACACTTAAGCATCTCG; Primers for COL16A1 was F: GGAAGGACTCAAATTGGAACAC, R: GATCTTCTTGATGGCAGACGTC; Primers for THY1 was F: CCAACTTCACCAGCAAATACAA, R: ACTTGACCAGTTTGTCTCTGAG; Primers for PPIB was F: TTCTTCATCACGACAGTCAAGA, R: TCACATCCTTCAGGGGTTTATC; Primers for BGN was F: GAACATGAACTGCATCGAGATG, R: ATTTTGTTGTGGTCTAGGTGGA.

Glyceraldehyde-3-phosphate Dehydrogenase (GAPDH) were used as internal controls and fold-changes were calculated by means of the $2^{-\Delta\Delta Ct}$ method.

Statistical Analysis

The expression profile of 24991 mRNAs was shown with their normalization by log₂ transformation. Twenty-four enrichment biological processes were screened, and the top-ranking process EMT was selected. Then, Cox analysis was used for further studies. Univariate Cox regression analysis was launched to screen the genes correlated to OS with $p < 0.05$. Subsequently, we used multivariate Cox proportional hazards regression analysis to verify the genes closely related to the clinical prognosis among the genes we obtained following the above procedure. ROC curve analysis was carried out to compare the predictive accuracy of the gene sets. Thereafter, a prognostic risk score formula was established on the basis of a linear combination of the expression levels weighted with regression coefficients originating from the multivariate Cox regression analysis after all of the genes were divided into risky ($HR > 1$) and protective ($0 < HR < 1$) groups.

$$\text{Risk score} = \sum_{i=1}^n \text{Exp}_i * \beta_i,$$

where n is the number of selected genes, Exp_i is the expression level of gene i , and β_i represents the regression coefficient of gene i . A total of 367 patients were classified into high-risk and low-risk groups according to the median risk coefficient as the boundary. Risk factors were equally weighed with common clinicopathological parameters to validate the prognostic significance by the Kaplan-Meier curves and the log-rank method. The expression levels of the chosen genes in normal and cancer tissues were confirmed by Student's t-test, and the alteration of optimal genes was displayed in various pathologic types of OCs online (<http://www.cbioportal.org/>). SPSS 16.0 and GraphPad Prism 7 were used for the statistical analysis in this study. A p value < 0.05 was set as the significant difference statistically.

Data of qRT-PCR are expressed as the means \pm standard deviation (SD) of 3 independent experiments. Statistical analyses were performed with GraphPad Prism 6.0 software (La Jolla, CA). A 2-sided Student's t test or 1-way analysis of variance (ANOVA) was used to ascertain differences between the 2 groups since the data exhibited the same normal distribution. A $P < 0.05$ was considered statistically significant.

3. Results

The flowchart of this study is shown in Fig. 1. We divided the 367 patients into two groups according to the stage from the TCGA dataset. A prognostic model was established through the EMT related genes and COX analysis was performed, which indicated that the risk model was an independent prognostic indicator of OC.

Gene Set Enrichment Analysis by GSEA

Out of a total of 24991 genes detected, 634 were upregulated, while 543 were downregulated (Fig. 2A). There were also 367 specimens from TCGA classified into two groups (OC patients in stage I-II and OC patients in stage III-IV). GSEA revealed that 24 out of the hallmark biological processes were selected for enrichment, for example, epithelial mesenchymal transition (EMT) ($p \neq 0$), oxidative phosphorylation ($p \neq 0$), adipogenesis ($p \neq 0$), myogenesis ($p \neq 0$), coagulation ($p = 0.003$), apoptosis ($p = 0.037$), and fatty acid metabolism ($p = 0.038$), which embodied significant differences between OC patients in stage I-II and OC patients in stage III-IV (Fig. 2B, Table 2). $\text{FDR} < 0.05$ and gene size ≥ 100 were considered as the cut-off criteria. The top-ranking gene set indicating the epithelial mesenchymal transition process ($p < 0.001$), including 197 mRNAs, was selected for further studies.

Table 2
Gene sets enriched in ovary cancer (367 samples).

GS follow link to MSigDB	SIZE	ES	NOM P-value	Rank at MAX
EPITHELIAL MESENCHYMAL TRANSITION	197	0.6	0	5691
OXIDATIVE PHOSPHORYLATION	183	0.53	0	7891
ADIPOGENESIS	190	0.41	0	10392
MYOGENESIS	199	0.39	0	10785
COAGULATION	136	0.39	0.003	10967
APOPTOSIS	159	0.32	0.037	5216
FATTY ACID METABOLISM	157	0.32	0.038	10137

Identification of a Six-mRNA Prognostic Signature Predicts Survival of OC Patients

Eighty-three genes were selected because their expression profile was enriched in stage III–IV OC. A heatmap of the 83 common differentially expressed genes in the good and poor prognosis groups is shown in Additional file 1: Fig. S1. Then, a total of 19 prognosis-associated mRNAs were obtained using univariate Cox regression analysis, and six of these mRNAs (TGFB1, SFRP1, COL16A1, THY1, PPIB, BGN) were selected for multivariate Cox regression analysis to construct a risk assessment model. We assigned these mRNAs as 4 risky ($HR > 1$) and 2 protective ($0 < HR < 1$) mRNAs (Table 3).

