

Identification and validation of core genes for serous ovarian adenocarcinoma via bioinformatics analysis

RURU ZHU, JISEN XUE, HUIJUN CHEN and QIAN ZHANG

Department of Gynaecology, The First Affiliated Hospital of Wenzhou Medical University,
Wenzhou, Zhejiang 325000, P.R. China

Received June 30, 2019; Accepted May 14, 2020

DOI: 10.3892/ol.2020.12007

Abstract. Ovarian cancer is a fatal gynaecological malignancy in women worldwide, and serous ovarian cancer (SOC) is considered the most common histological subtype of this malignancy. Thus, the present study aimed to identify the core genes for SOC via bioinformatics analysis. The GSE18520 and GSE14407 datasets were downloaded from the Gene Expression Omnibus (GEO) database to screen for differentially expressed genes (DEGs) and perform gene set enrichment analysis (GSEA). A protein-protein interaction (PPI) network was constructed to identify the core genes, while The Cancer Genome Atlas (TCGA) database was used to screen for prognosis-associated DEGs. Furthermore, clinical samples were collected for further validation of kinesin family member 11 (KIF11) gene. In the GEO analysis, a total of 198 DEGs were identified, including 81 upregulated and 117 downregulated genes compared SOC to normal tissue. GSEA across the two datasets demonstrated that 16 gene sets, including those involved in the cell cycle and DNA replication, were notably associated with SOC. A PPI network of the DEGs was constructed with 130 nodes and 387 edges. Subsequently, 20 core genes involved in the same top-ranked module were filtered out by submodule analysis. Survival analysis identified three predictive genes for SOC prognosis, including KIF11, CLDN3 and FGF13. KIF11 was identified as a core and predictive gene and thus was further validated using clinical samples. The results demonstrated that KIF11 was upregulated in tumour tissues compared with adjacent normal tissues and was associated with aggressive factors, including tumour grade, TNM stage and lymph node invasion.

In conclusions, the present study identified the core genes and gene sets for SOC, thus extending the understanding of SOC occurrence and progression. Furthermore, KIF11 was identified as a promising tumour-promoting gene and a potential target for the diagnosis and treatment of SOC.

Introduction

Ovarian cancer is one of the most common types of cancer among women worldwide, and it is the leading cause of mortality among gynaecological malignancies in USA, with 22,240 new cases and 14,030 mortalities in 2013 (1). Despite recent advances in therapeutic strategies against ovarian cancer, its overall mortality remains unchanged, approximately 6.8 per 100,000 females of age-adjusted annual cancer death rate in USA in the past decade (2). This is primarily due to late-stage diagnosis and lack of well-established early detection biomarkers. For instance, CA125 is extensively used as a biomarker for ovarian cancer, its low sensitivity and specificity decrease its efficiency as a diagnostic tool (3). A novel biomarker human epididymis protein 4 exhibits greater specificity and lower sensitivity compared with CA125, which causes it is unsuitable for general population screening (4). Thus, identifying novel biomarkers and understanding their underlying molecular mechanisms on ovarian cancer remains an urgent challenge.

Ovarian cancer comprises heterogeneous histological subtypes, including serous, endometrioid, clear cell and mucinous (5). The serous subtype is the most common and accounts for ~70% of all cases (6). As different histological subtypes display distinct clinical behaviours, an improved understanding of the underlying molecular mechanisms may be obtained by focusing on a certain subtype. Thus, the present study focused on serous ovarian cancer (SOC), which is characterized by tumours situated at the fallopian tube or ovarian surface epithelial-like cells (7).

In the past decade, high-throughput transcriptomics techniques such as microarray sequencing and RNA-seq have been extensively used for discovery and functional studies of several diseases, such as multiple sclerosis, schizophrenia and different types of cancer (8). Several online databases, including the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) are readily available for use of these techniques. Thus, the present study aimed to identify the

Correspondence to: Dr Qian Zhang, Department of Gynaecology, The First Affiliated Hospital of Wenzhou Medical University, 2 Fu Xue Road, Wenzhou, Zhejiang 325000, P.R. China
E-mail: qianzhang19112@outlook.com

Abbreviations: GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas; DEGs, differentially expressed genes; GSEA, gene set enrichment analysis

Key words: core genes, bioinformatics, serous ovarian adenocarcinoma

core genes that may act as potential biomarkers or treatment targets for SOC. To achieve this aim, various bioinformatics approaches, accompanied with online datasets were used, and clinical validation of patient samples were assessed.

