A 19-miRNA Support Vector Machine classifier and a 6-miRNA risk score system designed for ovarian cancer patients

JINGWEI DONG and MINGJUN XU

Department of Anesthesiology, Beijing Obstetrics and Gynecology Hospital, Dongcheng, Beijing 100001, P.R. China

Received October 29, 2018; Accepted March 18, 2019

DOI: 10.3892/or.2019.7108

Abstract. Ovarian cancer (OC) is the most common gynecologic malignancy with high incidence and mortality. The present study aimed to develop approaches for determining the recurrence type and identify potential miRNA markers for OC prognosis. The miRNA expression profile of OC (the training set, including 390 samples with recurrence information) was downloaded from The Cancer Genome Atlas database. The validation sets GSE25204 and GSE27290 were obtained from the Gene Expression Omnibus database. Prescreening of clinical factors was conducted using the survival package, and the differentially expressed miRNAs (DE-miRNAs) were identified using the limma package. Using the Caret package, the optimal miRNA set was selected to build a Support Vector Machine (SVM) classifier. The miRNAs and clinical factors independently related to prognosis were analyzed using the survival package, and the risk score system was constructed. Finally, the miRNA-target regulatory network was built by Cytoscape software, and enrichment analysis was performed. There were 46 DE-miRNAs between the recurrent and non-recurrent samples. After the optimal 19-miRNA set was selected for constructing the SVM classifier, 6 DE-miRNAs (miR-193b, miR-211, miR-218, miR-505, miR-508 and miR-514) independently related to prognosis were further extracted to build the risk score system. The neoplasm cancer status was independently correlated with the prognosis and conducted with stratified analysis. Additionally, the target genes in the regulatory network were enriched in the regulation of actin cytoskeleton and the TGF-β signaling pathway. The 6-miRNA signature may serve as a potential biomarker for OC prognosis, particularly for recurrence.

Introduction

Ovarian cancer (OC) is the most lethal gynecologic malignancy with a 5-year overall survival (OS) of ~47%, which has almost never changed over the past 20 years (1). In 2015, 1.2 million women suffered from OC, and the disease led to 161,100 deaths worldwide (2). The symptoms of OC are inconspicuous and non-specific thus most cases are diagnosed at a later stage (3). Therefore, early diagnosis and treatment of OC are critical for improving the outcomes of the disease, and prognosis mainly depends on the disease degree, tumor subtypes and medical conditions (2,4). Understanding the underlying mechanisms of OC could facilitate the development of advanced treatment approaches.

MicroRNAs (miRNAs) play important roles in OC pathogenesis and progression. By comparison of the transcriptome data from different tissues with genome-scale biomolecular networks, miR-124-3p was identified as a potential biomarker for OC (5). miR-27a is considered as an oncogene which inhibits forkhead box O1 (FOXO1) in OC (6), while miR-34a serves as a suppressor by downregulating histone deacetylase 1 (HDAC1) (7). miR-409-3p was found to enhance the cisplatin-sensitivity of OC cells by inhibiting autophagy controlled by Fip200 (8).

In addition, a growing number of studies have shown that the dysregulation of miRNAs is associated with the prognosis of OC. The miR-200 family members have been identified as prognostic indicators for the disease stage, tumor histology and survival of OC (9). For example, miR-200b-429 may be a promising marker for OC survival, and the low expression of miR-200 indicates a poor prognosis and plays a regulatory role in the tumor (10). An upregulated serum miR-221 expression level is correlated with tumor stage and grade of epithelial ovarian cancer (EOC), which serves as an independent factor for a poor prognostic in EOC (11). Serum miR-141 and miR-200c can distinguish OC patients from healthy controls, and they may be utilized as markers for predicting the prognosis of OC (12). A high expression level of miR-203 has been reported as a candidate marker that predicts the progression and adverse outcome of patients with EOC (13,14). Serum miR-21 expression was found to be increased in EOC patients, and it may function as a novel marker for the diagnosis and prognosis of EOC (15). The expression of miR-150 is higher in primary serous OC than in omental metastases, and its lower expression is associated with...
shorter progression free-survival in metastatic tissues (16). Nevertheless, the miRNAs related to the recurrence of OC have not been fully revealed.