Table 3

The detailed information of five prognostic mRNAs significantly associated with overall survival in patients with stomach adenocarcinoma.

mRNA	Ensemble ID	Location	HR	B(Cox)	P
TGFB1	ENSG00000120708	Chr5: 136,028,988 – 136,063,818	1.1324	0.1243	0.0175
SFRP1	ENSG00000104332	Chr 8: 41,261,962 – 41,309,473	1.0707	0.0683	0.0217
COL16A1	ENSG00000084636	Chr1: 31,652,263 – 31,704,319 re	1.1528	0.1422	0.0249
THY1	ENSG00000154096	Chr11: 119,417,378 – 119,424,985	1.0853	0.0819	0.0504
PPIB	ENSG00000166794	Chr15: 64,155,812 – 64,163,205	0.7530	-0.2837	0.0069
BGN	ENSG00000182492	Chr X: 153,494,980 – 153,509,546	0.8216	-0.1965	0.0393

A prognostic model was developed to predict prognosis according to the gene expression and regression coefficients of the six genes. Risk score = $0.1243 \times \text{expression of TGFB1} + 0.0683 \times \text{expression of SFRP1} + 0.1422 \times \text{expression of COL16A1} + 0.0819 \times \text{expression of THY1} - 0.2837 \times \text{expression of PPIB}$

0.1965*expression of BGN. After every patient was endowed with a risk score, the median risk score value was regarded as the cut-off criterion, and the patients were divided into low-risk and high-risk groups (Fig. 3A). The distribution of the patient relapse status is also shown in Fig. 3A. The mortality increased with increasing risk score among these patients. A heat map (Fig. 3B) was used to display the six mRNA expression profiles. The expression of risky type mRNAs (TGFB1, SFRP1, COL16A1, THY1) was obviously upregulated as the risk score of OC patients increased, and the expression of protective type mRNAs (PPIB, BGN) was downregulated.

The cBioPortal database was used to detect alterations in the six selected genes in 367 patients. The results indicated that TGFB1 included 0.9% alteration, with 2 examples of amplification, 2 examples of deep deletions and 1 example of a missense mutation (unknown significance). SFRP1, COL16A1, THY1, PPIB and BGN showed 4%, 3%, 1%, 1.2% and 4% alterations, respectively (Fig. 3C). Similarly, the alteration types in selected genes varied in different OC pathologies (Fig. 3D).

Validation of the Six mRNAs for the Survival Prediction by Kaplan-Meier Curves and the expression profile of the six mRNAs

The Kaplan-Meier curves were used to validate the prognostic prediction of OC by every mRNA. The results showed that besides SFRP1, the risk score was a specific indicator between high and low risk groups ($p < 0.0001$). The patients in high risk group were definitely faced with lower survival rate (Fig. 4A). The original data of expression profile of the six mRNAs without log transformation in high and low risk groups were displayed in Fig. 4B. Furthermore, the expression levels of the six selected genes were tested in normal ovary tissues and OC tissues (Fig. 4C). In addition, the expression level of every mRNA was detected in ovarian cancer tissues and normal tissues. It indicated that risky genes like TGFB1 ($p = 0.0002$), SFRP1 ($p = 0.0344$), COL16A1 ($p = 0.0449$) and THY1 ($p < 0.0001$) were overexpressed in ovarian cancer tissues than normal tissues while protective genes like PPIB ($p = 0.0128$) and BGN ($p = 0.0047$) were low expressed (Fig. 4D).

EMT process was validated to be associated with high risk group of OC via GSEA database

According to the high and low risk score group through GSEA database, we came to a heatmap displayed various biological processes associated with OC, from which we could see that EMT was included (Fig. 5A). The prognostic signaling pathways were evaluated by KEGG and GO pathways, from which we concluded that the prognosis-related genes were enriched in ligand receptor activity, etc. Figure 5B and C displays the GO and KEGG pathway enrichment plots, respectively, for OC.

The Risk Score was Identified as an Independent Prognostic Indicator in OC

Univariate and multivariate analyses were carried out together to compare the risk score to other common clinical pathological parameters (Table 4). The dataset indicated that risk score, stage and cancer status were independent prognostic indicators since their p values were < 0.05 not only in univariate but also in multivariate analyses. Importantly, the risk score was the most obvious clinical parameter related to

mortality for the patients in the high-risk score group, who were 1.408 times more likely to suffer from death than were those in the low-risk score group.

Table 4
Univariable and multivariable analyses for each clinical feature.

Clinical feature	Number	Univariate analysis			Multivariate analysis		
		HR	95%CI of HR	P value	HR	95%CI of HR	P value
Riskscore	367	1.408	1.02–1.944	0.038	1.545	1.098–2.173	0.012
Stage	367	0.449	0.27–0.747	0.002	0.475	0.265–0.849	0.012
Cancer Status	318	0.431	0.309–0.603	0	0.421	0.300-0.594	0
Grade	366	1.227	0.749–2.009	0.417			
age	367	0.871	0.638–1.189	0.385			
venous invasion	102	1.026	0.611–1.722	0.924			
lymphatic invasion	147	1.086	0.686–1.72	0.725			

Validation of the Six-mRNA Signature for the Survival Prediction by Kaplan-Meier Curves

The Kaplan-Meier curves and long rank method were used to validate the prognostic prediction of OC by clinical parameters (risk score, stage, age, cancer status, grade, venous invasion and lymphatic invasion). The results indicated that patients with high risk scores had poor prognoses. The patients with tumors and in cancer stage III-IV were definitely faced with a high risk of poor prognosis, and the accuracy of our analysis was further verified (Fig. 6A). The time-dependent ROC curve of each parameter demonstrated the sensitivity and specificity of the 5-year OS prediction (Fig. 6B). The AUC of the ROC curve verified the accuracy of our prognostic model. Furthermore, we performed a data stratification analysis on the entire cohort, and 367 patients were stratified based on their clinical parameters. According to the significant results above, patients with tumors and stage III cancers in grade 3 were more likely associated with a shorter survival time (Fig. 6C).

The Kaplan-Meier curves was used to validate the prognostic prediction of OC by the risk score in an independent GEO cohort (GSE9891). Although there are more than a dozen GEO databases for ovarian cancer, we chose GSE9891 database that covered the six genes matched with survival data in our study.

It was apparent that patients with high risk scores had poor prognoses ($p < 0.0001$), which was consistent with the result we obtained above in Fig. 6D. As we elucidated above, every patient was endowed with a risk score. The median risk score value was regarded as the cut-off criterion, and the patients were divided into low-risk and high-risk groups similarly through GEO database. The distribution of the patient relapse status is shown in Fig. 6E. Meanwhile, The Kaplan-Meier curves were also used to validate the prognostic prediction of OC in colon cancer (Fig. 6F) and hepatocellular cancer (Fig. 6G). The results showed that the risk score was not a significantly indicator between high and low risk groups ($p = 0.1158$ and $p = 0.3675$), which indicated the specificity of the risk score model in ovarian cancer. PPI networks obtained from the STRING database and visualized by Cytoscape software helped us to identify the hub genes among the prognosis-related genes. The core genes (degrees ≥ 15) (Additional file 2: Fig. S2B) were further submitted to PPI network analysis, which indicated that TGFBI, SFRP1, and COL16A1 were clearly at the center of the network (Additional file 2: Fig. S2A).