Materials and methods

Data collection and identification of differentially expressed genes (DEGs). The gene expression profiles of the GSE18520 (9) and GSE14407 (10) datasets were downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo>) to identify the DEGs. Both datasets were based on the GPL570 Affymetrix Human Genome U133A 2.0 Array platform (Affymetrix; Thermo Fisher Scientific, Inc.). The GSE18520 dataset included 53 SOC samples and 10 normal samples (9), and the GSE14407 dataset included 12 SOC samples and 12 normal samples (10), and all those normal samples were derived from non-ovarian disease subjects. In addition, GSE12470 dataset, including 43 SOC and 10 normal tissue samples, was downloaded from the GEO database to validate gene expression derived from the aforementioned datasets (11).

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. The EGSEA package (version 3.5.0), which integrates multiple gene set enrichment analyses (GSEA) into a single ensemble framework (12), was used to perform GO and KEGG pathway enrichment analyses. $P < 0.05$ was considered to indicate statistically significant difference.

Protein-protein interaction (PPI) network construction. The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (<https://string-db.org/>) was used to acquire the information on the PPIs between the DEGs, and Cytoscape software 3.6.1 (<https://cytoscape.org/>) was used to construct the PPI network. DEGs with at least 10 interactions were considered as initial candidate core genes. Subsequently, the MCODE (version 1.4.2) plugin in Cytoscape was used to identify the modules that contained the core genes within the PPI network (degree cut-off ≥ 2 ; node score cut-off ≥ 0.2 ; K-core ≥ 2 ; max depth, 100).

Identification of genes associated with overall survival. The TCGA-OV RNA-seq datasets (<https://cancergenome.nih.gov>) were accessed and downloaded using the TCGAbiolinks package (version 2.14.1) with the function of 'GDCquery' (project = 'TCGA-OV', data.type = 'Gene expression quantification' and platform = 'Illumina HiSeq') and 'GDCdownload' (13). Individuals with histological subtypes other than serous adenoma or with incomplete data on tumour stage, sex, age at diagnosis, cancer site and survival outcomes were excluded. Following exclusion, a total of 300 serous ovarian adenocarcinoma samples were included in the present study to perform survival analysis. Cox regression analysis was performed to identify the DEGs with potential prognostic values.

Patient enrolment. Given that some SOC samples did not have the adjacent tissues in the selected tissue repository, 40 tumour tissues and 10 adjacent normal tissues were collected from patients with SOC between July 2015 and July 2018. All

patients underwent surgery at the Department of Gynecology, The First Affiliated Hospital of Wenzhou Medical University (Wenzhou, China). Samples were taken within 10 min after excision, and immediately stored in liquid nitrogen until further experimentation. All diagnoses were histologically confirmed and tumour grades were re-evaluated by two independent pathologists who were blinded to the result. The tumours were graded according to the Silverberg grading system (14). The original clinical data were obtained from hospital medical records and included patient age, tumour size, Federation of Gynaecology and Obstetrics stage (15), histological grade, serum CA125 level and lymph node metastasis status. None of the patients had received radiotherapy, chemotherapy or immunotherapy prior to surgery. The present study was approved by the Research Ethics Committee of The First Affiliated Hospital of Wenzhou Medical University (Wenzhou, China), and written informed consent was provided by all participants prior to the study.

Reverse transcription-quantitative (RT-q)PCR. Total RNA was extracted from SOC and adjacent normal tissues using the ReliaPrep RNA Cell Miniprep system (Promega Corporation). RNA was eluted in 60 μ l RNase-free water, and RNA concentration was measured using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Inc.). Total RNA was reverse transcribed into cDNA using the iScript Reverse Transcription kit, and qPCR was subsequently performed using the SYBR[®]-Green Supermix on a Bio-Rad CFX 96 PCR instrument (all from Bio-Rad Laboratories, Inc.). All processes were performed according to the manufacturer's protocols. The primer sequences used for qPCR were designed using Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome) and purchased from Sigma-Aldrich; Merck KGaA (Table I). The following thermocycling conditions were used for qPCR: Initial denaturation 1 min at 95°C; 33 of cycles of denaturation at 95°C for 30 sec, annealing at 60°C for 20 sec and elongation at 72°C for 30 sec; and final extension at 72°C for 5 min. Relative expression levels were calculated using the $2^{-\Delta\Delta C_q}$ method (16) and normalized to the internal reference gene GAPDH. All experiments were performed in triplicate. Then, the SOC samples were divided into two groups based on mean kinesin family member 11 (KIF11) expression, namely high (\geq mean expression) and low ($<$ mean expression) groups.