Thus, exploring the correlation between miRNAs and the development and recurrence of OC is critical for improving the curative effects and prognosis of OC patients. Based on the miRNA expression profile of OC in the public database, the miRNAs correlated with the recurrence of OC were screened, and then a classifier was constructed to recognize the recurrence of OC. Combined with the prognostic information of the samples, a risk score system was constructed based on the expression levels of significant miRNAs. The present study may provide a theoretical basis for the prognostic prediction and targeted therapy of OC, particularly recurrent OC.

Materials and methods

Data source and prescreening of clinical factors. The miRNA expression profile of OC (the training set) was downloaded from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) database (September 10, 2018), based on the Illumina HiSeq 2000 RNA Sequencing platform. The data in the ‘0a07b199-d93d-4202-a63a-b38e39dc5ca4.mirbase21.mirnas.quantification.txt’ file that is level 3 was downloaded and used. Then we used the encoding information to obtain the sample information. There were 415 OC samples with available clinical information in the training set, of which 390 had information regarding recurrence: 170 were non-recurrent OC samples and 220 were recurrent. The human reference genome hg38/GRCh38 was used to annotate the expression information.

Meanwhile, other relevant datasets were searched from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database with the keywords ‘ovarian cancer’ and ‘Homo sapiens’. The inclusive criteria were: i) the samples in the dataset had recurrent information; ii) the samples had prognostic information; iii) the total number of samples was no <50; and iv) the dataset was an miRNA expression profile. Based on these criteria, only two datasets, GSE25204 (17) and GSE27290 (18), were selected and used as validation datasets. The GSE25204 dataset was based on the Illumina Human v2 MicroRNA expression beadchip platform and included 85 OC samples with recurrent information (validation set 1). The GSE27290 dataset was based on the Agilent-015508 Human miRNA Microarray platform (pre-commercial version 6.0) and contained 58 OC samples with recurrent information (the validation set 2).

The clinical information of all the samples in the training set was statistically analyzed. In order to determine the basis for grouping, the univariate and multivariate Cox regression analysis in the R survival package (19) was used to select the DE-miRNAs significantly related to prognosis, with the threshold of log-rank P<0.05. The DE-miRNAs significantly related to recurrent prognosis were further selected to perform the follow-up analyses.

Data standardization and differential expression analysis. The miRNA expression profile matrices of the three datasets were stacked, and the matrix for each dataset was scaled according to the expression level. The unit specification was scaled as follows to provide a sample vector:

\[ v = (v_1, ..., v_n), \]

\[ v_{\text{normalized}} = \frac{1}{\|v\|_2} \]

where \(\|v\|_2\) is the 2-norm of the vector (\(l_2\) norm).

Using the sqrt [sum(data^2)] function in R (20), the square root of the eigenvalue of matrix \(B = A^TA\) was obtained. The purpose of this normalization was to obtain the sample values scaled to 1. Using median scaling, the expression level of each miRNA was centralized and normalized according to the median and median absolute deviation (MAD). Specifically, an eigenvector \(x = (x_1, ..., x_n)\):

\[ \text{mad}(x) = \text{median}(\{|x_i - \text{median}(x)|, x_i \in x\}) \]

\[ x_{\text{scaled}} = (x - \text{median}(x)) \cdot \frac{1}{\text{mad}(x)} \]

For the training set, the samples were grouped according to the recurrence condition of the samples. Using the R package limma (21) (version 3.34.7, https://bioconductor.org/packages/release/bioc/html/limma.html), the differentially expressed miRNAs (DE-miRNAs) between the recurrent and non-recurrent OC samples were selected. The false discovery rate (FDR) <0.05 and log_fold_change (FC) >0.263 were set as the thresholds for the significant differences. According to the expression levels of the DE-miRNAs in the training datasets, the bidirectional hierarchical clustering for expression levels of these DE-miRNAs was performed based on the centered Pearson correlation algorithm, using the heatmap package (22) (version 1.0.8, https://cran.r-project.org/web/packages/heatmap/index.html) in R.