4. Discussion

In clinical practice, although two patients with the same clinical and pathological parameters received the same treatment, the prognosis was sometimes discriminating. Robust biomarkers are emerging to stratify high-risk groups among these patients. For example, the expression level of miR-205 was an independent variation selected to predict the events of poor clinical evolution[16]. The soluble VEGFR-2 level could predict the malignancy of ovarian neoplasms and poor prognosis in epithelial OC[17]. The InHIFCAR-mediated mechanism for HIF-1 activation was of great value in prognosis and potential therapeutic strategy for oral carcinoma[18, 19]. Nevertheless, increasing studies support the fact that the regulation of genes in biological processes is complex, with multiple genes interacting with each other to form a network. Therefore, the expression level of a single gene is not enough to predict the prognosis of these patients[20–22].

EMT together with the reverse process mesenchymal-epithelial transition (MET) were initially researched in embryonic morphogenesis[23, 24]. In EMT, cancer cells lose their epithelial features and acquire invasive properties. Activation of the EMT process elicits changes in multiple fundamental aspects of cellular physiology, for example, wound healing, tissue fibrosis, drug resistance and the hotspot aspect - carcinoma progression[24, 25]. Over the last decade, the EMT process has been well studied, and there have also been arguments about the mechanism underlying cancer[26]. As previously described, EMT might control the expression of immune checkpoint inhibitors and promote immune evasion in nonsmall cell lung carcinoma[27]. miR-34a acts on SNAIL to regulate EMT in breast cancer and lung cancer cells[28]. The current EMT markers are not consistently related to poor prognosis because the cellular context is complex.

Currently, the wide application of genomic profiling is gradually becoming a standard tendency in clinical oncology and will undoubtedly accelerate the progress in diagnosis and therapy[29–31]. The age of big data has arrived, and the endless screening of the hub molecules has emerged. For example, researchers presented a data-interlinked platform called BIOOPENER, which enabled the query of different types of

mutations and genomic alterations to contribute to molecular and clinical insights on cancer, resulting in the discovery of three potential pathways that cause promoter changes in gynecological cancers[32]. Gene signatures containing several genes are superior to individual biomarkers that contribute to OC prognostic prediction and can be applied widely[33–35].

In the present study, vital bioinformatics analyses, such as GSEA, were used to revealed that EMT was potentially the most affected pathway in OC with $p < 0.05$. KEGG, GO and PPI networks were carried out to display the prognosis-related biological processes and hub genes. To narrow the area and improve the predictive efficiency, univariate Cox regression and multivariate Cox regression analyses were performed, and out of 197 mRNAs, six mRNAs, namely, TGFBI, SFRP1, COL16A1, THY1, PPIB and BGN, were associated with the prognosis of OC patients. We were able to distinguish low-risk and high-risk samples with a relatively high prognosis accuracy consisting of these six mRNAs. Among these mRNAs, TGFBI was investigated as a highly induced transcript during EMT in A549 cells, acting as the ceRNA for miR-21 to modulate EMT[36]. SFRP1 exhibited a tumor-promoting function by selectively activating TGF β signaling in gastric cancer cells and thus activating EMT progress[37]. Furthermore, we investigated the prognostic value of the signature with other clinical pathologic parameters and in different subgroups. Based on the Cox analysis, the signature exhibited forceful prediction for risk score and stage in OC patients, revealing that the signature was a prospective independent marker. Actually, under the confirmation of the ROC curve, this six-gene signature exhibited a more powerful prediction for OS than did individual genes, declaring that this signature is a promising independent marker for OC. The exprelssion level of these genes detected by qRT-PCR validated that the risky genes were overexpressed in ovarian cancer tissues and the protective genes were low expressed. The Kaplan-Meier curves validated that the six-gene model was developed as a nomogram to predict the survival probability of OC patients. Finally, GEO cohort was exhibited to confirm the prognosis predictive capacity of this signature additionally, and the survival result was consistent with the tendency above. In fact, it is common that OC metastasis to the liver, colon and other important viscera. But the risk score of the six-gene signature was not significant in these two cancers, which indicated that risk score was specific to OC prognosis. To our knowledge, this study is the first to build a prognostic gene signature in OC.

Despite the novel findings proposed by our study of candidates for OC prognosis, there are still limitations that require further investigation. This study is a retrospectively designed study, and the results would be more convincing if the sample size could be enlarged.

Conclusion

The six signature genes TGFBI, SFRP1, COL16A1, THY1, PPIB and BGN might be potential biomarkers for predicting the prognosis of OC patients. To some degree, our study might provide some clues for further investigation into the biological processes, clinical diagnosis and therapy of these genes.

Abbreviations

OC: Ovarian cancer; TCGA: The Cancer Genome Atlas; GSEA: gene set enrichment analysis; EMT: epithelial-to-mesenchymal transition; OS: overall survival; GEO: Gene Expression Omnibus; DLL3: delta-like protein 3; VEGF: vascular endothelial growth factor; AJCC: American Joint Committee on Cancer; AUC: area under curve; ROC; receiver operating characteristic; PDAC: pancreatic ductal adenocarcinoma; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; ChIP: Chromatin Immunoprecipitation; PPI: protein-protein interaction; MET: mesenchymal-epithelial transition.

Declarations

Acknowledgments

This study was funded by the National Natural Science Foundation of China (No. 81872123); University innovation team of Liaoning Province; Special Professor of Liaoning Province; "Major Special Construction Plan" for Discipline Construction of China Medical University in 2018 (No. 3110118029); Outstanding Scientific Fund of Shengjing Hospital (No. 201601).

Authors' contributions

X.P. analyzed the data, drafted the manuscript, prepared all Fig.s and tables. X.X.M. designed the study. All authors read and approved the final manuscript.

