Development of a novel transcription factors-related prognostic signature for serous ovarian cancer

He Li

Hunan Cancer Hospital

Zhao-Yi Liu

Hunan Cancer Hospital

Naiyuan Wu

Hunan Cancer Hospital

Yong-Chang Chen

Hunan Cancer Hospital

Quan Cheng

Xiangya Hospital

Jing Wang (✉ wangjing0081@hnca.org.cn)

Hunan Cancer Hospital <https://orcid.org/0000-0001-7683-6259>

Primary research

Keywords: Transcription factors, serous ovarian cancer, prognosis, risk score, model

Posted Date: October 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-82117/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on March 30th, 2021. See the published version at <https://doi.org/10.1038/s41598-021-86294-z>.

Abstract

Background

Ovarian cancer (OC) is one of the most lethal gynecological cancer globally. Serous ovarian cancer (SOC) is most common pathological type of ovarian cancer. Growing evidence suggests that transcription factors (TFs) play vital roles in serous ovarian cancer (SOC). In the present study, we aimed to develop a transcription factors-related prognostic signature to better predict the survival in patients with SOC.

Materials and methods

The TFs mRNA expression profiles of 564 SOC subjects in the TCGA database, and 70 SOC subjects in the GEO database were screened. Lasso cox regression and consensus clustering analysis were utilized to construct a novel TFs related signature and cluster model, respectively. Genomic alternative analysis was used to elaborate on the association between the TFs related prognostic signature and genomic aberrations. GSEA analysis was applied to identify the biological functional difference between high risk-score group and low-risk score group.

Results

A 17-TFs related prognostic signature was constructed using lasso cox regression and validated in the TCGA and GEO cohorts. Consensus clustering analysis was applied to establish a cluster model. The 17-TFs related prognostic signature, risk score, and cluster models were effective at accurately distinguishing the prognosis of SOC. The GSEA assay results suggested that there was a significant difference in the inflammatory and immune response pathways between the high-risk and low-risk score groups. Immune infiltration, immunotherapy, and chemotherapy responses were analyzed due to the significant difference in the regulation of lymphocyte migration and T cell-mediated cytotoxicity between the two groups; indicating that patients with low-risk score were more likely to respond anti-PD-1, etoposide, paclitaxel, and veliparib but not to gemcitabine, doxorubicin, docetaxel, and cisplatin. Also, the prognostic nomogram model revealed that the risk score was a good prognostic indicator for SOC patients.

Conclusions

In conclusion, we explored the prognostic values of TFs in SOC and developed a 17-TFs related prognostic signature to predict the survival of SOC patients.

Background

Ovarian cancer (OC) is one of the most lethal gynecological cancer globally, with 295,414 new cases and 184,799 deaths in 2018 [1]. The disease encompasses a heterogeneous group of malignancies with epithelial ovarian cancer (EOC) accounting for 90% of all cases. Based on tumor cell histology, EOC is classified as serous, endometrioid, mucinous, clear cell, others, or unspecified. Data shows that more than

50% of EOC is serous ovarian cancer (SOC) [2]. Approximately 60% of OC patients are diagnosed with late-stage disease, partly due to limited test methods. Platinum-based chemotherapy is the main treatment for OC patients; a high proportion of OC patients suffer from platinum resistance, leading to insignificant improvement in the 5-year survival rate in the past decade [3].

Multiple prognostic biomarkers have been discovered in OC [4]. For instance, CA125 was the first identified prognostic biomarker and is used clinically [5]. The mutation status of BRCA1/2 is proven to be associated with patient prognosis and chemosensitivity [6]. Several promising prognostic OC biomarkers have been recently discovered such as; VEGF, HE4, mesothelin, M-CSF, osteopontin, kallikrein(s), bikunin, EphA2, and soluble EGF receptor [4]. With the development of multi-omics, numerous genomic variations, miRNAs, lncRNAs, and metabolites have been suggested to serve as prognostic biomarkers [7–12].

It is difficult to blame OC on a single specific issue due to its heterogeneity [13]. Models integrating multiple genes spring up in OC and other cancers as the global gene expression data becomes more feasible to use. So far, autophagy, DNA repair pathway, and the Notch pathway-related prognostic signatures have been developed for OC [14–16]. Gene methylation panel, miRNAs-based signature, miRNA-lncRNA signature were established because of the vital role of epigenetics in OC prognosis [17–19]. Recently, a novel and non-invasive radiomic prognostic signature was derived based on computed tomography (CT) images, which significantly improved prognostic methods and could be exploited to guide personalized therapy of OC [20].

Transcription factors participate in various biological processes including cancer proliferation, invasion and migration, cell cycle, apoptosis, EMT, differentiation, and drug resistance [21]; therefore, numerous TFs have been suggested to be potential treatment targets and prognostic biomarkers in various cancers [21, 22]. However, it is not clear whether TFs can be utilized to predict the prognosis of SOC. Therefore, we aimed to explore the prognostic values of TFs in SOC and to identify a novel TFs related prognostic signature for SOC.

Materials And Methods

Datasets and preprocessing

Gene expression profiles and corresponding clinical information were downloaded from the TCGA data source (<https://xena.ucsc.edu>) and the GEO data portal (<https://www.ncbi.nlm.nih.gov/geo/>). All TCGA samples were randomly divided into two equal cohorts: training cohort (n=282) and validation cohort (n=282). The total TCGA data (n=564) were used to validate as sum cohort. The GEO cohort (n=70) was used as the external validation set. The clinical characteristics of all subjects were listed in Additional file 1: Table S1. All eukaryotic TFs were downloaded from MSigDB (http://software.broadinstitute.org/gsea/msigdb/gene_families.jsp).