Construction of Support Vector Machine (SVM) classifier. In the training set, the Cox regression analysis in the R survival package (19) was used to select the DE-miRNAs significantly related to prognosis, with the threshold of log-rank P<0.05. The DE-miRNAs significantly related to recurrent prognosis were further selected to perform the follow-up analyses.

Recursive feature elimination (RFE) is an integrated machine learning method, which considers the selection of subset as an optimization problem (23). Using the RFE algorithm in the R package Caret (24) (version 6.0-76, https://cran.r-project.org/web/packages/caret), the optimal miRNA set was filtered from the training dataset. In the 100-fold cross validation, the miRNA with the highest accuracy was selected as the signature miRNA.

SVM is a supervised classification algorithm of machine learning, which discriminates sample types by estimating the probability that a sample belongs to a certain category (25). For the training set, the SVM classifier was constructed based on the optimal miRNA set using the SVM method (Core: Sigmoid Kernel; Cross: 100-fold cross validation) in the R package e1071 (26) (version 1.6-8, https://cran.r-project.org/web/packages/e1071).

The performance of the SVM classifier was separately evaluated in the training set and the validation sets using 4 valuation indicators (Concordance index, C-index; Brier score;
Log-rank P-value of Cox-proportional hazard (PH) regression; and area under the receiver operating characteristic (ROC) curve, AUC. The C-index and Brier score were calculated using the R package survcomp (27) (version 1.30.0, http://www.bioconductor.org/packages/release/bioc/html/survcomp.html). Using the R package survival (19), the Kaplan-Meier (KM) curves for the two groups classified by the SVM classifier were generated, and the log-rank P-value of the difference between the two groups was calculated. Furthermore, the indicators of ROC curves (sensitivity, Sen; specificity, Spe; positive prediction value, PPV; negative prediction value, NPV) were calculated using the R package pROC (28) (version 1.12.1, https://cran.r-project.org/web/packages/pROC/index.html).

### Construction of the risk score system

Based on the multivariate Cox regression analysis in the R survival package (19), the prognosis-associated miRNAs were further analyzed to identify the DE-miRNAs independently related to prognosis. The log-rank P<0.05 was set as the threshold.

Based on the regression coefficients of the independent prognostic miRNAs, the risk score system was constructed, and the risk score of each sample was obtained according to the following formula:

\[ \text{Risk score} = \sum \text{Coef}_{\text{DE-miRNA}} \times \text{Exp}_{\text{DE-miRNA}} \]

where Coef\text{DE-miRNA} represents the regression coefficient, and Exp\text{DE-miRNA} indicates the expression level of the corresponding miRNA.

For the training set, the samples were divided into high- and low-risk groups with the median of risk scores as the cut-off point. Using the KM curve analysis in the R survival package (19), the correlation between the risk score system and prognosis was evaluated. Meanwhile, the risk score system was confirmed in the validation sets.

### Stratified analysis of clinical factors

Using the univariate and multivariate Cox regression analysis in the R survival package (19), the clinical factors independently correlated with prognosis in the training set were screened out. Combined with the high- and low-risk samples determined by the risk score system, stratified analysis was further carried out.

### miRNA-target regulatory network analysis and enrichment analysis

The risk scores of the mRNA-sequencing samples matched with the miRNA-sequencing samples were calculated using the risk score system. Based on the risk scores, the samples in the training set were divided into high- and low-risk groups. Using the R package limma (21), the differentially expressed genes (DEGs) between the two groups were selected, with the thresholds of FDR <0.05 and |log2FC| >0.263. Based on the starBase database (29) (version 3.0, http://starbase.sysu.edu.cn/), the miRNA-mRNA regulatory interactions in at least one of the five databases, targetScan, picTar, RNA22, PITA, and miRanda, were selected. Then, the correlation of the expression levels of the miRNAs and target DEGs in the matched samples were calculated, and the interactions with significant negative correlations were selected. Subsequently, the miRNA-target regulatory network was visualized using the Cytoscape software (30) (version 6.8, https://david.ncifcrf.gov/), the functional and pathway enrichment analyses were carried out, with P<0.05 as the screening criterion.