Funding

This study was funded by the National Natural Science Foundation of China (No. 81872123); University innovation team of Liaoning Province; Special Professor of Liaoning Province; "Major Special Construction Plan" for Discipline Construction of China Medical University in 2018 (No. 3110118029); Outstanding Scientific Fund of Shengjing Hospital (No. 201601).

Availability of data and materials

All data or related information supporting the conclusions of the review is included in the article.

Ethics approval and consent to participate

This study was approved by the committee of Shengjing Hospital (2018PS251K).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang, 110021, PR China.

Supplemental Figures

Additional file 1: Fig. S1. Heatmap of 83 common differentially expressed genes in the good and poor prognosis groups.

Additional file 2: Fig. S2. (A) PPI network analysis of prognosis-related genes. (B) Quantity of the genes correlated with the hub genes.

References

1. Baldwin LA, Chen Q, Tucker TC, White CG, Ore RN, Huang B. Ovarian Cancer Incidence Corrected for Oophorectomy. *Diagnostics (Basel)*. 2015;7.
2. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136,E359-86.
3. Torre LA, Trabert B, Desantis CE, Miller KD, Samimi G, Runowicz CD, et al. Ovarian cancer statistics. *Histopathology*. 2018;68,284-96.
4. Tarver T. Cancer Facts & Fig.s 2012. American Cancer Society (ACS). *J Consum Health Internet*. 2012;16,366-7.
5. Minlikeeva AN, Freudenheim JL, Cannioto RA, Szender JB, Eng KH, Modugno F, et al. History of hypertension, heart disease, and diabetes and ovarian cancer patient survival: evidence from the ovarian cancer association consortium. *Cancer causes control*. 2017;28,469-86.

6. Testa U, Petrucci E, Pasquini L, Castelli G, Pelosi E. Ovarian Cancers: Genetic Abnormalities, Tumor Heterogeneity and Progression, Clonal Evolution and Cancer Stem Cells. *Medicines (Basel)*. 2018;5.
7. Nakagawa H, Fujita M. Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Sci*. 2018;109,513-22.
8. Schuh A, Dreau H, Knight SJL, Ridout K, Mizani T, Vavoulis D, et al. Clinically actionable mutation profiles in patients with cancer identified by whole-genome sequencing. *Cold Spring Harb Mol Case Stud*. 2018;4.
9. Song Z, Huang Y, Zhao Y, Ruan H, Yang H, Cao Q, et al. The Identification of Potential Biomarkers and Biological Pathways in Prostate Cancer. *J Cancer*. 2019;10, 1398-408.
10. Wang J, Zhang K, Liu Z, Wang T, Shi F, Zhang Y, et al. Upregulated delta-like protein 3 expression is a diagnostic and prognostic marker in endometrial cancer: A retrospective study. *Medicine*. 2018;97,e13442.
11. Jayson GC, Zhou C, Backen A, Horsley L, Marti-Marti K, Shaw D, et al. Plasma Tie2 is a tumor vascular response biomarker for VEGF inhibitors in metastatic colorectal cancer. *Nat Commun*. 2018;9,4672.
12. Raman P, Maddipati R, Lim KH, Tozeren A. Pancreatic cancer survival analysis defines a signature that predicts outcome. *PloS one*. 2018;13,e0201751.
13. Mathur R, Rotroff D, Ma J, Shojaie A, Motsinger-Reif A. Gene set analysis methods: a systematic comparison. *BioData Mining*. 2018;11,8.
14. Zhang Y, Topham DJ, Thakar J, Qiu X. FUNNEL-GSEA: FUNctioNal ELastic-net regression in time-course gene set enrichment analysis. *Bioinformatics*. 2017;33,1944-52.
15. Dean JL, Zhao QJ, Lambert JC, Hawkins BS, Thomas RS, Wesselkamper SC. Editor's Highlight: Application of Gene Set Enrichment Analysis for Identification of Chemically Induced, Biologically Relevant Transcriptomic Networks and Potential Utilization in Human Health Risk Assessment. *Toxicol Sci*. 2017;157,85-99.
16. Cañueto J, Cardeñoso-Álvarez E, García-Hernández JL, Galindo-Villardón P, Vicente-Galindo P, Vicente-Villardón JL, et al. MicroRNA (miR)-203 and miR-205 expression patterns identify subgroups of prognosis in cutaneous squamous cell carcinoma. *Br J Dermatol*. 2017;177,168-78.
17. Sallinen H, Heikura T, Koponen J, Kosma VM, Heinonen S, Ylä-Herttua S, et al. Serum angiopoietin-2 and soluble VEGFR-2 levels predict malignancy of ovarian neoplasm and poor prognosis in epithelial ovarian cancer. *BMC cancer*. 2014;14,696.
18. Shih JW, Chiang WF, Wu ATH, Wu MH, Wang LY, Yu YL, et al. Long noncoding RNA LncHIFCAR/MIR31HG is a HIF-1 α co-activator driving oral cancer progression. *Nat Commun*. 2017;8,15874.
19. Bolha L, Ravnik-Glavač M, Glavač D. Long Noncoding RNAs as Biomarkers in Cancer. *Disease markers*. 2017;2017,7243968.
20. Chen YL, Ge GJ, Qi C, Wang H, Wang HL, Li LY, et al. A five-gene signature may predict sunitinib sensitivity and serve as prognostic biomarkers for renal cell carcinoma. *J Cell Physiol*. 2018;233.