L1-penalized (Lasso) cox regression

A robust likelihood-based survival modeling was applied to reduce the dimension and obtain robust survival-associated TFs. The Lasso regression was used to identify the TFs with independent prognostic value. Through 5,000 cross-validations, a set of prognostic genes and their LASSO coefficients were obtained according to the highest lambda value, using the Lasso method (lambda 1st) with 10-fold cross-validation. Based on the LASSO coefficients, we established a TFs-related prognostics risk score model [23]. The association between risk score, TFs expression and clinical characters was illustrated by heat maps

Receiver operating characteristic (ROC)

ROC curve and area under the curve (AUC) were applied to evaluate the prediction accuracy of the risk score model in the training, test, sum, and GEO cohorts. Besides, it was also used to estimate the prediction performance of the nomogram.

Overall survival analysis

The SOC patients were divided into high-risk and low-risk score groups according to the median TFs related prognostic signature or median TFs mRNA expression. Survival curves were illustrated by the Kaplan-Meier method with the long-rank test. The overall survival rate of high and low clusters was also calculated. Furthermore, stratified survival analyses were performed based on different ages, FIGO stage, tumor grade, and residual tumor size.

Genetic alterations analysis

Genetic alterations including somatic mutations and somatic copy number alterations (SCNAs) were investigated. All the samples obtained from the TCGA database were divided into four groups based on the values of the TFs-related prognostic signature and risk score. The Top 20 genes with the highest mutation frequency in the lowest and highest group were selected. The SCNAs of both groups were identified using the Genomic Identification of Significant Targets in Cancer 2.0 (GISTIC) [24].

Gene set enrichment analysis (GSEA)

The GSEA annotation was adopted using the “clusterProfiler” package, to identify potential differences in biological function between SOC patients with high-risk and low-risk scores [25].

Consensus clustering analysis

Samples were divided into different clusters using consensus clustering analysis with the R package “ConsensusClusterPlus” [26]. The subsampling parameter was 80% with 1000 times and k (the number of clusters) ranged from 2 to 10. The optimal number of clusters was selected based on cumulative distribution function plots and consensus matrices [27]. Clinical features and different expressions of TFs between clusters were visualized by heat maps.

Gene set variation analysis (GSVA)

The 17-TFs signature for each SOC sample was established using gene sets enrichment analysis (GSEA). The heat map was used to illustrate the correlation between the 17-TFs, their mRNA expression and the clinical features of SOC. The GSEA package was utilized to reveal the functional enrichment status of these TFs in SOC. The cutoff of the Correlation coefficient was set as 0.5 ($p < 0.05$) [28]. Gene Ontology (GO) terms, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were obtained from MSigDB [29].

Immune infiltration, immunotherapy and chemotherapy response prediction

The infiltration of 22 subtypes of tumor-infiltrating immune cells (TIICs) was calculated by CIBERSORT algorithm, to better understand the immune response function of TFs in SOC [30]. Dysfunction and Exclusion (TIDE) algorithm was employed to estimate the individual likelihood of responding to immunotherapy [31].

There are currently no approved immune therapies for ovarian cancer. Therefore, the response to anti-CTAL-4 or anti-PD-1 therapy was analyzed based on the treatment results of melanoma patients in the subclass analysis [32].

The chemotherapy response of each patient was evaluated using the Genomics of Drug Sensitivity in Cancer database (GDSC, <https://www.cancerrxgene.org>). Drugs that are used for SOC treatment were selected for assessment. The half-maximal inhibitory concentration (IC₅₀) was calculated and represented the drug response. The R package 'pRRophetic' was used with 10-fold cross-validation and other parameters by default [33].

Prognostic Model construction based on clinical features and risk score

Univariate cox proportional hazard regression analysis was performed to select risk factors ($p < 0.05$). A multivariate COX model based on the selected features and the nomogram chart was constructed using the "RMS" in R software to predict 5-year overall survival [34]. The accuracy of the risk model was assessed using the calibration curve and AUC. Finally, the prognostic model value was calculated according to the formula, value = $\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$ (β , regression coefficient; X, prognostic factors).

Statistical analysis

All statistical analyses were performed using R software (version 3.5.3). Significant quantitative differences between groups were analyzed using the two-tailed Students' t-test, whereas differences among groups were analyzed using the one-way ANOVA. The chi-square test was used to analyze the clinical features of different cohorts. Kaplan–Meier method with log-rank test was used to calculate the overall survival rate of SOC patients. Univariate and multivariate Cox regression analyses were performed to evaluate the prognostic value of the risk score and clinical features. The partition around medoids (PAM) algorithm was conducted in the consensus clustering analysis. The correlation between two variables was analyzed with Spearman rank ($p < 0.05$).

Results

Clinical characteristics of the study patients

The TFs mRNA expression profiles of 564 SOC subjects in the TCGA database and 70 SOC subjects in the GEO database were screened. All patients from the TCGA database were randomly separated into training (n=282) and test (n=282) cohorts and there was no difference between two cohorts in age ($p=0.4104$), FIGO stage ($p=0.6542$), Grade ($p=0.6890$) and living status ($p=0.6393$) (Additional file 2: Table S2). Patients from the GEO database were used as the external validation cohort. The workflow indicated in Fig. 1 shows the process of the study.

Construction and validation of TFs-related risk signature for SOC

Univariate cox regression analysis was performed to construct a TFs-related risk signature of SOC; 88 TFs were selected ($p<0.05$) (Additional file 3: Table S3) and employed into the lasso cox regression model (Fig. 2a-b). Finally, a TFs-related risk signature comprising 17 different TFs (*CBX5*, *CREB3*, *FOXJ1*, *FOXK2*, *IRF4*, *LHX2*, *RB1*, *SPDEF*, *STAT2*, *TBX2*, *TEAD1*, *TFAM*, *TRIM38*, *ZHX3*, *ZNF124*, *ZNF8* and *ZXDB*) was constructed in the training group (Fig. 2c). The risk score of each patient was calculated according to their 17-TFs expression. Patients were divided into high-risk and low-risk groups, based on the median risk value as a cutoff point (cutoff = -2.047), to investigate the association between the TFs-related risk signature and prognosis of SOC patients (Fig. 2d). The results showed that high-risk TFs-related risk signature was associated with poor overall survival (OS) in the training cohort (Fig. 2d-e). The prediction feature was assessed by the ROC curve and the AUC value of 0.803 (Fig. 2f). Similarly, patients with high-risk scores had a poor OS in the test cohort, sum cohort, and external validation cohort ($p<0.0001$, $p<0.0001$, $p=0.031$, respectively, Fig. 3). All these findings indicate that there was a significant association between the TFs-related risk signature and SOC patients' prognosis.