Statistical analysis. R software 3.6.1 (<https://www.r-project.org>) (17) and Bioconductor packages (<http://www.bioconductor.org>) were used to analyze the GEO and TCGA datasets and produce all figures (18). Unpaired Student's t-test was used to compare continuous variables, whereas χ^2 test was used to compare categorical variables. Univariate Cox regression analysis was performed to determine genes with potential prognostic value in patients with SOC, which were further subjected to multivariate Cox regression analysis.

The Bioconductor edgeR package (version 3.28.1) (<https://www.bioconductor.org/packages/release/bioc/html/edgeR.html>) was used to assess gene expression levels, which was consistent with the method used to analyse the GEO datasets. The limma package (version 3.42.2) (<https://www.bioconductor.org/packages/release/bioc/html/limma.html>) in R software was used

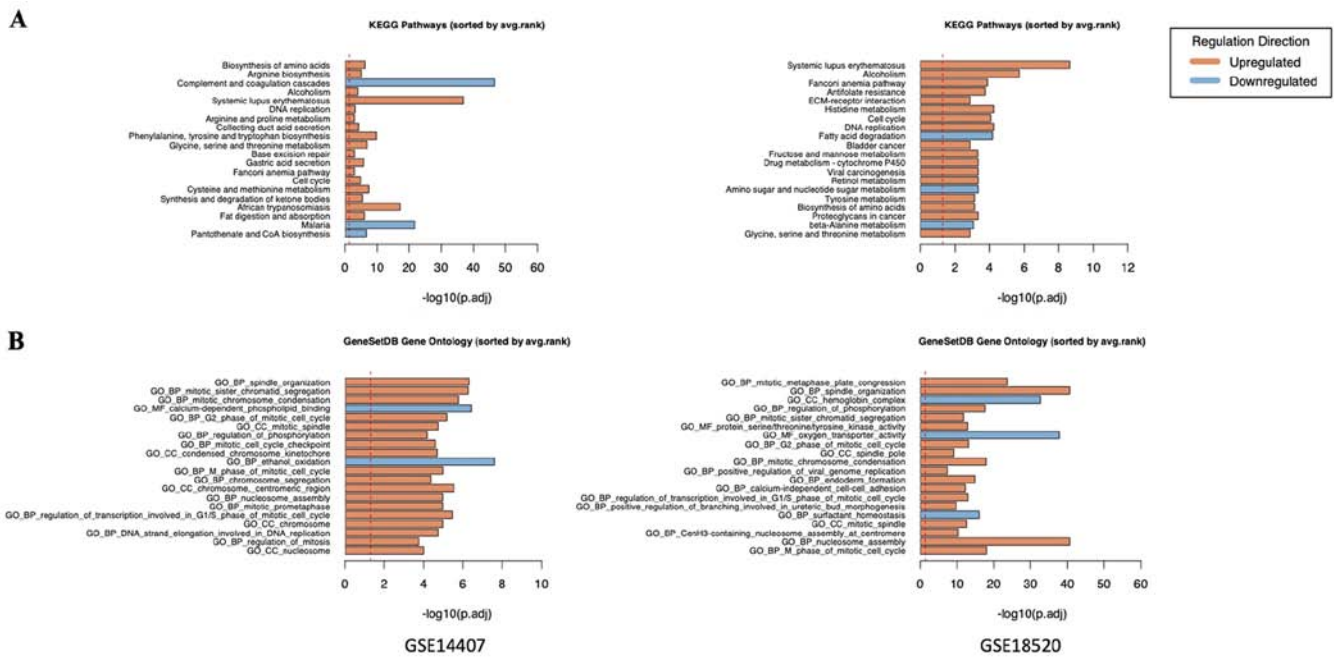


Figure 1. Gene set enrichment analysis based on entire gene expression, including DEGs or non-DEGs. (A) KEGG pathway enrichment analysis. (B) GO gene set enrichment analysis. DEGs, differentially expressed genes; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; avg, average.

to screen DEGs between tumour and normal tissue samples, with adjusted $P < 0.05$ and \log_2 fold-change (FC) > 1 set as the cut-off criteria. Overall survival analysis was performed using the survminer package (version 0.4.6) (<https://cran.r-project.org/web/packages/survminer/index.html>) and log-rank test. For survival analysis with log-rank method, we first determined the optimal cut-off value for continuous gene expression by the maximally selected rank statistics from the maxstat package (version 0.7-25) (<https://cran.r-project.org/package=maxstat>). $P < 0.05$ was considered to indicate a statistically significant difference.