21. Qiu Z, Sun W, Gao S, Zhou H, Tan W, Cao M, et al. A 16-gene signature predicting prognosis of patients with oral tongue squamous cell carcinoma. *Peer J*. 2017;5,e4062.
22. Zuo S, Dai G, Ren X. Identification of a 6-gene signature predicting prognosis for colorectal cancer. *Cancer Cell Int*. 2019;19,6.
23. Hay ED. The mesenchymal cell, its role in the embryo, and the remarkable signaling mechanisms that create it. *Dev Dyn*. 2005;233,706-20.
24. Shibue T, Weinberg RA. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat Rev Clin Oncol*. 2017;14,611-29.
25. Nieto MA, Huang RY, Jackson RA, Thiery JP. EMT: 2016. *Cell*. 2016;166,21-45.
26. Jolly MbK, Ware KE, Gilja S, Somarelli JA, Levine H. EMT and MET: necessary or permissive for metastasis? *Mol Oncol*. 2017;11,755-69.
27. Asgarova A, Asgarov K, Godet Y, Peixoto P, Nadaradjane A, Boyer-Guittaut M, et al. PD-L1 expression is regulated by both DNA methylation and NF- κ B during EMT signaling in non-small cell lung carcinoma. *Oncoimmunology*. 2018;7,e1423170.
28. He SJ, Xiang CQ, Zhang Y, Lu XT, Chen HW, Xiong LX. Recent progress on the effects of microRNAs and natural products on tumor epithelial-mesenchymal transition. *Onco Targets Ther*. 2017;10,3435-51.
29. Wheeler DA, Wang L. From human genome to cancer genome: the first decade. *Genome Res*. 2013;23,1054-62.
30. Ocak S, Sos ML, Thomas RK, Massion PP. High-throughput molecular analysis in lung cancer: insights into biology and potential clinical applications. *Eur Respir J*. 2009;34,489-506.
31. Ouellet V, Aprikian A, Bergeron A, Brimo F, Bristow RG, Chevalier S, et al. The Terry Fox Research Institute Canadian Prostate Cancer Biomarker Network: an analysis of a pan-Canadian multi-center cohort for biomarker validation. *BMC Urol*. 2018;18,78.
32. Jha A, Khan Y, Mehdi M, Karim MR, Mehmood Q, Zappa A, et al. Towards precision medicine: discovering novel gynecological cancer biomarkers and pathways using linked data. *J Biomed Semantics*. 2017;8,40.
33. Liu XP, Yin XH, Meng XY, Yan XH, Wang F, He L. Development and Validation of a 9-Gene Prognostic Signature in Patients With Multiple Myeloma. *Front Oncol*. 2018;8,615.
34. Qiao GJ, Chen L, Wu JC, Li ZR. Identification of an eight-gene signature for survival prediction for patients with hepatocellular carcinoma based on integrated bioinformatics analysis. *Peer J*. 2019;7,e6548.
35. Shahid M, Choi TG, Nguyen MN, Matondo A, Jo YH, Yoo JY, et al. An 8-gene signature for prediction of prognosis and chemoresponse in non-small cell lung cancer. *Oncotarget*. 2016;7,86561-72.
36. Liu Y, Xue M, Du S, Feng W, Zhang K, Zhang L, et al. Competitive endogenous RNA is an intrinsic component of EMT regulatory circuits and modulates EMT. *Nat Commun*. 2019;10,1637.

Figures

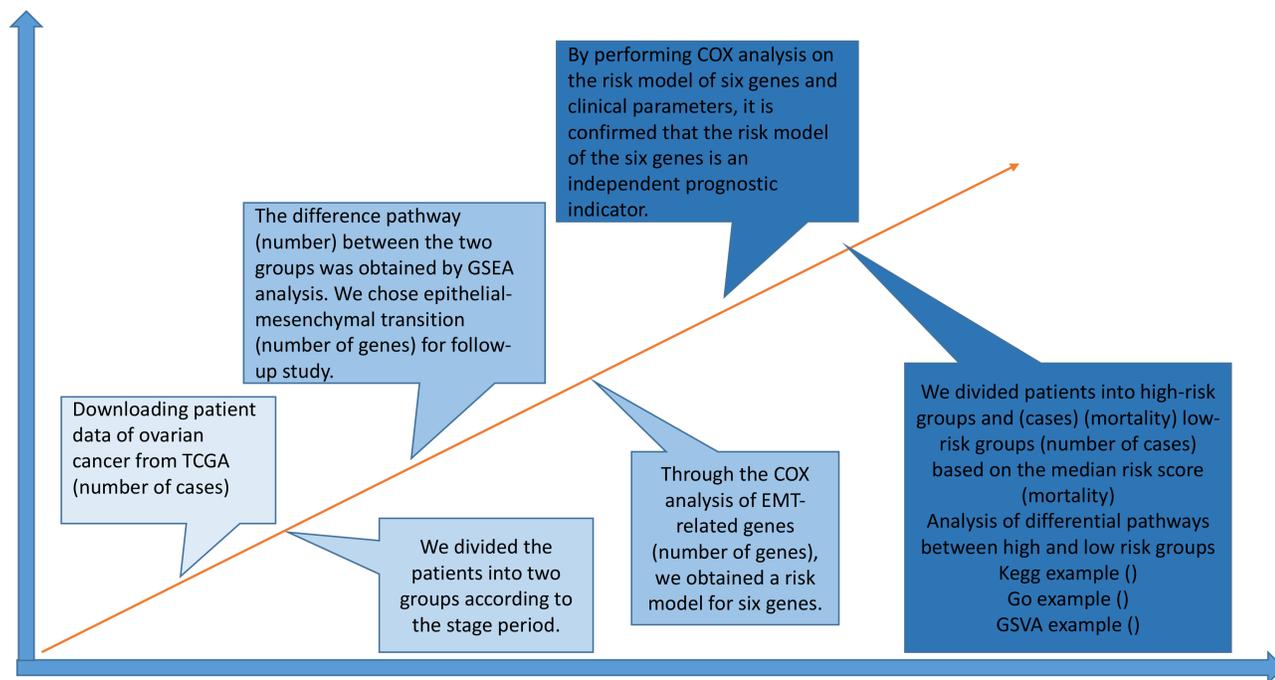


Figure 1

Flowchart of this study.