The association between the risk signature and various clinicopathologic features including grade, FIGO stage, and residual tumor size was analyzed. The results showed that there was no difference in risk score among different grades ($p>0.05$, Additional file 4: Figure S1). Patients with FIGO I and II had higher risk scores compared with patients with FIGO III and IV, ($p<0.01$, Additional file 4: Figure S1). There was also a significant association between residual tumor size and risk score ($p<0.001$, Additional file 4: Figure S1). A stratified analysis based on age, grade, FIGO grade, and residual tumor size was constructed. The high-risk group had shorter OS than the low-risk group in all these stratified cohorts, indicating that the TFs-related risk signature could be used to accurately distinguish the prognosis of SOC patients, without considering the clinical parameters ($p<0.01$, Fig. 4).

Association between the TFs-related risk signature and genetic alterations

Genetic alterations including somatic mutations and SCNAs were identified in the lowest risk (1st) and highest risk (4th) groups. There was no difference between the two groups in the mutation frequency of TP53, TIN, MUC16, DNH3, CSMD3, FAT3, FLG, and APOB (Fig. 5a-b). Patients in the lowest risk group had

a higher probability of somatic mutation in the DNAB3 gene, while a higher mutation frequency of KMT2C and USH2A occurred in the highest risk group (Fig. 5a-b). We also analyzed the SCNAs between both groups (Fig. 5c-d). The G values between the highest-risk and lowest risk groups were significantly different, although the amplification (including 19q.12, 8q24.21, 3q36.2 and 11q14.1) and focal peaks (including 19p13.3, 22q13.32, 5q12.1, and 17q11.2) were detected. Additionally, several regions of amplifications harboring multiple oncogenes, such as 8q24.21 (PTV1, MYC) 19q13.2 (NFKB1B), 19p13.12 (NOTCH3) and 15q26.3 (IGF1R) were only detected in the highest risk group. Moreover, focal deletion peaks were detected in the group with high-risk scores such as 9p21.3 (CDKN2A, CDKN2B) and 10q23.31 (FAS) (Fig. 5c-d).

Consensus clustering analysis to construct the cluster model

All SOC samples from the TCGA database were divided into different clusters using consensus clustering analysis; two, was found to be the optimal number of clusters (Additional file 5: Figure S2). The principal component analysis (PCA) explored the significant difference in mRNA expression of the 17 TFs between the two clusters (TF1 vs TF2) (Fig. 6a). There was a significant difference in OS between the two clusters ($p=0.00023$, Fig. 6b). The TF1 cluster was related to a lower risk score, while it had no association with age, FIGO stage, grade, and residual tumor size (Fig. 6b-c).

Pathways enrichment analysis to identify the biological function of TFs in SOC

Firstly, GSEA analysis was performed to identify the differences in biology function among patients with high-risk and low-risk scores. The results indicate differences in several pathways including; chronic inflammatory response, positive regulation of vascular endothelial growth factor, regulation of T cell-mediated cytotoxicity, tumor necrosis factor biosynthetic process, MyD88-dependent toll-like receptor signaling pathway, positive regulation of coagulation, and regulation of lymphocyte migration (Fig. 7a). Then, GSVA analysis was executed to reveal the functional enrichment status of the 17-TFs in SOC. Signal pathways with a high correlation coefficient and statistical significance were selected from GO gene sets and KEGG pathway analysis. The higher gene set enrichment scores of all the selected signaling pathways were associated with higher risk scores, except the glutathione metabolic process and IL-4 response pathways. Some pathways, including the TGF- β signaling and IL-4 response, were mainly related to the inflammatory response (Fig. 7b). Cell proliferation, invasion, and migration-related pathways including; PI3K signaling, cell cycle arrest, Wnt signaling, Notch signaling, mTOR signaling, MAPK signaling, and VEGF signaling, were also enhanced in the high-risk score group. The biological processes involved in each TF were also analyzed, indicating that they had different biological functions in SOC (Additional file 6: Figure S3).

Different immune infiltration, immunotherapy and chemotherapy response between high-risk score and low-risk score groups

There was a significant difference between the high-risk and the low-risk score groups in the regulation of lymphocyte migration and T cell-mediated cytotoxicity. Therefore, immune infiltration, immunotherapy,

and chemotherapy responses were analyzed. Data showed that the distribution of aDC, B cells, NK CD56bright cells, NK CD56dim cells, T cells, TFH and Th2 cells was lower in the high-risk than the low-risk score group, whereas mast cells, NK cells, and Tcm cells infiltration was higher in the high-risk than the low-risk score group (Fig. 8a). The difference in immune infiltration between two clusters (TF1 vs TF2) was also determined (Additional file 7: Figure S4). The potential response to immunotherapy for each patient was assessed using the TIDE algorithm. The results suggest that patients with low-risk scores are more sensitive to immunotherapy than those with high-risk scores ($p=0.0003$, Fig. 8b). Subsequently, the response of anti-CTL4 and anti-PD-1 therapy was explored; patients in the low-risk score group were more likely to respond to anti-PD-1 therapy ($p<0.001$, Fig. 8c).