### Results

**Prescreening of clinical factors and differential expression analysis.** The clinical information of the 415 OC samples in the training set was performed with statistical analysis, and
then the clinical factors significantly associated with prognosis were screened. The age, tumor recurrence, and neoplasm cancer status were found to be the clinical factors significantly related to prognosis (Table I and Fig. 1). To identify the recurrence prognosis-associated miRNAs, the samples in this study were grouped based on the recurrence information.

For the training set, a total of 46 DE-miRNAs (18 upregulated and 28 downregulated) were identified between the recurrent and non-recurrent OC samples (Fig. 2A). The clustering heatmap was drawn based on the expression levels of the DE-miRNAs, which indicated that the samples were clearly divided into two types (Fig. 2B).

Construction of the SVM classifier. Using the Cox regression analysis, 24 prognosis-associated miRNAs were selected in the training set. Using RFE algorithm, the optimal miRNA set involving 19 miRNAs (including miR-135b, miR-139, miR-151, miR-187, miR-193b, miR-210, miR-211, miR-218, miR-219, miR-30b, miR-30d, miR-365, miR-505, miR-506, miR-508, miR-509, miR-513c, miR-514 and miR-760) was selected (Fig. 3). Based on the optimal 19-miRNA set, the SVM classifier was constructed. Then, the performance of the SVM classifier in the training set and the validation sets was assessed using the 4 valuation indicators aforementioned. The results showed that the C-index values were >0.80, and Brier score values <0.1 in both the training and validation sets (Table II). As shown in the confusion table diagrams that indicated the sample classification based on the SVM classifier.

Figure 1. The Kaplan-Meier (KM) survival curves for the clinical factors significantly associated with prognosis. (A) The KM curve for age (blue and red curves separately represent the tumor samples with age <60 and >60 years). (B) The KM curve for tumor recurrence (blue and red curves separately represent the tumor samples with information on non-recurrence and recurrence). (C) The KM curve for neoplasm cancer status (blue and red curves separately represent the tumor-free samples and the samples with tumor).

Figure 2. Results of the differential expression analysis for screening differentially expressed miRNAs (DE-miRNAs). (A) The volcano plot of the DE-miRNAs (blue and red dots represent DE-miRNAs with FDR <0.05 and |log2FC| >0.263). (B) Clustering heatmap of the DE-miRNAs (green and red sample bars separately represent the samples with information on non-recurrence and recurrence).

Figure 3. Accuracy curve for screening the optimal miRNA set. The horizontal axis represents the number of miRNA variables, and the vertical axis represents the cross-validation accuracy. The marked content is the number of miRNAs corresponding to the optimal miRNA set.
classifier, the 19-miRNA set could distinguish well the recurrent samples from the non-recurrent (Fig. 4). The AUC values of the training set and the validation sets were >0.9 (Fig. 4 and Table II). The KM curves suggested that the predictive results of the SVM classifier were significantly related to prognosis (P<0.05; Fig. 4). These results indicated that the 19-miRNA-based classifier could accurately determine the recurrence type of the OC samples.

Table II. Evaluation indicators for the Support Vector Machine (SVM) classifier in the training set and the validation sets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>C-index</th>
<th>Brier score</th>
<th>Log rank P-value</th>
<th>ROC AUC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set (TCGA, n=390)</td>
<td>0.942</td>
<td>0.021</td>
<td>4.72x10^{-12}a</td>
<td>0.905</td>
<td>0.847</td>
<td>0.946</td>
<td>0.914</td>
<td>0.829</td>
</tr>
<tr>
<td>Validation set 1 (GSE25204, n=85)</td>
<td>0.899</td>
<td>0.066</td>
<td>7.98x10^{-7}a</td>
<td>0.941</td>
<td>0.783</td>
<td>0.903</td>
<td>0.75</td>
<td>0.918</td>
</tr>
<tr>
<td>Validation set 2 (GSE27290, n=58)</td>
<td>0.842</td>
<td>0.063</td>
<td>6.39x10^{-6}a</td>
<td>0.902</td>
<td>0.875</td>
<td>0.905</td>
<td>0.778</td>
<td>0.95</td>
</tr>
</tbody>
</table>

ROC, receiver operating characteristic; AUC, area under the receiver operating characteristic curve; PPV, positive prediction value; NPV, negative prediction value. aSignificant P-value.
Construction of risk score system. Combining the optimal 19-miRNA set with the recurrence prognosis information of the samples, 6 independent prognosis-related DE-miRNAs (miR-193b, miR-211, miR-218, miR-505, miR-508 and miR-514) were identified (Table III).