Results

Identification of DEGs between tumour and normal tissue samples. The limma package in R software was used to screen DEGs between tumour and normal tissue samples, with adjusted $P < 0.05$ and \log_2 FC > 1 set as the cut-off criteria. A total of 410 and 804 DEGs were screened from the GSE18520 and GSE14407 datasets, respectively. After filtering overlapping DEGs with consistent differential expression direction, 198 DEGs were identified across the two datasets, including 81 upregulated and 117 downregulated DEGs (Table S1).

GO and KEGG pathway enrichment analyses. GO and KEGG pathway enrichment analyses were performed using the EGSEA package in R software to identify the molecular mechanisms of SOC. The results demonstrated that 'systemic lupus erythematosus', 'alcoholism', 'fanconi anemia pathway', 'cell cycle', 'DNA replication', 'biosynthesis of amino acid' and 'glycine, serine and threonine metabolism' were among the top 20 significantly enriched pathways across the two GEO datasets (Fig. 1A). In addition, GO enrichment analysis indicated that the genes significantly enriched in 'spindle

organisation', 'mitotic sister chromatid segregation', 'mitotic chromosome condensation', 'phase of mitotic cell cycle', 'mitotic spindle', 'regulation of phosphorylation', 'mitotic cell cycle M phase', 'nucleosome assembly' and 'regulation of transcription involved in the G1/S phase of the mitotic cell cycle' were among the top 20 significant gene sets in the two GEO datasets (Fig. 1B).

Identification of core genes using PPI and modular analyses. Cytoscape software and the STRING database were used to identify the core genes. The 198 DEGs were uploaded onto Cytoscape to construct the PPI network, which comprised 130 nodes and 387 edges. Of the 198 DEGs, 68 were extracted from the PPI network, of which 24 critical node genes were identified with the criterion filtering degree ≥ 10 (each core gene had at least 10 interactions; Fig. 2). Then, the MCODE plugin was used to detect significant modules, and the module with the highest degree contained 20 nodes and 189 edges (Fig. 3). Those 20 genes were considered as core SOC genes, included thyroid hormone receptor interactor 13 (TRIP13), RAD51 associated protein 1 (RAD51AP1), DLG associated protein 5 (DLGAP5), kinesin family member 11 (KIF11), kinesin family member 20A (KIF20A), family with sequence similarity 83 member D (FAM83D), protein regulator of cytokinesis 1 (PRC1), cell division cycle associated 5 (CDCA5), BUB1 mitotic checkpoint serine/threonine kinase B (BUB1B), CDC28 protein kinase regulatory subunit 2 (CKS2), kinesin family member 4A (KIF4A), maternal embryonic leucine zipper kinase (MELK), kinesin family member 15 (KIF15), cell division cycle 20 (CDC20), forkhead box M1 (FOXM1), centrosomal protein 55 (CEP55), TTK protein kinase (TTK), ubiquitin conjugating enzyme E2 C (UBE2C), centromere protein F (CENPF) and kinesin family member 14 (KIF14), all of which were significantly associated with the 'cell cycle' ($P = 0.001$), 'ATP-dependent microtubule motor activity,