Chemotherapy is the main treatment method for SOC and it plays a key role in patients' survival. The different response to chemotherapy between the high-risk and low-risk score groups was noted. The cell line data obtained from GDSC database was used to train the predictive model and the IC_{50} values for etoposide, paclitaxel, veliparib, gemcitabine, doxorubicin, docetaxel, and cisplatin of each patient with SOC were determined; the values were calculated by the predictive model. The results indicate that there was a significant difference in the estimated IC_{50} between the high-risk and low-risk score groups; patients with low-risk scores showed better response to etoposide, paclitaxel, and veliparib ($p=3.33e-6$; $p=0.003$; $p=0.040$, respectively, Fig. 8d). However, the response to gemcitabine, doxorubicin, docetaxel, and cisplatin ($p=0.077$; $p=0.131$; $p=0.169$; $p=0.182$, respectively) was similar for patients in the two groups (Additional file 8: Figure S5).

Modeling prognostic nomogram for overall survival

Only 503 SOC patients were included in modeling prognostic nomogram for overall survival due to the loss of clinical data. First, the independent prognostic factors were identified using univariate and multivariate cox regression (Table 1). The results indicated that risk score ($p<0.001$, HR=10.0841), residual tumor size (1-10 mm: $p<0.01$, HR=1.5889; 10-20 mm: $p=0.008$, HR=2.0247; >20 mm: $p<0.001$, HR=2.0447), and age ($p=0.02$, HR=0.7586) were all independent prognostic factor for 5-year overall survival. The FIGO stage was adopted as one of the independent factors given that the FIGO III+IV parameters were statistically significant in univariate cox regression ($p=0.03$, HR=2.884), and showed marginal statistical difference in multivariate cox regression ($p=0.133$, HR=1.7485). There was no association between grade, stage, and 5-year overall survival of patients. Therefore, risk score, residual disease, age, and FIGO stage were used to construct the nomogram, and each variable was assigned a score on the corresponding axis. The probabilities of the outcomes were determined by the location of the total score on the survival axes, which was calculated by adding up all the variables (Fig. 9a). The nomogram-predicted 5-year overall survival closely corresponded with the observations (Fig. 9b). Calibration curves were used to evaluate the accuracy of the nomogram with an AUC value of 0.736 (Fig. 9c). The prognostic model value was calculated according to prognostic factors and regression coefficient. Patients were divided into two groups by setting the median prognostic value as cutoff; there was a significant difference in the 5-year overall survival rate between the groups ($p<0.001$, Fig. 9d).

Discussion

Ovarian cancer has the highest mortality rate among gynecological cancers. It is reported that more than 90% of OC patients suffer from the EOC subtype. Based on tumor cell histology, ECO is classified as serous (> 50% of cases), endometrioid, mucinous, clear cell, and others or unspecified. Approximately 60% of OC patients are diagnosed with late-stage disease due to the lack of early and effective test methods, leading to poor clinical outcomes. Moreover, the 5-year survival rate has marginally improved in the past decade. Transcription factors play vital roles in various biological processes including tumorigenesis and drug resistance, and may serve as biomarkers to predict the prognosis of SOC.

In this study, a 17-TFs related prognostic signature and cluster model was introduced to predict the clinical outcome of SOC. In detail, 564 SOC patients from the TCGA database were randomly divided into the training and test cohorts. The 70 SOC patients from GEO were deemed as the external validation set. A 17-TFs related prognostic signature was obtained in the training cohort by successively performing univariate cox regression and Lasso cox regression. The 5-year overall survival rate was significantly different between the high-risk and low-risk score groups. The 17-TFs related prognostic signature showed high prediction performance with a high AUC value (AUC = 0.803), and this was validated in the test, sum, and GEO cohorts. Genomic alterations assay results showed that the risk score of SOC patients was associated with genomic aberrations. Besides, the cluster model was constructed using consensus clustering analysis. Function annotation was performed using GSEA and GSVA analysis. The GSVA results revealed that the prognostic TFs were mainly related to the TGF- β , PI3K, mTOR, MAPK, VEGF and Wnt signaling pathways, which play important roles in cell proliferation, invasion, and migration. The TGF- β signaling pathway, which participates in immune response, was screened from both GO gene sets and the KEGG pathway. The GSEA results indicate that there were significant differences in the chronic inflammatory and immune responses between the high-risk and low-risk score groups. The regulation of lymphocyte migration, immune infiltration, immunotherapy, and chemotherapy response between the high-risk score and low-risk score groups were explored due to the big difference in regulating T cell-mediated cytotoxicity. The distribution of several immune cells including T cells, B cells, and NK cells was found to be different between the two groups. Moreover, patients with low-risk scores showed higher response to immunotherapy, especially anti-PD-1 therapy, and chemotherapy drugs including etoposide, paclitaxel, and veliparib. Finally, a nomogram model was constructed combined with clinical data and risk scores. The prediction utility was very high with an AUC value of 0.736 in the TCGA cohort. Multiple gene signatures have been developed to predict overall survival, recurrence, and platinum sensitivity in OC. Some gene signatures were also used to predict the overall survival outcome in SOC [14, 35, 36]. Liu *et al.* developed a seven gene-based novel signature to predict clinical outcomes of High-Grade Illc Serous Ovarian Carcinoma with an AUC of 0.715 in the TCGA cohort [35]. A genomic rearrangement signature was constructed in High-Grade Serous Ovarian Cancer (HGSOC) with an AUC of 0.63 in the TCGA cohort [36]. A novel autophagy-related prognostic signature was exploited for SOC with an AUC of 0.703 [14]. All these known gene signatures have lower prediction capacities, compared with the 17-TFs related prognostic signature.