Combined with the regression coefficients of the 6 independent prognostic miRNAs, the risk score system for OC was constructed. The formula for calculating the risk score of each sample was:

\[
\text{Risk score} = (0.2618) \times \exp(\text{hsa-miR-193b}) + (0.1507) \times \exp(\text{hsa-miR-211}) + (0.1588) \times \exp(\text{hsa-miR-218}) + (-0.1959) \times \exp(\text{hsa-miR-505}) + (-0.1704) \times \exp(\text{hsa-miR-508}) + (0.2526) \times \exp(\text{hsa-miR-514})
\]

Table III. The 6 differentially expressed miRNAs independently related to prognosis.

<table>
<thead>
<tr>
<th>ID</th>
<th>Coef</th>
<th>P-value</th>
<th>Hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-193b</td>
<td>0.2618</td>
<td>0.0431*</td>
<td>1.2993 (1.008-1.674)</td>
</tr>
<tr>
<td>hsa-miR-211</td>
<td>0.1507</td>
<td>0.0331*</td>
<td>1.1627 (1.012-1.336)</td>
</tr>
<tr>
<td>hsa-miR-218</td>
<td>0.1588</td>
<td>0.0179*</td>
<td>1.1721 (1.028-1.337)</td>
</tr>
<tr>
<td>hsa-miR-505</td>
<td>-0.1959</td>
<td>0.0052*</td>
<td>0.8221 (0.675-0.902)</td>
</tr>
<tr>
<td>hsa-miR-508</td>
<td>-0.1704</td>
<td>0.0181*</td>
<td>0.8433 (0.657-0.983)</td>
</tr>
<tr>
<td>hsa-miR-514</td>
<td>0.2526</td>
<td>0.0211*</td>
<td>1.2874 (1.066-1.913)</td>
</tr>
</tbody>
</table>

CI, confidence interval. *Significant P-value.

With the median of risk scores as the cut-off point, the samples were classified into high- and low-risk groups. For the training and the validation sets, the KM curves showed that the high- and low-risk groups determined by the risk score system were significantly associated with the actual recurrence prognosis information (Fig. 5).
Stratified analysis of the clinical factors. In the training set, although the age, tumor recurrence, and neoplasm cancer status were all identified as prognosis-associated clinical factors, only the neoplasm cancer status was considered as an independent prognostic factor relating to the recurrence, based on the multivariate Cox regression analysis (Table IV and Fig. 6A). To analyze the correlation between the neoplasm cancer status and recurrence prognosis separately in the high- and low-risk groups, stratified analysis was further performed for the neoplasm cancer status (Fig. 6B and C).

miRNA-target regulatory network analysis and enrichment analysis. In total, we identified 615 DEGs (400 upregulated and 215 downregulated) between the high- and low-risk groups. Based on the StarBase database, the target genes were predicted for the 6 independent prognostic miRNAs. The overlapping genes between the target genes and the DEGs were obtained after comparison, and 601 miRNA-mRNA regulatory interactions were selected. Then, 218 interactions with significant negative correlations were retained for constructing the miRNA-target regulatory network (including miR-193b, miR-211, miR-505, miR-508, and miR-514) (Fig. 7). In addition, the target genes in the regulatory network were enriched in 25 functional terms (such as blood vessel development and vasculature development) and 6 pathways (such as regulation of actin cytoskeleton and TGF-β signaling pathway) (Table V).