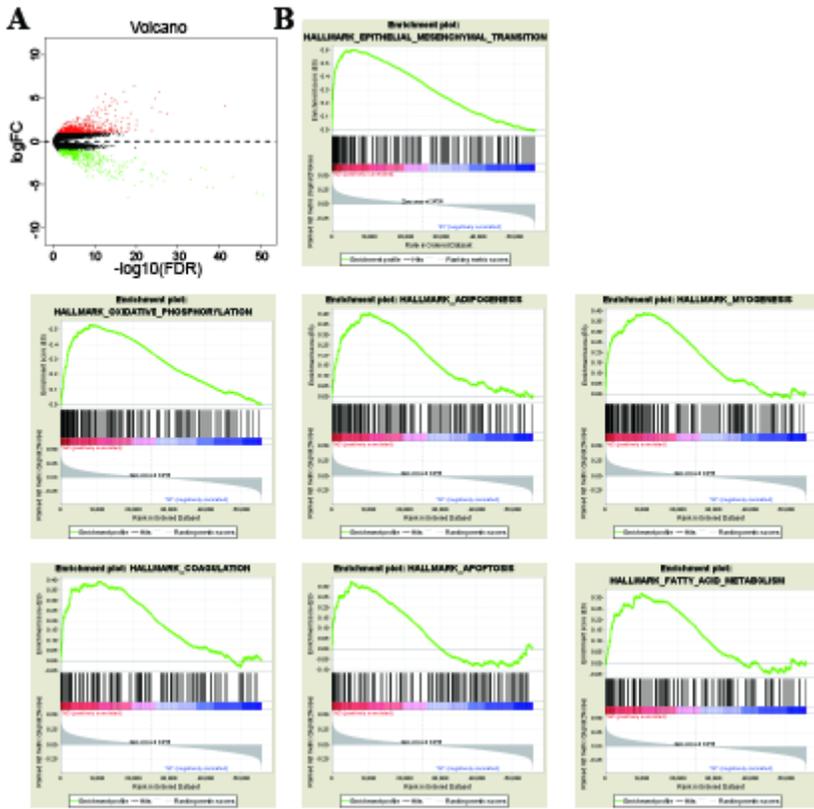


Figure 2

(A) Differentially expressed mRNAs identified using the edgeR package. (B) Enrichment plots of 7 gene sets that were functionally enriched in OC.

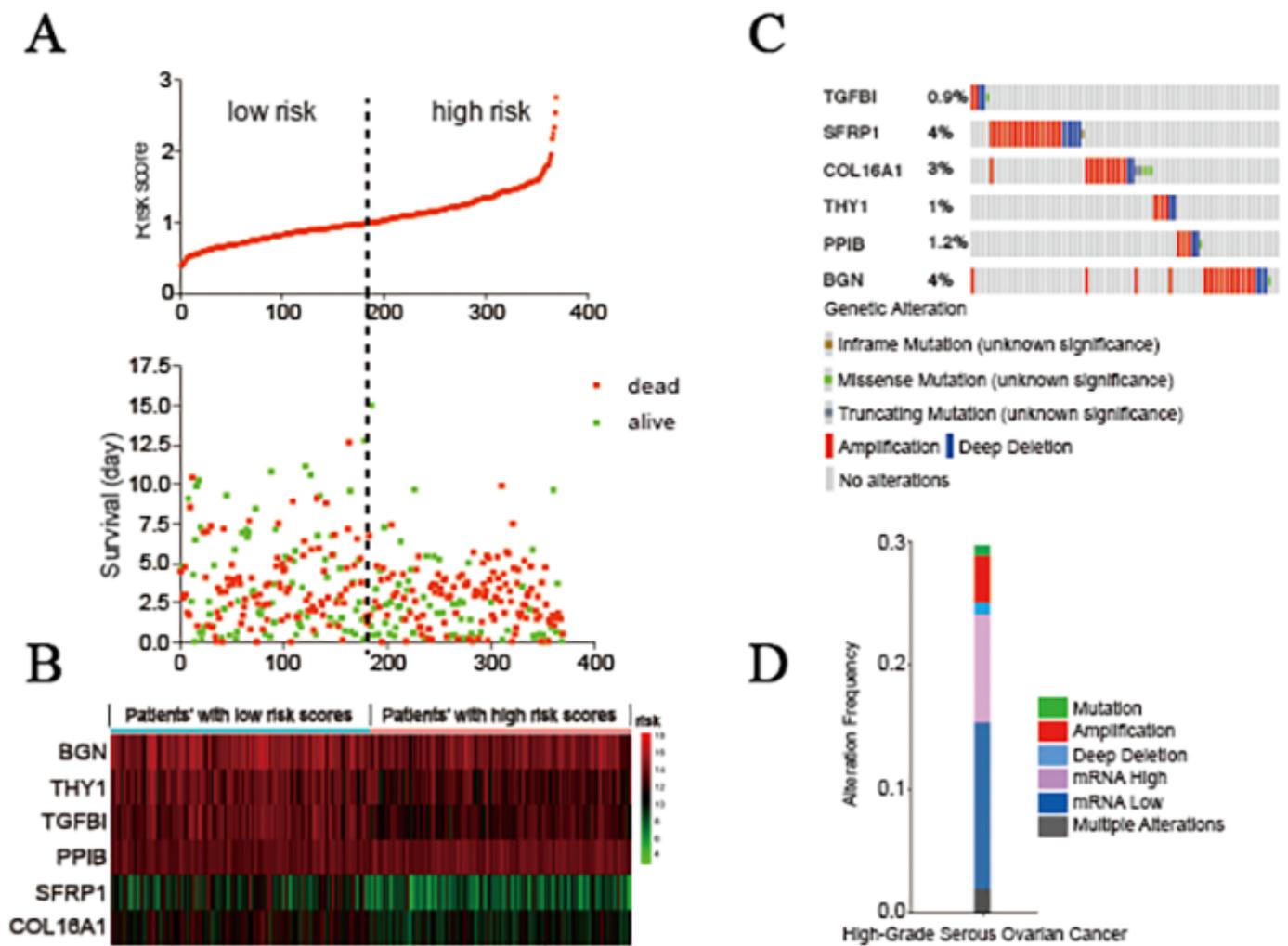


Figure 3

The six-mRNA signature related to risk score predicts the OS and was associated with survival in OC patients. (A) mRNA risk score distribution and survival days of patients. (B) A heatmap of five gene expression profiles. (C) Selected gene alterations in the clinical samples. (D) Selected gene-specific alterations in the detailed cancer type.