TFs play important roles in various biological processes including cancer proliferation, invasion and migration, cell cycle, apoptosis, EMT, differentiation, and drug resistance in OC. The 17 TFs, except FOXK2, LHX2, SPDEF, TRIM38, ZHX3, and ZXDB, were reported to be involved in the malignant phenotype of OC. The CBX5, a member of chromobox (CBX) family, was suggested to function as a prognostic biomarker and potential target for various cancers; including gastric cancer, breast cancer, and lung cancer [37–40]. The expression of CBX5 was elevated in gastric cancer tumor tissues and it promoted cell proliferation, migration, and invasion [38]. It was highly expressed in breast cancer tissues and correlated to relapse-free survival (RFS) for the patients [39]. Besides, CBX5 was identified as a potential target in regulating lung cancer survival and the stem-like properties of lung CD133⁺-tumor stem-like cells (TSLCs) [40]. It was reported that BRD4 suppression enhanced the cytotoxicity of CHK1 inhibition in HGSOC, which was mediated by the increased expression of CBX5 [41]. The CREB3 marker was identified as an oncogene that plays an important role in tumor metastasis in various cancers; including prostate cancer, breast cancer, and osteosarcoma [42–44]. It is reported that the CREB3 antigen may serve as an effective biomarker with high diagnostic sensitivity and specificity in OC, although its prognostic role is unclear [45]. The FOXK2 and FOXJ1 are members of the forkhead box (FOX) transcription factors. Growing evidence shows that FOXK2 participates in a range of key processes in cancer cells, such as cell proliferation, invasion and migration, DNA damage, and autophagy [46]. The function of FOXK2 is context and tumor-specific, and it might serve as an oncogene and tumor suppressor in different cancers [47]. It was reported that FOXK2 inhibited cell proliferation and invasion in breast cancer, non-small cell lung cancer (NSCLC), clear-cell renal cell carcinoma, gastric cancer, and indicated favorable prognosis [48–51]. However, it promoted cell growth and indicated unfavorable prognosis in colorectal cancer and hepatocellular carcinoma [52, 53]. The role of FOXK2 in OC is currently unknown. Meanwhile, FOXJ1, negatively regulated by NANOG, exerts anti-cancer proliferation, migration, and invasion activity in OC [54]. Mainly expressed in plasma cells, IRF4 was reported to be associated with tumor stage, histological grade, and completeness of chemotherapy [55]. Multiply evidence has shown that TEAD1 and TFAM are oncogenes in OC, which is concordant with our results [56, 57]. The role of ZNF124 in OC is little known, but it was found to inhibit apoptotic death and function as an oncogene in hematopoietic cells [58]. The pRB protein, coded by the RB1 gene, is a famed tumor suppressor in OC, which is similar to our results, and was related to acquired chemotherapy resistance and OC recurrence [59, 60]. The ectopic expression of TBX2 is associated with conferred resistance to platinum-based chemotherapy, indicating that it may serve as a predictive marker of the efficacy of platinum-based chemotherapy for patients with OC [61]. Recently, STAT2 was identified to be highly expressed in OC, and the patients involved had poor OS and PFS. It was reported that STAT2 functions as an oncogene by increasing the expression of EGFR [62].

This study has some limitations: 1.) Venous and lymphatic invasion were not considered because the data was missing for most patients. 2.) The sample size in the external validation set was small and the predictive ability of the 17-TFs related prognostic signature was not validated in the samples.

In conclusion, we obtained a TFs related prognostic signature and constructed a clinical model for predicting prognosis of SOC patients, which was able to effectively distinguish high-risk populations from

the TCGA and GEO datasets.

Conclusions

We identified a 17-TFs related prognostic signature (CBX5, CREB3, FOXJ1, FOXK2, IRF4, LHX2, RB1, SPDEF, STAT2, TBX2, TEAD1, TFAM, TRIM38, ZHX3, ZNF124, ZNF8 and ZXDB) to predict survival of SOC. Then, function annotation was performed using GSEA and GSVA analysis, indicating that there were significant differences in the chronic inflammatory and immune responses between high-risk score and low-risk score groups. Besides, the regulation of lymphocyte migration, immune infiltration, immunotherapy, and chemotherapy response between the high-risk score and low-risk score groups were significantly different. Finally, a nomogram model was constructed combined with clinical data and risk scores, indicating that the risk score was a good prognostic indicator for SOC patients.

Abbreviations

OC: ovarian cancer; SOC:serous ovarian cancer; TFs:transcription factors; EOC:epithelial ovarian cancer; TCGA:The Cancer Genome Atlas; ROC:receiver operating characteristic; GSEA:Gene Set Enrichment Analysis; GSVA:gene set variation analysis; OS:overall survival; HR:hazard ratio; MSigDB:Molecular Signatures Database; SCNAs:somatic copy number alterations; KEGG:Kyoto Encyclopedia of Genes and Genomes; TILCs:tumor-infiltrating immune cells

Declarations

Availability of data and materials

All data generated or analyzed during this study were included in this published article and its additional files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that there are no potential conflicts of interest.

Funding

This research was supported by the National Natural Science Foundation of China (NO.81972836), Natural Science Foundation of Hunan Province (2020JJ5338) and National Key R&D Program (2016YFC1303703).

Authors' contributions

HL, ZYL, QC and JW designed the idea and wrote the manuscript. NYYW and YCC contributed the data collection and analysis. QC provided analysis tools. All authors revised the manuscript and gave approval for the version of the manuscript to be published.

Acknowledgements

We would like to thank Dr. Quan Cheng for the bioinformatics technology support.

References

1. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, Jemal A, Kramer JL, Siegel RL. Cancer treatment and survivorship statistics, 2019. *Cancer J Clin.* 2019;69(5):363–85.
2. Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, Gaudet MM, Jemal A, Siegel RL. Ovarian cancer statistics, 2018. *Cancer J Clin.* 2018;68(4):284–96.
3. Holmes D. The problem with platinum. *Nature.* 2015;527(7579):218–9.
4. Huang J, Hu W, Sood AK. Prognostic biomarkers in ovarian cancer. *Cancer Biomark A.* 2010;8(4–5):231–51.
5. Felder M, Kapur A, Gonzalez-Bosquet J, Horibata S, Heintz J, Albrecht R, Fass L, Kaur J, Hu K, Shojaei H, et al. MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress. *Mol Cancer.* 2014;13:129.
6. Ben David Y, Chetrit A, Hirsh-Yechezkel G, Friedman E, Beck BD, Beller U, Ben-Baruch G, Fishman A, Levavi H, Lubin F, et al. Effect of BRCA mutations on the length of survival in epithelial ovarian tumors. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology.* 2002;20(2):463–6.
7. Krzystyniak J, Ceppi L, Dizon DS, Birrer MJ. Epithelial ovarian cancer: the molecular genetics of epithelial ovarian cancer. *Ann Oncol.* 2016;27(Suppl 1(Suppl 1):i4–10.
8. Hentze JL, Høgdall CK, Høgdall EV. Methylation and ovarian cancer: Can DNA methylation be of diagnostic use? *Molecular clinical oncology.* 2019;10(3):323–30.
9. Singh A, Gupta S, Sachan M. Epigenetic Biomarkers in the Management of Ovarian Cancer: Current Prospectives. *Frontiers in cell developmental biology.* 2019;7:182.