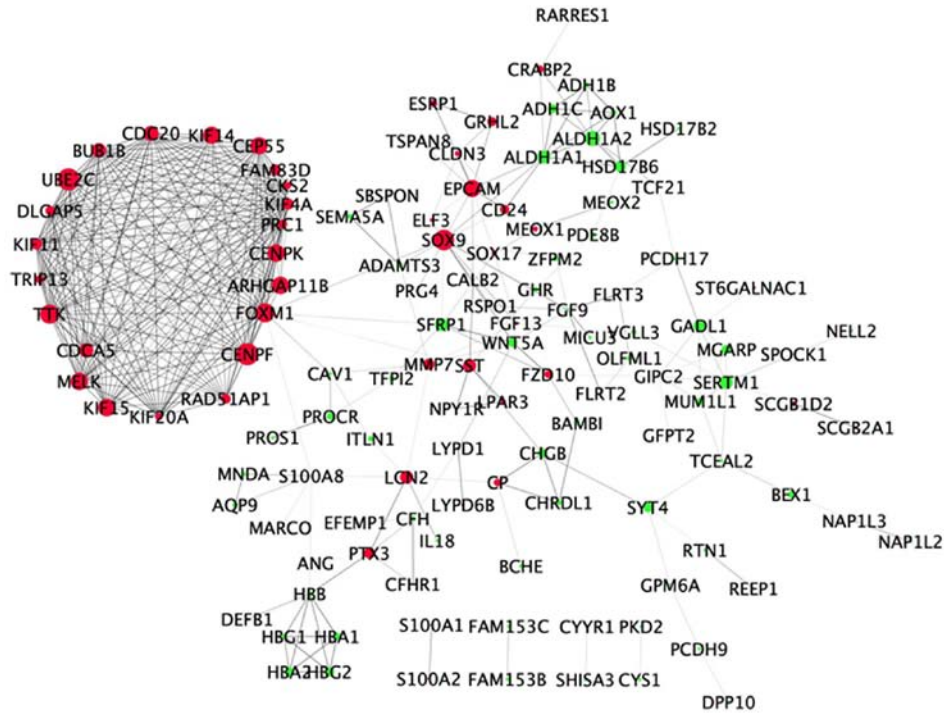


Figure 2. Protein-protein interaction network of differentially expressed genes. Red nodes represent upregulated genes; green nodes represent downregulated genes. Node size indicates the number of interactions. Lines represent interactions between the nodes. The width of the line indicates the combined score from the Search Tool for the Retrieval of Interacting Genes/Proteins database.

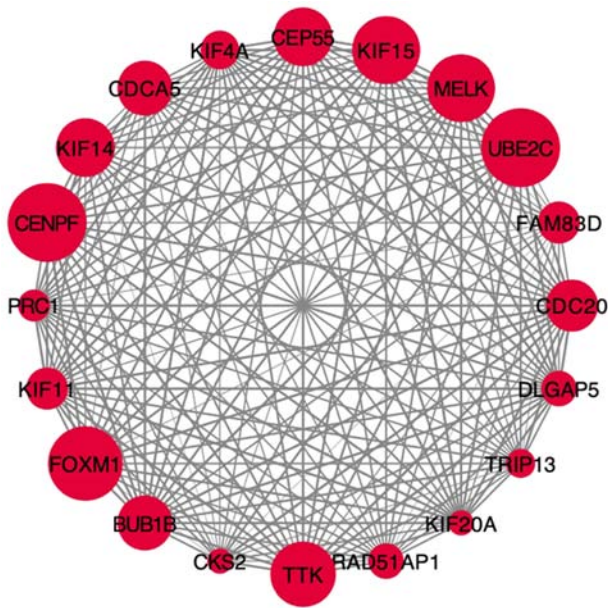


Figure 3. A significant module selected from the protein-protein interaction network. Red nodes represent upregulated genes. Node size indicates the number of interactions. Lines represent interactions between the nodes. The width of the line indicates the combined score from the Search Tool for the Retrieval of Interacting Genes/Proteins database.

plus-end-directed' ($P=0.001$), 'microtubule binding' ($P<0.001$), 'microtubule motor activity' ($P<0.001$), 'tubulin binding' ($P<0.001$), 'motor activity' ($P<0.001$), 'ATPase activity' ($P<0.001$), 'kinesin binding' ($P=0.003$), 'ATP-dependent microtubule motor activity' ($P=0.003$) and 'Ubiquitin mediated proteolysis' ($P=0.016$) (Table SII).

Identification of genes predicting overall survival. Univariate Cox regression analysis was performed to determine the potential prognostic genes for SOC based on TCGA data. A total of 10 significant genes were identified, including butyrylcholinesterase (BCHE) ($P=0.032$), proteoglycan 4 (PRG4) ($P=0.011$), cwcv and kazal like domains proteoglycan 1 (SPOCK1) ($P=0.038$), aquaporin 9 (AQP9) ($P=0.036$), fibroblast growth factor 13 (FGF13) ($P=0.024$), calbindin 2 (CALB2) ($P=0.033$), nucleosome assembly protein 1 like 2 (NAP1L2) ($P=0.002$), claudin 3 (CLDN3) ($P=0.011$), grainyhead like transcription factor 2 (GRHL2) ($P=0.029$) and KIF11 ($P=0.017$), which were subsequently subjected to multivariate Cox regression analysis with adjustment for age and TNM stage (19) (Table II). Multivariate analysis demonstrated that high expression of CLDN3 independently showed a favourable prognosis ($P=0.028$), whereas high expression of FGF13 ($P=0.021$) and KIF11 ($P=0.015$) independently predicted an inferior overall survival (Table II).