Discussion

In the present study, we identified 46 DE-miRNAs between the recurrent and non-recurrent ovarian cancer (OC) samples. Nineteen prognosis-associated miRNAs were used to construct an SVM classifier, among which 6 were deregulated and independently related to prognosis. A risk score system based on the 6 miRNAs had a high accuracy for risk prediction in both the training and validation sets. The neoplasm cancer status was a clinical factor independently correlated with recurrence.

miR-193b serves as a tumor suppressor in many cancer types. Its role in OC has recently been investigated. The epigenetic silencing of miR-193a-3p could promote OC progression by targeting the growth factor receptor-bound protein-7 (GRB7) (32). miR-193b-3p has an antitumor effect in OC cells by inhibiting the p21-activated kinase 3 (33). Downregulation of miR-193b could induce OC metastasis (34). These results indicate that miR-193b may be a tumor suppressor in OC. Moreover, low expression of miR-193b is associated with a poor prognosis of OC patients (35). In the present study, miR-193b was one of the 6 miRNA signatures

### Table IV. Cox regression analysis for screening the clinical factor independently correlated with prognosis in the training set.

<table>
<thead>
<tr>
<th>Clinical characteristics</th>
<th>HR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>1.012</td>
<td>0.999-1.024</td>
<td>6.38x10⁻²</td>
</tr>
<tr>
<td>Neoplasm histologic grade (G1/G2/G3/G4/-)</td>
<td>1.338</td>
<td>0.927-1.932</td>
<td>1.19x10⁻¹</td>
</tr>
<tr>
<td>Pathological stage (II/III/IV/-)</td>
<td>1.005</td>
<td>0.753-1.342</td>
<td>9.73x10⁻¹</td>
</tr>
<tr>
<td>Neoplasm cancer status (tumor free/with tumor/-)</td>
<td>0.279</td>
<td>0.129-0.601</td>
<td>5.30x10⁻⁴</td>
</tr>
<tr>
<td>Lymphatic invasion (yes/no/-)</td>
<td>1.078</td>
<td>0.678-1.714</td>
<td>7.52x10⁻¹</td>
</tr>
<tr>
<td>Venous invasion (yes/no/-)</td>
<td>0.619</td>
<td>0.365-1.049</td>
<td>7.21x10⁻²</td>
</tr>
</tbody>
</table>

HR, hazard ratio; CI, confidence interval; *indicates significant P-value.

Figure 6. Kaplan-Meier (KM) survival curves for neoplasm cancer status. (A) The KM curve showing the correlation between neoplasm cancer status and recurrence prognosis in all samples. (B) The KM curve showing the correlation between neoplasm cancer status and recurrence prognosis in the low-risk group. (C) The KM curve showing the correlation between neoplasm cancer status and recurrence prognosis in the high-risk group. Blue and red curves represent the tumor-free samples and the samples with tumor, respectively.
that could predict recurrence of OC, suggesting that its expression also might be linked to recurrence.

Currently, only a few studies have reported the correlations between miR-218 and OC. It was reported that miR-218 prevents the proliferation and invasion in OC by downregulating its target gene, runt-related transcription factor 2 (RUNX2) (36). In colon adenocarcinoma, the long noncoding RNA MNX1-AS1 could promote progression. It acts as a competing endogenous RNA (ceRNA) of miR-218-5p and upregulates SEC61A1, the downstream target gene of miR-218-5p (37). MNX1-AS1 was also found to facilitate the progression of OC (37). However, it is unclear whether MNX1-AS1 also has this competing relationship with miR-218-5p. In the present study, miR-218 was another important miRNA identified to be related to the recurrence of OC, indicating it may be a novel predictive factor for OC recurrence.

It has been found that downregulation of the tumor suppressor miR-211 induces the overexpression of cyclin-dependent kinase 6 (CDK6) and cyclin D1 (CCND1), which contributes to the proliferation of epithelial ovarian cancer (EOC) cells (38). The high expression of PHF19 (PHD finger protein 19) was found to be related to the poor prognosis of OC patients, and it is a target of miR-211. By competing with the lncRNA MALAT1, miR-211 suppresses the expression of PHF19, and thus functions as a suppressor in OC development (39). This suggests that miR-211 is linked to OC prognosis by regulating the downstream targets. However, there are no studies reporting the role of miR-211 in the recurrence of OC. Based on the present study, it was one of the 6 miRNA signatures for OC recurrence. Therefore, miR-211 may be a potential biomarker indicative of the recurrence of OC.

