Prognostic signature of ovarian cancer based on 14 tumor microenvironment genes

Xiazi Nie
Gansu Provincial Hospital

Lina Song
Gannan Tibetan Autonomous Prefecture State People’s Hospital

Xiaohua Li
People’s Hospital of Gangu County, Gansu Province

Yirong Wang
Gansu Maternal and Child Health Care Hospital

Bo Qu (✉ qubo291@outlook.com)
Gansu Provincial Hospital  https://orcid.org/0000-0002-3411-5290

Research article

Keywords: Ovarian cancer, tumor microenvironment, prognosis, nomogram

DOI: https://doi.org/10.21203/rs.3.rs-56329/v1

License: ☑️  This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background

Ovarian cancer is one of the lethal gynecological in women. Tumor microenvironment (TME) is emerging as a pivotal biomarker for patients' therapeutic sensitivity and prognosis. In this study, we proposed to explore the prognostic role of TME-related genes in ovarian cancer.

Methods

The data of whole genome expression profiles and detailed clinicopathological information of three cohorts of ovarian cancer patients from the Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). Univariate Cox regression analysis was used to screen TME-related genes with significantly prognostic value based on TCGA cohort. LASSO Cox regression analysis was adapted to the construction of prognostic model. Ovarian cancer cohorts from GEO were used as validation set for verifying the reliability of the prognostic model. Relative infiltrating proportion of 22 immune cells were estimated through CIBERSORT software.

Results

This study identified a total of 14 TME-related genes that finally incorporated into the prognostic model. The risk score that calculated through the prognostic model was proved as an independent prognostic signature in ovarian cancer. Nomogram that contains TNM stage and risk score could reliably predict the long-term overall survival probability. Additionally, risk score was significantly associated with the relative infiltrating proportion of several immune cells in ovarian cancer and mRNA levels of some immune checkpoint genes.

Conclusions

This study constructed a prognostic model for ovarian cancer, which was closely associated with the prognosis and immune status. This should provide novel clue for prognosis study in ovarian cancer.

Background

Ovarian cancer is the most lethal gynecological malignancy[1], following cervical cancer and uterine body cancer, the mortality of the disease accounts for up to 5% in female cancer-related deaths in the worldwide[2]. Malignant ovarian lesions include primary lesions and secondary lesions. In the primary lesions, epithelial ovarian carcinoma is the most common among malignant ovarian tumors (70% of all ovarian malignancies); also including stromal tumors of the ovary, germ-cell tumors, sex-cord stromal tumors, and other more rare types[3]. Unfortunately, early ovarian cancer causes minimal, nonspecific, or
no symptoms due to its anatomy features[4]. Therefore, the most of cases are diagnosed in an advanced stage.[5, 6]. Epithelial ovarian carcinoma has been shown to have considerable complexity and heterogeneity in biology, drug response, and survival time, representing a major obstacle for its precision medicine practice.

Recent studies have shown that the tumor microenvironment (TME), composed of a variety of immune cells and stromal cells, is crucial in the occurrence and development of cancer and the response of cells to chemotherapy[7]. At the same time, the rise of immunotherapy, including immune checkpoint inhibitors, shows that the assessment of TME heterogeneity and the remodeling of the immune microenvironment will have broad application in and profound impact on the future cancer treatment[8]. Recently, some studies have shown prognostic markers in tumor microenvironment for tumor prognosis. The study describes the comprehensive features of gastric cancer TME-related genes characteristics and provide new strategies for cancer treatment[9]. Another present study found that many elements of TME other than tumor epithelial cells influence the progression of non-small cell lung cancer[10]. Consequently, new treatment strategies and paradigms are of great need for these patients. Unfortunately, there are currently no approved immune therapies for ovarian cancer. Therefore, mining TME-related genes in the immune microenvironment significantly changed the treatment prospects of ovarian cancer. TME-related genes may denote a solution to overwhelmed drug resistance and improve the results of the ovarian cancer patients[11].