10. Llauredó M, Majem B, Altadill T, Lanau L, Castellví J, Sánchez-Iglesias JL, Cabrera S, De la Torre J, Díaz-Feijoo B, Pérez-Benavente A, et al. MicroRNAs as prognostic markers in ovarian cancer. *Molecular cellular endocrinology*. 2014;390(1–2):73–84.
11. Luo P, Liu XF, Wang YC, Li ND, Liao SJ, Yu MX, Liang CZ, Tu JC. Prognostic value of abnormally expressed lncRNAs in ovarian carcinoma: a systematic review and meta-analysis. *Oncotarget*. 2017;8(14):23927–36.
12. Wheeler LJ, Watson ZL, Qamar L, Yamamoto TM, Sawyer BT, Sullivan KD, Khanal S, Joshi M, Ferchaud-Roucher V, Smith H, et al. Multi-Omic Approaches Identify Metabolic and Autophagy Regulators Important in Ovarian Cancer Dissemination. *iScience*. 2019;19:474–91.
13. Kossaï M, Leary A, Scoazec JY, Genestie C. Ovarian Cancer: A Heterogeneous Disease. *Pathobiol J ImmunoPathol Mol Cell Biol*. 2018;85(1–2):41–9.
14. An Y, Bi F, You Y, Liu X, Yang Q. Development of a Novel Autophagy-related Prognostic Signature for Serous Ovarian Cancer. *J Cancer*. 2018;9(21):4058–71.
15. Sun H, Cao D, Ma X, Yang J, Peng P, Yu M, Zhou H, Zhang Y, Li L, Huo X, et al. Identification of a Prognostic Signature Associated With DNA Repair Genes in Ovarian Cancer. *Frontiers in genetics*. 2019;10:839.
16. Chen F, Liu N. A 10-gene expression signature of Notch pathway predicts recurrence in ovarian carcinoma. *Oncology letters*. 2015;10(3):1704–8.
17. Guo W, Zhu L, Yu M, Zhu R, Chen Q, Wang Q. A five-DNA methylation signature act as a novel prognostic biomarker in patients with ovarian serous cystadenocarcinoma. *Clinical epigenetics*. 2018;10(1):142.
18. Li L, Gu H, Chen L, Zhu P, Zhao L, Wang Y, Zhao X, Zhang X, Zhang Y, Shu P. Integrative Network Analysis Reveals a MicroRNA-Based Signature for Prognosis Prediction of Epithelial Ovarian Cancer. *BioMed research international*. 2019;2019:1056431.
19. Zhang Y, Ye Q, He J, Chen P, Wan J, Li J, Yang Y, Li X. Recurrence-Associated Multi-RNA Signature to Predict Disease-Free Survival for Ovarian Cancer Patients. *BioMed research international*. 2020;2020:1618527.
20. Wei W, Liu Z, Rong Y, Zhou B, Bai Y, Wei W, Wang S, Wang M, Guo Y, Tian J. A Computed Tomography-Based Radiomic Prognostic Marker of Advanced High-Grade Serous Ovarian Cancer Recurrence: A Multicenter Study. *Front Oncol*. 2019;9:255.
21. Sirotkin AV. Transcription factors and ovarian functions. *Journal of cellular physiology*. 2010;225(1):20–6.
22. Safe S, Abbruzzese J, Abdelrahim M, Hedrick E. Specificity Protein Transcription Factors and Cancer: Opportunities for Drug Development. *Cancer prevention research (Philadelphia Pa)*. 2018;11(7):371–82.
23. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. *Biometrical journal Biometrische Zeitschrift*. 2010;52(1):70–84.

24. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhir R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome biology*. 2011;12(4):R41.
25. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102(43):15545–50.
26. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26(12):1572–3.
27. Datta S, Datta S. Comparisons and validation of statistical clustering techniques for microarray gene expression data. *Bioinformatics*. 2003;19(4):459–66.
28. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform*. 2013;14:7.
29. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems*. 2015;1(6):417–25.
30. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nature methods*. 2015;12(5):453–7.
31. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nature medicine*. 2018;24(10):1550–8.
32. Roh W, Chen PL, Reuben A, Spencer CN, Prieto PA, Miller JP, Gopalakrishnan V, Wang F, Cooper ZA, Reddy SM, et al: **Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance**. *Science translational medicine* 2017, 9(379).
33. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome biology*. 2014;15(3):R47.
34. Nick TG, Hardin JM. Regression modeling strategies: an illustrative case study from medical rehabilitation outcomes research. *The American journal of occupational therapy: official publication of the American Occupational Therapy Association*. 1999;53(5):459–70.
35. Liu G, Chen L, Ren H, Liu F, Dong C, Wu A, Liu Z, Zheng Y, Cheng X, Liu L. Seven Genes Based Novel Signature Predicts Clinical Outcome and Platinum Sensitivity of High Grade IIIc Serous Ovarian Carcinoma. *Int J Biol Sci*. 2018;14(14):2012–22.
36. Hillman RT, Chisholm GB, Lu KH, Futreal PA. Genomic Rearrangement Signatures and Clinical Outcomes in High-Grade Serous Ovarian Cancer. *J Natl Cancer Inst*. 2018;110(3):265–72.
37. Claerhout S, Lim JY, Choi W, Park YY, Kim K, Kim SB, Lee JS, Mills GB, Cho JY. Gene expression signature analysis identifies vorinostat as a candidate therapy for gastric cancer. *PloS one*. 2011;6(9):e24662.
38. Guo J, Zhang Z, Pan L, Zhou Y. Identification of miR-758-3p as Potential Modulator of CBX5 Expression in Gastric Cancer. *Technology in cancer research treatment*. 2018;17:1533033818816061.