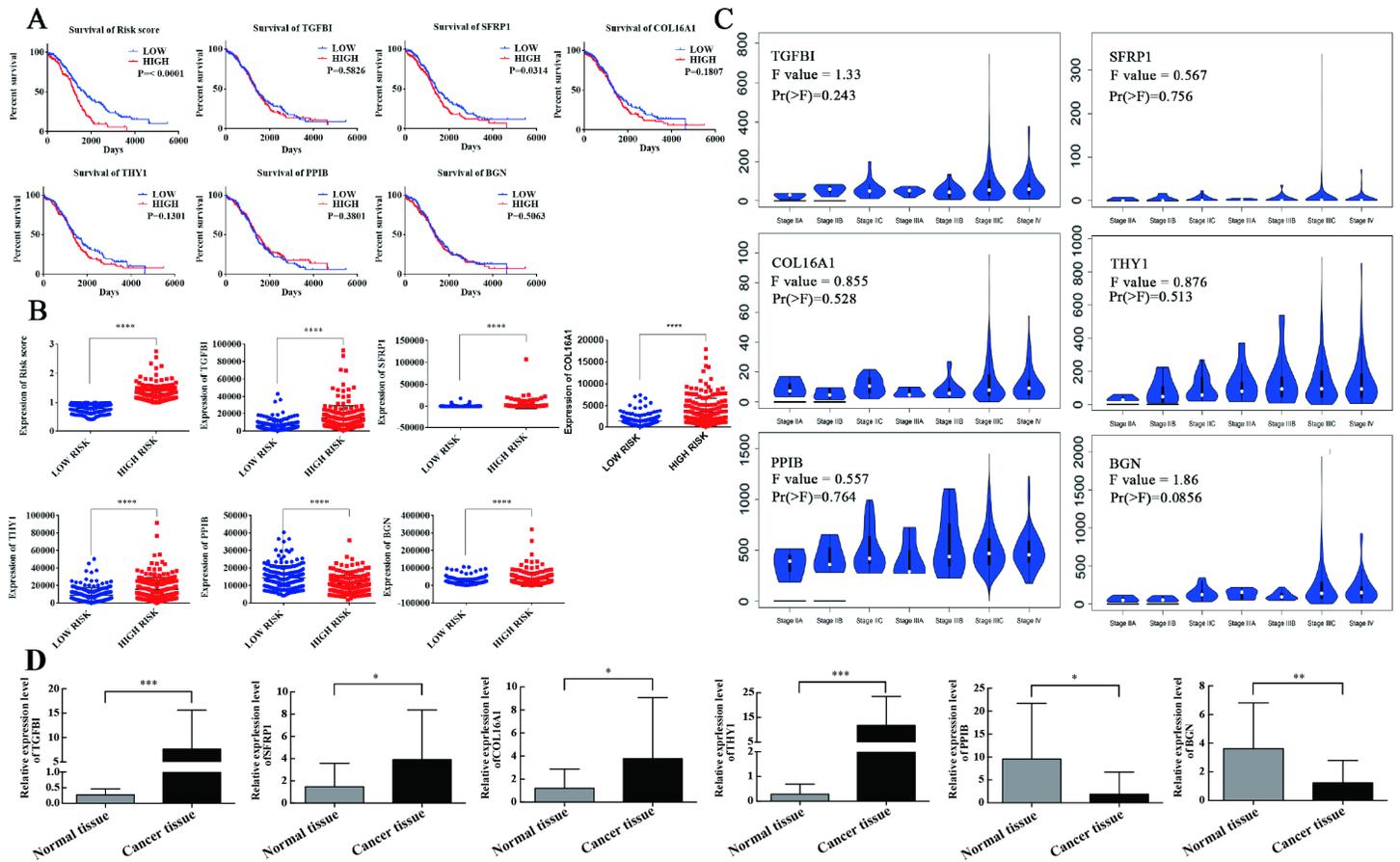


Figure 4

Survival validation and expression profile of the 6 genes grouped by high and low risk. (A) Prognostic prediction of OC by every mRNA in high and low risk groups. (B) Gene expression of the 6 genes in high and low risk groups. (C) Differential expression of 6 genes. (D) The expression level of six mRNAs detected by qRT-PCR.