In this study, we analyzed the expression of 760 tumor microenvironment-related genes in ovarian cancer patients. Meanwhile, we constructed a risk score model by hub TME-related genes to predict the prognosis of ovarian cancer and verified evaluation efficiency of the model in the multiple level, which provides new ideas for improving the prognosis of ovarian cancer.

Methods

Study population and TME-related genes

In the present study, 386 ovarian cancer patients' mRNA expression profile data and corresponding clinical information were retrieved from The Cancer Genome Atlas (TCGA, https://tcga-data.nci.nih.gov/tcga/). Data set with incomplete survival information were excluded. 375 of 386 patients have complete survival information and were used in the following analysis. The detailed clinical information of these 375 patients is shown in Table 1. In addition, the data sets GSE26193 and GSE63885 were retrieved from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/). GSE26193 contains information of 107 ovaries cancer patients with complete survival information, and GSE63885 contains information of 101 ovarian cancer patients, 78 of which have complete survival information. MRNA levels of ovarian cancer samples in the two datasets of the GEO database were quantified by the Affymetrix Human Genome U133 Plus 2.0 Array platform. The 760 TME-related genes were obtained based on xCell (https://xcell.ucsf.edu/) and listed in Table S1.
Table 1
Clinicopathological characteristics of OV patients from TCGA database.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>OV Patients (N = 375)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO.</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>≤ 59 (Median)</td>
<td>198</td>
</tr>
<tr>
<td>&gt; 59 (Median)</td>
<td>177</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>white</td>
<td>326</td>
</tr>
<tr>
<td>Black or african american</td>
<td>24</td>
</tr>
<tr>
<td>asian</td>
<td>11</td>
</tr>
<tr>
<td>American indian or alaska</td>
<td>2</td>
</tr>
<tr>
<td>native hawaiian or other pacific islander</td>
<td>1</td>
</tr>
<tr>
<td>Unknown</td>
<td>11</td>
</tr>
<tr>
<td>Pathologic stage</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>1</td>
</tr>
<tr>
<td>ii</td>
<td>24</td>
</tr>
<tr>
<td>iii</td>
<td>293</td>
</tr>
<tr>
<td>iv</td>
<td>57</td>
</tr>
<tr>
<td>Survival Time</td>
<td></td>
</tr>
<tr>
<td>Long (&gt; 5 years)</td>
<td>75</td>
</tr>
<tr>
<td>Short (&lt; 5 years)</td>
<td>300</td>
</tr>
<tr>
<td>OS status</td>
<td></td>
</tr>
<tr>
<td>Dead</td>
<td>230</td>
</tr>
<tr>
<td>Alive</td>
<td>145</td>
</tr>
</tbody>
</table>

Construction Of Prognostic Signature

Based on the expression levels of 760 TME-related genes, univariate Cox regression analysis was performed, and genes with significant correlation with the prognosis of ovarian cancer were identified with P < 0.05 considered as statistically significant. Then the LASSO Cox regression analysis was performed using R language glmnet package to further optimize the TME-related genes that associated with the prognosis of ovarian cancer[12]. Risk score for each sample was calculated with the optimized genes base the following formula:

Risk score = \( \sum_{i=1}^{n} Coef_i * X_i \)
In this formula, $\text{Coef}_i$ is the risk coefficient of each factor calculated by the LASSO-Cox model, $X_i$ is the expression value of each factor, and in this study refers to the expression level of mRNA of TME-related genes. Then, R package survival, survminer and two-sided log-rank test were applied to determine the optimal cutoff value of risk score for stratifying patients. According to the cutoff value, patients were attributed into Low-Risk group and High-Risk group.

**Survival Analysis**

Kaplan-Meier analysis was performed to estimate the overall survival rate of different groups using the R language survival package and survminer package. Log-rank test was performed to determine the significance of the difference in survival rate between the different groups. Multivariate Cox regression model was used to analyze whether Risk Score can predict the survival of ovarian cancer patients independently of other factors.