39. Liang YK, Lin HY, Chen CF, Zeng D. Prognostic values of distinct CBX family members in breast cancer. *Oncotarget*. 2017;8(54):92375–87.
40. Yu YH, Chiou GY, Huang PI, Lo WL, Wang CY, Lu KH, Yu CC, Alterovitz G, Huang WC, Lo JF, et al. Network biology of tumor stem-like cells identified a regulatory role of CBX5 in lung cancer. *Scientific reports*. 2012;2:584.
41. Pongas G, Kim MK, Min DJ, House CD, Jordan E, Caplen N, Chakka S, Ohiri J, Kruhlak MJ, Annunziata CM: **BRD4 facilitates DNA damage response and represses CBX5/Heterochromatin protein 1 (HP1)**. *Oncotarget* 2017, **8**(31):51402–51415.
42. Kim Y, Kim J, Jang SW, Ko J. The role of sLZIP in cyclin D3-mediated negative regulation of androgen receptor transactivation and its involvement in prostate cancer. *Oncogene*. 2015;34(2):226–36.
43. Howley BV, Link LA, Grelet S, El-Sabban M, Howe PH. A CREB3-regulated ER-Golgi trafficking signature promotes metastatic progression in breast cancer. *Oncogene*. 2018;37(10):1308–25.
44. Wu Y, Xie Z, Chen J, Chen J, Ni W, Ma Y, Huang K, Wang G, Wang J, Ma J, et al. Circular RNA circTADA2A promotes osteosarcoma progression and metastasis by sponging miR-203a-3p and regulating CREB3 expression. *Mol Cancer*. 2019;18(1):73.
45. Antony F, Deantonio C, Cotella D, Soluri MF, Tarasiuk O, Raspagliesi F, Adorni F, Piazza S, Ciani Y, Santoro C, et al. High-throughput assessment of the antibody profile in ovarian cancer ascitic fluids. *Oncoimmunology*. 2019;8(9):e1614856.
46. Chen Y, Wu J, Liang G, Geng G, Zhao F, Yin P, Nowsheen S, Wu C, Li Y, Li L, et al. CHK2-FOXK axis promotes transcriptional control of autophagy programs. *Sci Adv*. 2020;6(1):eaax5819.
47. Nestal de Moraes G, Carneiro LDT, Maia RC, Lam EW, Sharrocks AD. **FOXK2 Transcription Factor and Its Emerging Roles in Cancer**. *Cancers* 2019, **11**(3).
48. Shan L, Zhou X, Liu X, Wang Y, Su D, Hou Y, Yu N, Yang C, Liu B, Gao J, et al. FOXK2 Elicits Massive Transcription Repression and Suppresses the Hypoxic Response and Breast Cancer Carcinogenesis. *Cancer cell*. 2016;30(5):708–22.
49. Chen S, Jiang S, Hu F, Xu Y, Wang T, Mei Q. Foxk2 inhibits non-small cell lung cancer epithelial-mesenchymal transition and proliferation through the repression of different key target genes. *Oncol Rep*. 2017;37(4):2335–47.
50. Zhang F, Ma X, Li H, Zhang Y, Li X, Chen L, Guo G, Gao Y, Gu L, Xie Y, et al. FOXK2 suppresses the malignant phenotype and induces apoptosis through inhibition of EGFR in clear-cell renal cell carcinoma. *International journal of cancer*. 2018;142(12):2543–57.
51. Liu X, Wei X, Niu W, Wang D, Wang B, Zhuang H. Downregulation of FOXK2 is associated with poor prognosis in patients with gastric cancer. *Mol Med Rep*. 2018;18(5):4356–64.
52. Du F, Qiao C, Li X, Chen Z, Liu H, Wu S, Hu S, Qiu Z, Qian M, Tian D, et al. Forkhead box K2 promotes human colorectal cancer metastasis by upregulating ZEB1 and EGFR. *Theranostics*. 2019;9(13):3879–902.

53. Lin MF, Yang YF, Peng ZP, Zhang MF, Liang JY, Chen W, Liu XH, Zheng YL. FOXK2, regulated by miR-1271-5p, promotes cell growth and indicates unfavorable prognosis in hepatocellular carcinoma. *Int J Biochem Cell Biol.* 2017;88:155–61.
54. Siu MK, Wong ES, Kong DS, Chan HY, Jiang L, Wong OG, Lam EW, Chan KK, Ngan HY, Le XF, et al. Stem cell transcription factor NANOG controls cell migration and invasion via dysregulation of E-cadherin and FoxJ1 and contributes to adverse clinical outcome in ovarian cancers. *Oncogene.* 2013;32(30):3500–9.
55. Heimes AS, Schmidt M, Jäkel J, Almstedt K, Gebhard S, Weyer-Eiberich V, Elger T, Krajnak S, Brenner W, Hasenburg A, et al. A retrospective analysis of immunohistochemically determined IRF4 (interferon regulating factor 4) expression in a consecutive cohort of 114 ovarian cancer patients. *Archives of gynecology obstetrics.* 2019;299(1):239–46.
56. Zhou Y, Huang T, Cheng AS, Yu J, Kang W, To KF. **The TEAD Family and Its Oncogenic Role in Promoting Tumorigenesis.** *International journal of molecular sciences* 2016, 17(1).
57. Signorile A, De Rasmio D, Cormio A, Musicco C, Rossi R, Fortarezza F, Palese LL, Loizzi V, Resta L, Scillitani G, et al: **Human Ovarian Cancer Tissue Exhibits Increase of Mitochondrial Biogenesis and Cristae Remodeling.** *Cancers* 2019, 11(9).
58. Kuramoto K, Uesaka T, Kimura A, Kobayashi M, Watanabe H, Katoh O. ZK7, a novel zinc finger gene, is induced by vascular endothelial growth factor and inhibits apoptotic death in hematopoietic cells. *Cancer research.* 2000;60(2):425–30.
59. Patch AM, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S, Nones K, Cowin P, Alsop K, Bailey PJ, et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature.* 2015;521(7553):489–94.
60. Du ZH, Bi FF, Wang L, Yang Q. Next-generation sequencing unravels extensive genetic alteration in recurrent ovarian cancer and unique genetic changes in drug-resistant recurrent ovarian cancer. *Mol Genet Genom Med.* 2018;6(4):638–47.
61. Tasaka R, Fukuda T, Shimomura M, Inoue Y, Wada T, Kawanishi M, Yasui T, Sumi T. TBX2 expression is associated with platinum-sensitivity of ovarian serous carcinoma. *Oncology letters.* 2018;15(3):3085–90.
62. Chen X, Huang J, Lü Y. [High expression of STAT2 in ovarian cancer and its effect on metastasis of ovarian cancer cells]. *Nan fang yi ke da xue xue bao = Journal of Southern Medical University.* 2020;40(1):34–41.