The maximally selected rank statistics method was used to determine the optimal cut-off value for each gene. The samples were divided into high and low expression groups for each gene as follows: 64 high and 236 low CLDN3 expression samples, 265 high and 35 low FGF13 expression samples, and 214 high and 86 low KIF11 expression samples. The Kaplan-Meier plots demonstrated that high KIF11 ($P=0.023$) and FGF13 ($P=0.0068$) expression levels were significantly associated with a shorter overall survival time, whereas high CLDN3 ($P=0.039$) expression was significantly associated with a longer overall survival time (Fig. 4).

Association between KIF11 expression and clinicopathological characteristics of patients with SOC. Due to time constraints, only one of the core genes was validated in

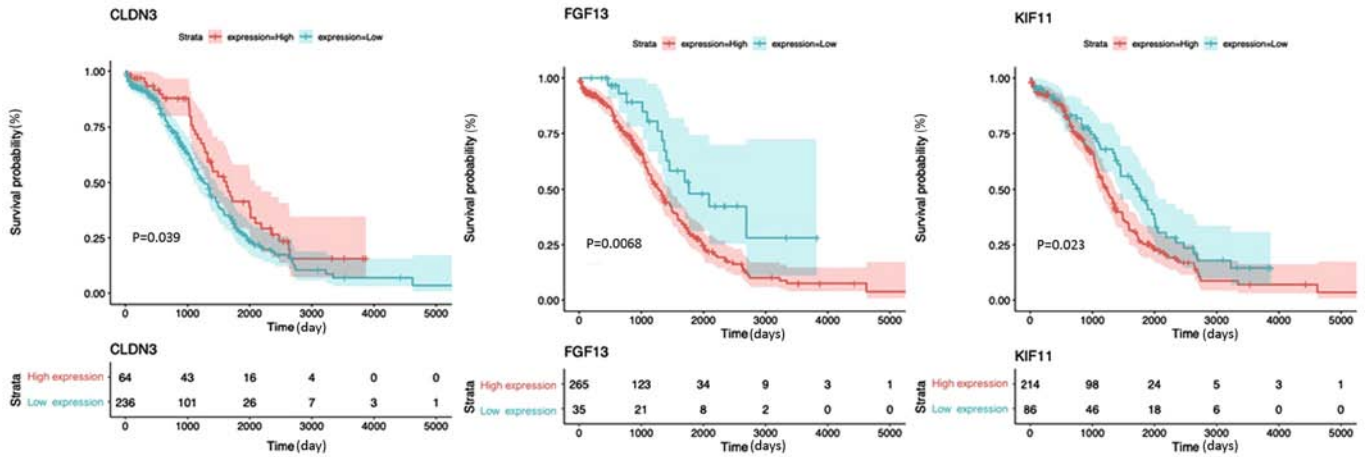


Figure 4. Overall survival analysis of the three core genes CLDN3, FGF13 and KIF11 in 300 serous ovarian cancer samples from The Cancer Genome Atlas database.

Table I. Forward and reverse primer sequences used for quantitative PCR.

Gene	Sequences (5'→3')
KIF11	F: GGTGGTGAGATGCAGGTAGG R: CAGAACGCGGAGAACCAAGA
GAPDH	F: CTGGGCTACACTGAGCACC R: AAGTGGTCGTTGAGGGCAATG

F, forward; R, reverse. KIF11, kinesin family member 11.

clinical samples. The representative gene was required to meet the following criteria: i) One of the core genes for SOC based on online datasets; ii) its expression was associated with prognosis; and iii) its high expression was positively associated with tumorigenesis and poor prognosis, which is essential to perform gene knockout experiments to further investigate functions *in vitro* and *in vivo*. KIF11 was the only core gene to meet these requirements and thus was selected for RT-qPCR analysis in 40 SOC and 10 adjacent normal tissue samples.

RT-qPCR analysis demonstrated that KIF11 expression was significantly upregulated in tumour tissues by ~3-fold compared with that in adjacent normal tissues (P<0.001) (Fig. 5A). Analysis of the GSE12470 dataset confirmed these results (P<0.01) (Fig. 5B). Patients were subsequently divided into two groups based on mean KIF11 expression, namely high (≥ mean expression) and low (< mean expression) expression groups. The association between KIF11 expression and clinicopathological characteristics of patients with SOC are presented in Table III. High KIF11 expression was significantly associated with a high tumour grade (P=0.042), TNM stage (P=0.024) and positive lymph node metastasis (P=0.026).