**Calculation Of Immune Cell Infiltration**

We used the software CIBERSORT\cite{13}to calculate the relative infiltration of 22 different kinds of immune cells in each cancer sample. CIBERSORT software will characterize the composition of immune infiltrating cells base on a deconvolution algorithm and 547 pre-set barcode genes based on the gene expression matrix. The sum of the proportion of all estimated immune cell types in each sample is equal to 1.

**Establishment Of Nomogram Prognosis Predictive Model**

Nomogram is widely used to predict the prognosis of cancer. In order to predict the 1-year, 3-year and 5-year survival probabilities of patients, the nomogram was established based on all the independent prognostic factors determined by multivariate Cox regression by using the R language rms package. A calibration curve of the nomogram was drawing to compare the predicted overall survival probability of nomogram and the actual overall survival probability.

**Statistical Analysis**

Kaplan-Meier method was used to estimate the overall survival rate of different groups. Log-rank was used to test the significance of the difference in survival rate between different groups. Wilcoxon signed-rank test method was used to compare the differences in immune cell infiltration in different groups, with $p < 0.05$ as the threshold of statistical significance. R software (v3.5.2) was used in performing statistical analysis in the current study.

**Results**
Construction and verification of prognostic models

Univariate Cox regression analysis was performed using the TCGA data set samples with 760 TME-related gene expression values as continuous variables, and the Hazard ratio (HR) of each gene was calculated. 48 genes were identified as prognostically significant genes with P value < 0.05 as the threshold. The protective genes with HR value less than 1 were conducive to prognosis, and those with HR value greater than 1 were hazardous genes with unfavorable prognosis (Fig. 1A). LASSO Cox regression analysis was performed on the 48 identified genes. According to the lambda value corresponding to the number of different genes in the LASSO Cox analysis, we determined to use 14 genes in the following analysis to achieve the smallest lambda value (Fig. 1B). The 14 genes are ELN, FBLN1, ANGPT2, FBL, COMP, PPL, PLD2, PDLIM4, EMP1, MYC, ITGAM, FRAT2, WDR45 and ADSL.

In order to obtain a unified cut-off value for patients’ stratification across different patients’ datasets, expression values of these genes were normalized to average of 0 and a standard deviation (SD) value of 1. Then, the normalized expression of each nuclear receptor gene is weighted with the regression coefficient of LASSO Cox regression analysis to establish a risk scoring model for predicting patients’ death risk. Risk Score = (0.032648133 * ELN) + (0.050148701 * FBLN1) + (-0.182538141 * ANGPT2) + (0.022643331 * FBL) + (0.008810390 * COMP) + (0.037391649 * PPL) + (0.029721887 * PLD2) + (0.055455835 * PDLIM4) + (0.018301503 * EMP1) + (0.008366183 * MYC) + (0.039140799 * ITGAM) + (0.018056371 * FRAT2) + (0.032383440 * WDR45) + (0.046671263 * ADSL). Base on this formula, the risk score of each patient in the TCGA data set and meta-GEO verification set (the combination of two sets of GEO data) was calculated. And according to the calculated optimal cut-off point (0.00515), these patients are divided into a high-risk group and a low-risk group. Survival analysis found that in the TCGA and meta-GEO datasets, high-risk ovarian cancer samples have poorer survival comparing to the low-risk samples (Fig. 1C and Fig. 1D). In addition, the expression levels of the 14 genes in the model between the high and low risk groups of the TCGA data set were significantly different (Fig. 1E). In conclusion, the risk score calculated by the 14 genes, ELN, FBLN1, ANGPT2, FBL, COMP, PPL, PLD2, PDLIM4, EMP1, MYC, ITGAM, FRAT2, WDR45 and ADSL can serve as a model to predict the death risk of patients with ovarian cancer.