miR-30a, miR-30e and miR-505 exhibit significantly lower expression in ovarian clear cell carcinoma (OCC) compared with those in elderly advanced ovarian papillary serous carcinoma (OPSC) patients, and the activating transcription factor 3 (ATF3) is the primary gene co-targeted by them (40). Overexpression of the tumor suppressors miR-130b-3p, miR-509-3p, miR-509-5p, miR-508-3p and miR-508-5p has been association with the improved survival of OC patients, and these miRNAs may alter the physical properties of OC cells via regulating the actin cytoskeleton (41). Moreover, by downregulating gene expression levels in the MAPK1/ERK signaling pathway, miR-508 acts as an inhibitor for cell proliferation, migration and invasion in OC cells (42). Reduced miR-514 is correlated with adverse prognosis of OC patients, and miR-514 can inhibit cell proliferation and lower cisplatin chemosensitivity in OC by regulating the ATP binding cassette subfamily (43). These findings indicate that the three miRNAs, miR-505, miR-508 and miR-514 may function as suppressors in OC development and their low expression could be associated with poor prognosis. Our results demonstrated that

Figure 7. miRNA-target regulatory network. Squares and circles represent miRNAs and mRNAs, respectively.
miR-505, miR-508 and miR-514 were three miRNA signatures in recurrent OC, suggesting that they may be the predictive indicators for OC recurrence.

For the target genes in the miRNA-target regulatory network (involving miR-193b, miR-211, miR-505, miR-508 and miR-514), they were significantly enriched in the regulation of the actin cytoskeleton and the TGF-β signaling pathway. The TGF-β signaling pathway functions in various cellular processes correlated with tumorigenesis, and the genetic variants in the pathway are related to OC risk and may help to identify high-risk individuals (44). TGF-β signaling can be suppressed by the accumulation of epigenetic modifications, which contributes to the oncogenesis of OC (45). The dynamic remodeling of the actin cytoskeleton is important for multiple cellular activities, and dysfunction of cytoskeletal proteins can lead to many diseases in humans (46). Therefore, miR-193b, miR-211, miR-505, miR-508 and miR-514 may influence the prognosis of OC via the regulation of the actin cytoskeleton and the TGF-β signaling pathway.

Although we performed comprehensive bioinformatic analyses using the miRNA expression profile of OC and confirmed the classification accuracy by the validation data sets, several limitations remain. First, the sample size with available recurrence information was small. Second, this study lacks validation experiments to validate the expression of these predictive miRNAs and interplayed target genes. Third, the accuracy of the SVM classifier and clinical value of the 6-miRNA risk score system should be further tested in OC patients. Therefore, further experiments should be prepared and conducted to support our findings.

In conclusion, the SVM classifier may be accurate in determining the recurrence status of OC patients. Moreover, the 6-miRNA risk score system may be effective in predicting the outcome of OC patients. Furthermore, miR-193b, miR-211,

Table V. Gene Ontology (GO) functional terms and pathways enriched in the target genes involved in the regulatory network.