Tables

Table 1. Univariable and multivariate cox regression analysis of factors affecting overall survival of SOC patients.

Variables	Univariable analysis		Multivariate analysis	
	HR (95% CI)	p value	HR (95% CI)	p value
Risk score	13.895 (7.637-25.281)	<0.001***	10.0841 (5.4247-18.7454)	<0.001***
Residual disease (1-10mm)	2.069 (1.471-2.912)	<0.001***	1.5889 (1.1173-2.2596)	<0.01**
Residual disease (10-20mm)	2.3 (1.389-3.809)	<0.001***	2.0247 (1.2066-3.3974)	0.008**
Residual disease (>20mm)	2.829 (1.925-4.157)	<0.001***	2.0447 (1.3745-3.0415)	<0.001***
Age (<50 year)	0.646 (0.514-0.813)	<0.001***	0.7586 (0.6009-0.9578)	0.020*
FIGO stage (Ⅲ+Ⅳ)	2.884 (1.428-5.823)	0.003**	1.7485 (0.8437-3.6237)	0.133
Grade stage (Ⅲ+Ⅳ)	1.297 (0.929-1.81)	0.126	1.0583 (0.7465-1.5003)	0.750

HR: Hazard Ratio; CI: Confidence Interval; * $p<0.05$; ** $p<0.01$; *** $p<0.001$

Figures

Figure 1

The workflow of this study.

Figure 2

A 17-TFs related prognostic signature was established in TCGA training cohort. (a-c) The lasso coefficients of the 17 TFs for overall survival in TCGA training cohort; (d) The distribution of risk score

and alive status in TCGA training cohort; (e, f) The 17-TF related prognostic signature predicts overall survival in SOC and the prediction performance was evaluated by calibration curve.

Figure 3

The 17-TFs related prognostic signature was verified in TCGA test cohort (a-c), TCGA sum cohort (d-f) and GEO cohort (g-i).

Figure 4

Construct a stratified analysis based on age (a, b), grade (c, d), FIGO grade (e, f) and residual resection status (g-j) to analyze the association between the TFs related prognostic signature and overall survival.

Figure 5

Genetic alterations including somatic mutations (a, b) and SCNAs (c, d) assigned to high-risk group (a, c) and low-risk group (b, d); The green line represents the significance threshold (q value=0.25).

Figure 7

Functional annotation of TFs related prognostic signature. (a, b) GSEA assay to identify the different pathways between high-risk score group and low-risk score group; (c, d) GSVA assay to explore the distribution of risk scores and clinical features (upper), and gene set enrichment of different pathways in GO terms and KEGG pathways (lower).

Figure 8

Difference between high-risk score group and low-risk score group in immune infiltration, immunotherapy and chemotherapy response prediction. (a) immune infiltration proportion of TIICs between high-risk score group and low-risk score group; (b) the TIDE value and response to immunotherapy of patients with SOC; (c) the response to anti-CTAL4 and anti-PD-1 therapy of patients with SOC in both groups; (d)

Estimated IC50 values indicate the efficiency of chemotherapy by etoposide, paclitaxel and veliparib in both groups.

Figure 9

Prognostic nomogram for predicting the 5-year overall survival of SOC patients. (a) Prognostic nomogram for glioma patients was created based on clinical risk factors; (b) The calibration curve of 5-year overall survival. The predicted probability of overall survival is plotted on the x-axis and the observed

OS is plotted on the y-axis; (c) ROC curve from the nomogram of 5-year overall survival; (d) Comparison of overall survival between two groups divided by the median value of the prognostic model (Low vs High).

Supplementary Files

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

- [supplementTable1.docx](#)
- [supplementTable1.docx](#)
- [supplementTable1.docx](#)
- [SupplementTable2.docx](#)
- [SupplementTable2.docx](#)
- [SupplementTable2.docx](#)
- [Supplementtable3.xlsx](#)
- [Supplementtable3.xlsx](#)
- [Supplementtable3.xlsx](#)
- [supplmentFigure1.tif](#)
- [supplmentFigure1.tif](#)
- [supplmentFigure1.tif](#)
- [supplementfigure2.tif](#)
- [supplementfigure2.tif](#)
- [supplementfigure2.tif](#)
- [supplmentFigure301.tif](#)
- [supplmentFigure301.tif](#)
- [supplmentFigure301.tif](#)
- [supplmentFigure40101.tif](#)
- [supplmentFigure40101.tif](#)
- [supplmentFigure40101.tif](#)
- [supplmentFigure501.tif](#)
- [supplmentFigure501.tif](#)
- [supplmentFigure501.tif](#)