Risk Score is an independent prognostic marker for ovarian cancer

Age, TNM Stage and Risk Score were adopted for multivariate Cox regression analysis to determine whether the risk score is an independent prognostic indicator (Fig. 2A). It was found that Risk Score and Stage were still significantly correlated with overall survival. Samples with a high-risk score were at greater risk of death (HR = 6.33, 95% CI: 3.50-11.45, P < 0.001).

In order to further explore the prognostic value of Risk Score in ovarian cancer samples with different clinicopathological factors (including age, TNM Stage), we regrouped ovarian cancer patients according to these factors and conducted Kaplan-Meier survival analysis. It was found that, the overall survival rates of samples in the high-risk group were significantly lower than those in the low-risk group in early
stage (Stage I + Stage II) and advanced stage (Stage III + Stage IV) samples (Fig. 2B and 2C); also, in samples \( \leq 59 \) years old and samples \( > 59 \) years old (Fig. 2D and Fig. 2E). These results indicate that the Risk Score can be used as an independent indicator to predict the prognosis of ovarian cancer patients.

**Nomogram model could reliably predict the long-term survival of ovarian cancer patients**

Both Stage and Risk Score, these two independent prognostic factors, was used to construct the nomogram model (Fig. 3A). Three lines were draw up to determine the points obtained from each factor in the nomogram for each patient. The sum of these points is located on the "Total Points" axis, and a line from the Total Points axis was drawn to determine the probability of survival of ovarian cancer patients for 1, 3, and 5 years. The corrected curve in the calibration chart is close to the ideal curve (a 45-degree line that passes through the origin of the coordinate axis and has a slope of 1), indicating that the predicted overall survival probability agrees well with the actual result (Fig. 3B and 3D).

**Immune Landscape In Low-risk And High-risk Ovarian Cancer Patients**

We used the CIBERSORT method combined with the LM22 feature matrix to estimate the difference in immune infiltration of 22 kinds of immune cells between high-risk and low-risk ovarian cancer patients. Results of immune cell infiltration in 375 ovarian cancer patients from TCGA were summarized (Fig. 4A). The change in the proportion of tumor infiltrating immune cells in different patients may represent the inherent characteristics of individual differences. There are significant differences in the relative proportion of infiltration of three types of immune cells such as M1 macrophages between the high-risk and low-risk groups (Fig. 4B), and these three significantly different immune cells have a correlation with Risk Score (Fig. 4C). Through PCA analysis, it was found that based on the three types of immune cells with significant differences, samples within low-risk and high-risk groups could be well separated (Fig. 4D).

The expression of immune checkpoints has become a biomarker for ovarian cancer patients to choose personalized immunotherapy[14]. The correlation between the patient's risk score and the key immune checkpoints (CTLA4, PD1, IDO1, TDO2, LAG3, TIGIT) were analyzed and found that the risk score was significantly related to most of them (Fig. 5A). Meanwhile, the expressions of CTLA4 and IDO1 in these 6 immune checkpoints were significantly different between ovarian cancer patients within high-risk and low-risk groups (Fig. 5B). The result implicated that the poor prognosis of patients with high-risk ovarian cancer may be partially contributed by the immunosuppressive microenvironment aberration.

**Discussion**

Ovarian cancer has the highest mortality rate among all gynecological malignancies. Nearly 80% of the patients are diagnosed after the disease has become symptomatic and has progressed to a late stage[15]. The focus of cancer treatment, especially cancer immunotherapeutic strategies, has shifted
from killing the cancer cell by conventional cytotoxic chemotherapy towards re-building the immune system to eliminate cancer cells and preventing the recurrence and relapse[16]. Here is a strong call for the methods to identify and develop clinically valuable gene signatures for OC prognosis, especially when based on comprehensive and unbiased whole-genome data. In this study, we first identified 760 TME-related genes were screened and eventually in the TCGA data set samples. By using the LASSO Cox analysis calculated, we finally determined to select 14 genes (ELN, FBLN1, ANGPT2, FBL, COMP, PPL, PLD2, PDLIM4, EMP1, MYC, ITGAM, FRAT2, WDR45 and ADSL) prognostic signature and scoring system showing strong discriminative power to separate patients with good or poor survival (Fig. 1). Among these, many are already reported contributing to ovarian cancers' pathophysiological features as a single gene. EMP1 promotes the proliferation and invasion of ovarian cancer cells by activating the MAPK pathway[17]. C-myc, a common over-activated oncogene by amplification, has been found in more than 70% advanced ovarian cancer patients[18, 19]. Integrin adhesion signaling has been reported playing a key role in macrophage activation, and Itgam knock out mice, tumor growth, and immunosuppressive cytokine mRNA levels are enhanced[20, 21].