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>Count</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO biology process</td>
<td>GO:0001568→blood vessel development</td>
<td>16</td>
<td>4.26x10⁻⁷</td>
</tr>
<tr>
<td></td>
<td>GO:0001944→vasculature development</td>
<td>16</td>
<td>5.81x10⁻⁷</td>
</tr>
<tr>
<td></td>
<td>GO:0001525→angiogenesis</td>
<td>10</td>
<td>9.95x10⁻⁵</td>
</tr>
<tr>
<td></td>
<td>GO:0007155→cell adhesion</td>
<td>22</td>
<td>1.72x10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>GO:0022610→biological adhesion</td>
<td>22</td>
<td>1.75x10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>GO:0048514→blood vessel morphogenesis</td>
<td>11</td>
<td>3.19x10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>GO:0006928→cell motion</td>
<td>14</td>
<td>6.66x10⁻³</td>
</tr>
<tr>
<td></td>
<td>GO:0051674→localization of cell</td>
<td>10</td>
<td>1.51x10⁻²</td>
</tr>
<tr>
<td></td>
<td>GO:0048870→cell motility</td>
<td>10</td>
<td>1.51x10⁻²</td>
</tr>
<tr>
<td></td>
<td>GO:0008219→cell death</td>
<td>16</td>
<td>3.56x10⁻²</td>
</tr>
<tr>
<td></td>
<td>GO:0016265→death</td>
<td>16</td>
<td>3.75x10⁻²</td>
</tr>
<tr>
<td></td>
<td>GO:0006915→apoptosis</td>
<td>14</td>
<td>3.87x10⁻²</td>
</tr>
<tr>
<td></td>
<td>GO:0012501→programmed cell death</td>
<td>14</td>
<td>4.27x10⁻²</td>
</tr>
<tr>
<td>GO cellular component</td>
<td>GO:0031012→extracellular matrix</td>
<td>14</td>
<td>3.11x10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>GO:0005578→proteinaceous extracellular matrix</td>
<td>12</td>
<td>1.89x10⁻³</td>
</tr>
<tr>
<td></td>
<td>GO:0099986→cell surface</td>
<td>11</td>
<td>1.03x10⁻²</td>
</tr>
<tr>
<td></td>
<td>GO:0044421→extracellular region part</td>
<td>20</td>
<td>2.47x10⁻²</td>
</tr>
<tr>
<td></td>
<td>GO:0005794→golgi apparatus</td>
<td>18</td>
<td>3.72x10⁻²</td>
</tr>
<tr>
<td>GO molecular function</td>
<td>GO:0030246→carbohydrate binding</td>
<td>13</td>
<td>1.10x10⁻³</td>
</tr>
<tr>
<td></td>
<td>GO:0005198→structural molecule activity</td>
<td>17</td>
<td>3.75x10⁻³</td>
</tr>
<tr>
<td></td>
<td>GO:0032555→purine ribonucleotide binding</td>
<td>32</td>
<td>2.72x10⁻²</td>
</tr>
<tr>
<td></td>
<td>GO:0032553→ribonucleotide binding</td>
<td>32</td>
<td>2.72x10⁻²</td>
</tr>
<tr>
<td></td>
<td>GO:0017076→purine nucleotide binding</td>
<td>33</td>
<td>2.89x10⁻²</td>
</tr>
<tr>
<td></td>
<td>GO:0001882→nucleoside binding</td>
<td>28</td>
<td>4.24x10⁻²</td>
</tr>
<tr>
<td></td>
<td>GO:0005524→ATP binding</td>
<td>26</td>
<td>4.52x10⁻²</td>
</tr>
<tr>
<td>Pathway</td>
<td>hsa04512:ECM-receptor interaction</td>
<td>8</td>
<td>6.16x10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>hsa04510:Focal adhesion</td>
<td>10</td>
<td>6.47x10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>hsa04810:regulation of actin cytoskeleton</td>
<td>7</td>
<td>4.57x10⁻³</td>
</tr>
<tr>
<td></td>
<td>hsa03010:ribosome</td>
<td>4</td>
<td>8.91x10⁻³</td>
</tr>
<tr>
<td></td>
<td>hsa04350:TGF-beta signaling pathway</td>
<td>3</td>
<td>2.87x10⁻²</td>
</tr>
<tr>
<td></td>
<td>hsa04142:lysosome</td>
<td>3</td>
<td>4.20x10⁻²</td>
</tr>
</tbody>
</table>

*Significant P-value.
miR-505, miR-508 and miR-514 may affect the prognosis of OC via regulation of the actin cytoskeleton and the TGF-β signaling pathway.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

The datasets analyzed in the present study are available from the corresponding author on reasonable request.

Authors’ contributions

JD performed the data analyses and wrote the manuscript. MX conceived and designed the study. JD and MX read and approved the final manuscript. All authors read and approved the manuscript and agree to be accountable for all aspects of the research in ensuring that the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethics approval and consent to participate

In the original studies that generated the datasets, the trials were approved by the local institutional review boards of all participating centers, and informed consent was obtained from all patients.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References


26. Meyer D: Support vector machines the interface to libsvm in Dynamic Languages, pp. 84-95, 2016.


46. Lee SH and Dominguez R: Regulation of actin cytoskeleton dynamics in cells. Mol Cells 29: 311-325, 2010. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.