Discussion

The pathogenesis of SOC is a complex process that is driven by specific genetic and epigenetic alterations. For example,

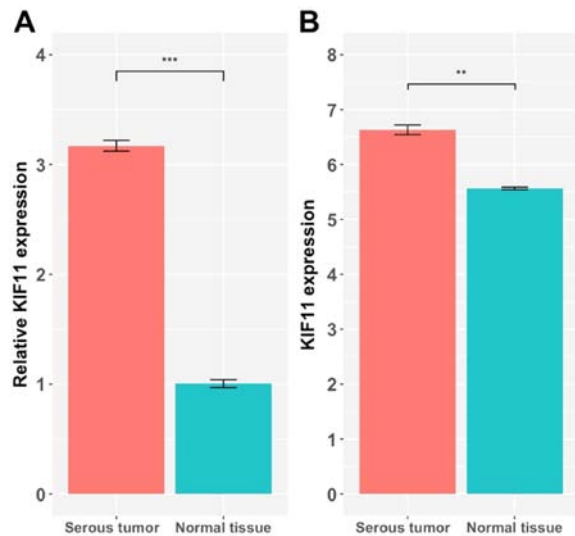


Figure 5. KIF expression between SOC and normal tissues. (A) Relative KIF11 expression between 40 SOC and 10 adjacent normal tissues from the clinical samples. (B) KIF11 expression between 43 SOC and 10 normal tissue samples from the GSE12470 dataset.

elevated ST6GAL1 expression and hypermethylation of RASSF1A have been demonstrated to play critical roles in ovarian cancer (20,21). The present study aimed to identify the core genes and significant gene sets for SOC to improve the understanding of the potential molecular mechanisms, which may help develop novel biomarkers or therapeutic targets for this disease. The present study also aimed to identify predictive genes for SOC prognosis.

Several core genes serve crucial roles in the tumorigenesis and progression of different types of cancer. For example, KIF11 is a member of the kinesin-like protein family and acts as a critical spindle motor protein during the establishment of a bipolar mitotic spindle (22). The inhibition of KIF11 blocks cell cycle progression, causes abnormal chromosome segregation and leads to cell death in head and neck squamous carcinoma (23). Similar results have been observed for glioblastoma, indicating that KIF11 is a potential anticancer target (24). Furthermore, KIF11 can promote the tumorigenesis

Table II. Univariate and multivariate analyses of genes with overall survival in TCGA SOC cohort.

Gene	Univariate analysis				Multivariate analysis ^b			
	HR	95% CI lower	95% CI upper	P-value	HR	95% CI lower	95% CI upper	P-value
BCHE	0.828	0.696	0.984	0.032 ^a	1.120	0.724	1.740	0.605
PRG4	0.911	0.847	0.979	0.011 ^a	1.140	0.734	1.770	0.563
SPOCK1	1.080	1.000	1.170	0.038 ^a	1.140	0.734	1.770	0.562
AQP9	1.080	1.010	1.160	0.036 ^a	1.120	0.722	1.740	0.613
FGF13	1.090	1.010	1.170	0.024 ^a	1.160	1.042	1.800	0.021 ^a
CALB2	1.070	1.010	1.140	0.033 ^a	1.120	0.724	1.740	0.606
CLDN3	0.883	0.806	0.968	0.008 ^a	0.920	0.729	0.980	0.028 ^a
KIF11	1.130	1.020	1.260	0.017 ^a	1.110	1.015	1.720	0.015 ^a
NAP1L2	1.210	1.070	1.360	0.002 ^a	1.130	0.729	1.750	0.584
GRHL2	1.070	1.010	1.130	0.029 ^a	1.130	0.727	1.750	0.590

We listed the 10 significant genes for univariate analysis. ^aP<0.05; ^bMultivariate analysis was performed with adjustment of age and TNM stage. HR, hazard ratio; CI, confidence interval; TNM, tumor-node-metastasis.

Table III. Association between KIF11 expression and clinicopathological characteristics of patients with serous ovarian cancer (n=40).