To discover a novel panel of prognostic of TEM-related genes, is the first phase in developing a reliable predictive model to predict the cancer patient outcomes in a clinical practice. After that with the prognostic model we generated in our study, risk factors can cooperate the weight of all 14 key genes was constructed in order to further explore the prognostic value of risk score. In order to further explore the prognostic value of risk score, we regrouped ovarian cancer patients with different clinicopathological factors (age, TNM Stage) and conducted Kaplan-Meier survival analysis. The results showed that the ovarian cancer of patients with a high-risk score have poorer survival comparing with low risk. This risk score can be used as an independent model of prognosis (Fig. 2). In addition, we constructed a nomogram that integrated clinicopathological stage features with risk score to predict the 1, 3, and 5-year survival probability of ovarian cancer patients. The nomogram findings showed that predicted overall survival probability agrees well with the actual result (Fig. 3). The key TEM-related genes showed the best accuracy in predicting the survival of ovarian cancer patients at any time point and would therefore be helpful for the precise treatment of ovarian cancer patients.

We applied settled algorithm CIBERSORT to estimate the fractions of immune cells. In the present studies, M1 macrophages, M2 macrophages and follicular helper T cells were associated with risk score. In contrast, naive B cells, CD8 T cells, CD4 memory activated T cells and eosinophils didn't show the correlation with risk score, which were the protective factors of patients. Immune cells are very complex and heterogeneous in vivo, different subsets of T cells play different roles in the immune response[22]. And Naive B cells is one of the mechanisms of autoimmune tolerance[23]. In order to express of immune checkpoints to help with personalized immunotherapy. We also proved that these risk scores are related to all key immune checkpoint genes named CTLA4, PD1, IDO1, TDO2, LAG3, TIGIT (Fig. 5). Lastly, CTLA4 and IDO1 expression was found to be highly associated with the risk score. The patients with low risk scores had increased expression levels of CTLA4 and IDO1 indicating a potentially high response rate to CTLA4 and IDO1. Though the early study showed CTLA4 blockade has shown minimal activity in ovarian cancer models[24], recent studies reported CTLA4 blockade can boost the expansion of tumor-reactive
CD8+ tumor-infiltrating lymphocytes in ovarian cancer [25]. In the present date, tailor targeted therapy against IDO1 may enhance the effectiveness of treatment. IDO1 inhibitors have been designed, screened, and tested in preclinical models of disease, but still requires verification in clinical trials and across cancer diagnoses.

**Conclusions**

In the present work, we are concerned with the genetic characteristics of the microenvironment. Our systematic analysis and assessment demonstrated that 14 TME related genes might play a pivotal role in the prognosis of ovarian cancer patients. Risk score may help to outline the prognosis of patients with ovarian cancer. Targeting on TME that proves to be successful in the preclinical phase, an enormous challenge still lies ahead for translating these strategies to clinical practice.

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

All the co-authors agree to the publication of the article.

**Availability of data and material**


**Competing interests**

The authors declare that they have no conflict of interest.

**Funding**

Not applicable.