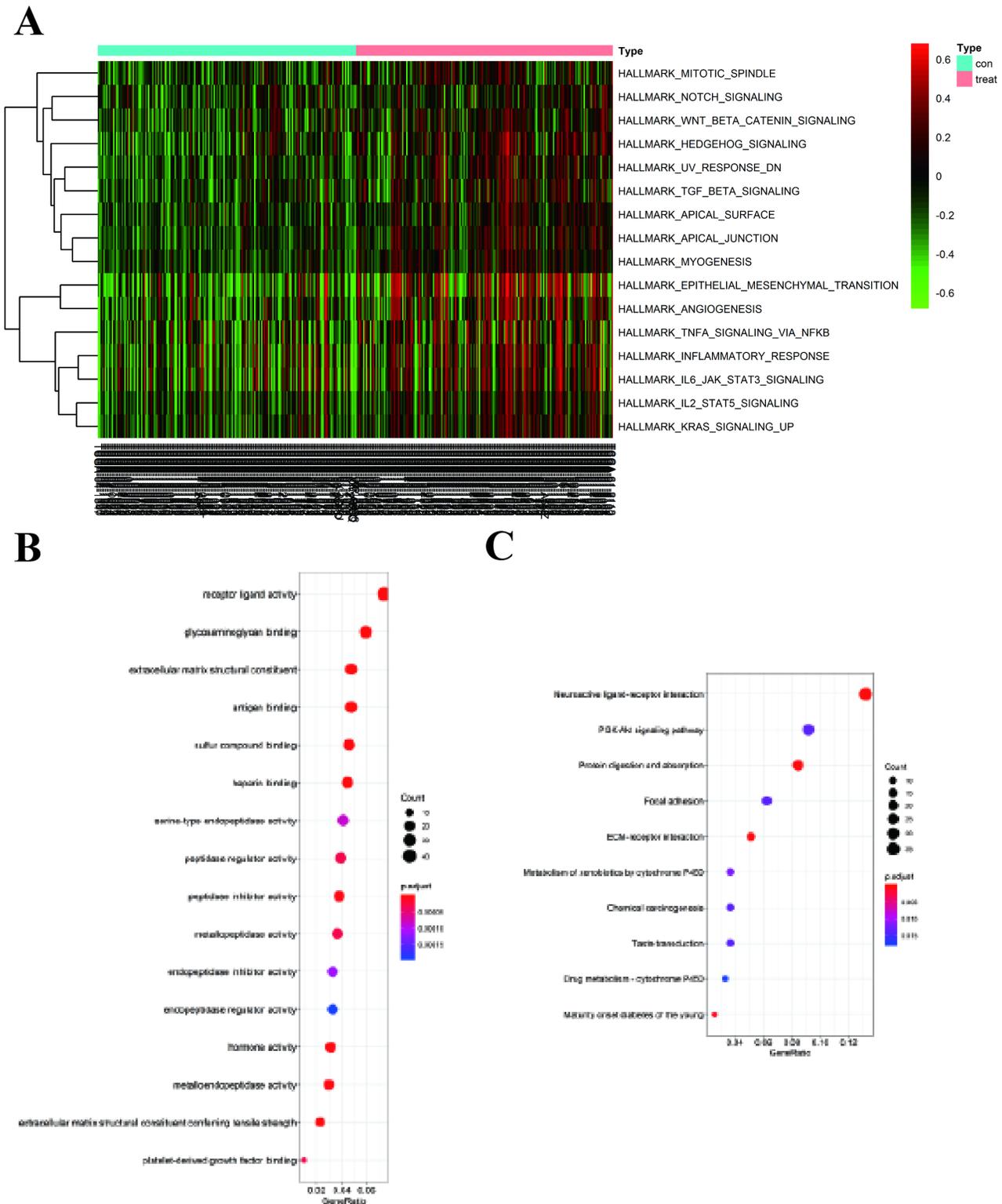


Figure 5

EMT process was associated with high risk group of OC. (A) Heatmap displayed various biological processes associated with OC. (B) GO and (C) KEGG pathways for the functional enrichment analysis.

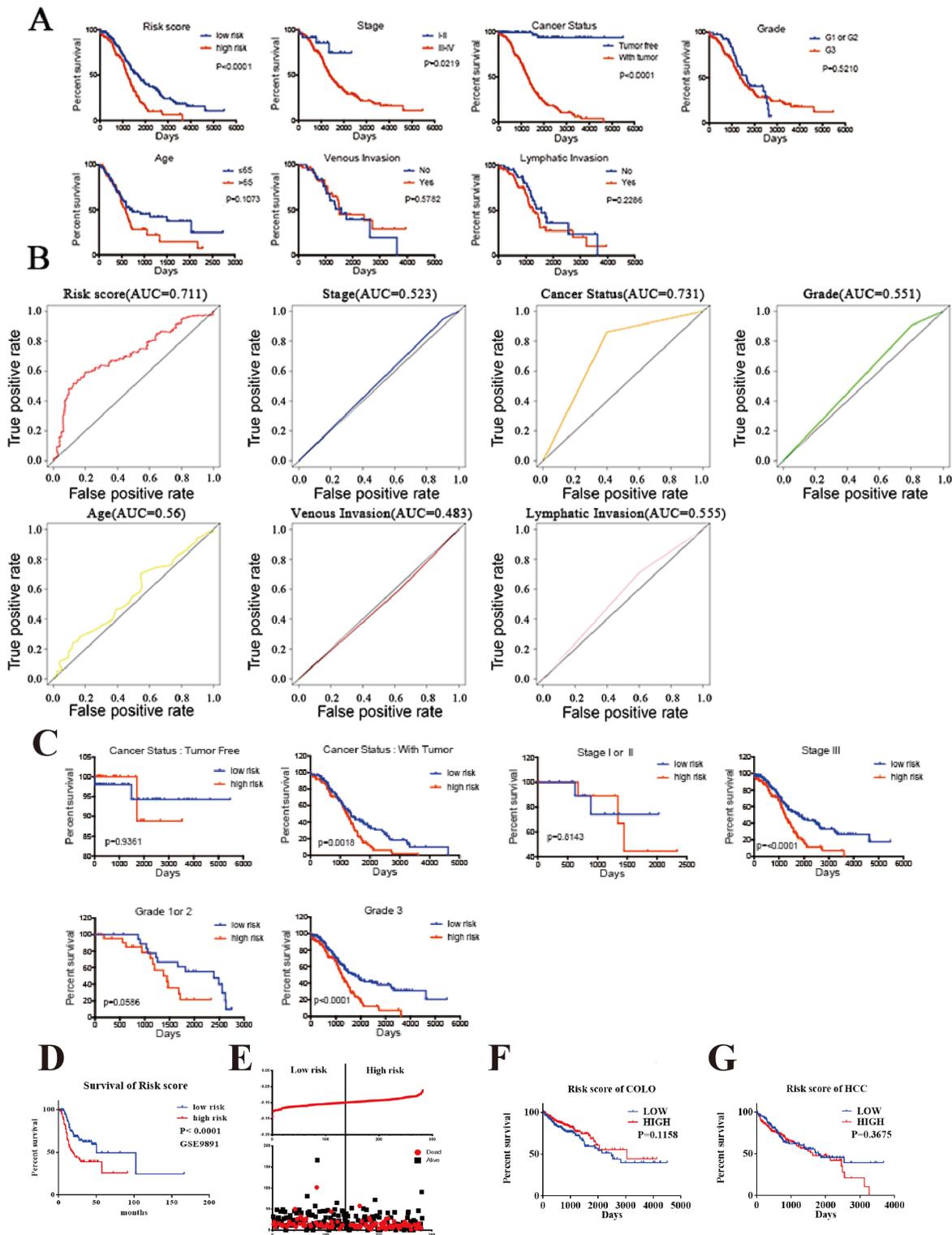


Figure 6

Kaplan-Meier survival analysis for the patients with OC in the TCGA dataset. (A) The Kaplan-Meier curve for patients divided into high-risk and low-risk groups and different clinical features that predict patient survival. (B) ROC curves of each parameter of the 5-year OS prediction. (C) Kaplan-Meier curves for prognostic value of risk-score signature for the patients divided by each clinical feature. (D) GEO

validation of the signature. (E) Distribution of OC patients with risk scores. (F) and (G) Prognostic prediction of OC in colon cancer and hepatocellular cancer.

Supplementary Files

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

- [SupplementaryFigure2.tif](#)
- [SupplementaryFigure1.tif](#)