**Authors' contributions**

Xiazi Nie, Bo Qu put forward the ideas of this article, written this article and analysed the data. Lina Song, Xiaohua Li and Yirong Wang helped for acquisition of data and analysis and interpretation of data. Bo Qu helped for revising the manuscript. All authors read and approved the final manuscript.

**Acknowledgments:**
References


Figures
Figure 1

Construction of ovarian cancer prognosis model. (A) Forest map of 48 single-factor analysis of tumor microenvironment (TME) genes significantly associated with the prognosis of ovarian cancer. HR, Hazard ratio. 95% CI, 95% confidence interval. (B) The graph of the tuning parameter lambda in the LASSO regression model. The horizontal axis is log (lambda), and the vertical axis is partial likelihood deviation. The Lambda value corresponding to the smallest value is the best, the Lambda value after taking Log and is below the dotted line, and the number correspondingly on the top is the number of variables. (C) Kaplan Meier survival curve in TCGA data set. The horizontal axis is time and the vertical axis is survival rate. Different colors represent different groups. The P value is based on the log-rank test. (D) Kaplan Meier survival curve in the meta-GEO data set. (E) The heat map of the mRNA expression of the 14 selected genes in the high-risk and low-risk samples in the TCGA dataset. The horizontal axis is the sample, and the vertical axis is the gene. Red means high expression, blue means low expression, and the color of the sample is marked with different colors on the heat map.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Hazard ratio</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>375</td>
<td>1.00 (0.99, 1.02)</td>
<td>0.46</td>
</tr>
<tr>
<td>Riskscore</td>
<td>375</td>
<td>6.33 (3.50, 11.45)</td>
<td>0.001</td>
</tr>
<tr>
<td>Stage I</td>
<td>1</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Stage II</td>
<td>24</td>
<td>0.11 (0.01, 0.92)</td>
<td>0.04</td>
</tr>
<tr>
<td>Stage III</td>
<td>293</td>
<td>0.13 (0.02, 0.96)</td>
<td>0.05</td>
</tr>
<tr>
<td>Stage IV</td>
<td>57</td>
<td>0.08 (0.01, 0.63)</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Figure 2

Risk score is an independent prognostic marker of ovarian cancer. (A) Multi-factor Cox regression analysis forest map. Compared with the reference sample, samples with a Hazard ratio greater than 1
have a higher risk of death, and samples with a Hazard ratio less than 1 have a lower risk of death. (B-C) Kaplan Meier survival curves of ovarian cancer samples with different stages. (D-E) Kaplan Meier survival curves of ovarian cancer samples of different age groups.

**Figure 3**

Nomogram predicts the survival of ovarian cancer patients. (A) Nomogram predicting the probability of overall survival at 1, 3, and 5 years in ovarian cancer patients. (B-D) The calibration curve for the nomogram predicting the probability of overall survival at 1, 3, and 5 years in ovarian cancer patients. The X axis represents the predicted survival rate indicated by the nomogram, and the Y axis represents the actual survival rate.
Figure 4

Immune infiltration in high-risk and low-risk ovarian cancer patients. (A) The relative immune infiltration rate in all patients. (B) Violin chart of immune cells with significant differences in the high-risk and low-risk groups. The horizontal axis is the high-risk and low-risk groups, and the vertical axis is the relative infiltration rate of immune cells. The p value is calculated by the Wilcoxon method. (C) The Chord diagram showing the correlation between the risk score and the three significantly different immune cells, the thicker the link between them represents the stronger the correlation. (D) PCA three-dimensional clustering map. Different colored dots represent different types of samples.
Figure 5

The relationship between important immune checkpoints and Risk Score. (A) The Chord diagram showing the correlation between the risk score and the expression of the five prominent immune checkpoints. The thicker the connection between them indicates the stronger the correlation. (B) Violin charts of immune checkpoints with significantly different expression levels in the high-risk and low-risk groups. Different colors represent the high-risk and low-risk groups. The vertical axis is the expression level. The p-value is calculated using the Wilcoxon method.

Supplementary Files

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

- `TableS1.xls`