Characteristic	KIF11 expression		P-value
	High, n (%)	Low, n (%)	
Age, years			>0.999
≥60	11 (55)	10 (50)	
<60	9 (45)	10 (50)	
Tumour diameter, cm			>0.999
≥5	13 (65)	12 (60)	
<5	7 (35)	8 (40)	
Tumour grade			0.042 ^a
1	3 (15)	9 (45)	
2	6 (30)	7 (35)	
3	11 (55)	4 (20)	
TNM stage			0.026 ^a
I-II	5 (25)	13 (65)	
III-IV	15 (75)	7 (35)	
Serum CA-125 level, U/ml			0.751
≥35	10 (50)	8 (40)	
<35	10 (50)	12 (60)	
Lymph node metastasis			0.024 ^a
Positive	12 (60)	4 (20)	
Negative	8 (40)	16 (80)	

^aP<0.05. TNM, tumor-node-metastasis.

of oesophageal squamous cell carcinoma and colorectal cancer cell lines (25). Inhibiting KIF11 expression significantly suppresses the proliferation of oral cancer cells (26). CEP55 encodes a protein involved in microtubule bundling,

which is required for cytokinesis (27). CEP55 knockdown can suppress breast cancer proliferation (28). In addition, upregulation of CEP55 is positively associated with the aggressive features and unfavourable prognosis of ovarian cancer (29). TRIP13 can promote invasion in colon cancer (30), whereas RAD51API is upregulated in ovarian cancer and is associated with cell proliferation (31). Taken together, these core genes are considered promising suppression/driver genes for SOC and are worth further investigation.

The present study identified SOC-associated gene sets in addition to the core genes. KEGG-annotated cell cycle and DNA replication were significantly elevated in SOC. The gene sets annotated by GO analysis to be associated with processes involved in the cell cycle, such as mitotic sister chromatid segregation, were also upregulated in SOC. Abnormal cell cycle progression is a critical trigger of tumorigenesis, and the occurrence of SOC is associated with cell cycle alteration (32). In addition, the results of the present study demonstrated that cell cycle and microtubule-associated gene sets were enriched in the top-ranked module comprising all core genes. Microtubules are a crucial component of the cell cytoskeleton, which is involved in a number of functions and processes in cancer, such as the cell cycle (33). Thus, it was hypothesized that microtubules serve crucial roles in SOC, which may facilitate the progression of SOC with more specific mechanisms.

A total of three independent DEGs associated with overall survival, namely FGF13, CLND3 and KIF11, were identified in the present study. FGF13 encodes fibroblast growth factor (FGF)13, which is a member of the FGF family (34). The members of this family are involved in various mitogenic and cell survival activities, including morphogenesis, tumour growth and invasion (34). Elevated FGF13 expression mediates resistance to platinum-based drugs in cervical cancer cells (35). In addition, a significant association exists between high cytoplasmic FGF13 staining and biochemical recurrence risk, and the mRNA and protein expression levels of FGF13 in prostate cancer cell lines are higher compared with those

in prostate epithelial cell lines (36). FGF13 can promote the capability of lung cancer cells to resist the stressful effects of cancer-associated deregulation (37). CLND3 encodes Claudin 3, which is associated with the tight junction representing cell-to-cell adhesion (38). The association between CLND3 and cancer remains controversial; upregulation of CLND3 has been reported to act as a biomarker and a predictor of poor prognosis in different types of cancer, including triple-negative breast, oesophageal and gastric cancer (39,40). By contrast, downregulation of CLDN3 has been demonstrated to be associated with the progression and metastasis of lung squamous cell carcinoma (41). Another study has confirmed that the loss of CLND3 induces Wnt/ β -catenin activation, which is exacerbated through STAT-3 activation, leading to a poor prognosis in patients with colon cancer (42). Notably, KIF11 was identified as a core gene and an overall survival predictive gene in the present study. Consistent with the results of the present study, KIF11 is associated with poor overall survival in non-small cell lung cancer (43).

Previous studies have also used online transcriptomics data to investigate the molecular mechanisms of ovarian cancer. For example, Xu *et al* (44) used three online GEO datasets and identified key genes and pathways specific to ovarian cancer. Of note, KIF11 was also identified as one of the critical gene in the above study, which may be explained by the high proportion of serous subtypes among all ovarian cancer. In addition, their study failed to provide any *in vitro* or *in vivo* evidence. Similar to the present study, Guo *et al* (45) used transcriptomics, long non-coding RNA (lncRNA) and transcription factor datasets to develop a lncRNA-based risk score model to predict SOC prognosis. However, utilising multiple datasets may lead to misleading results when neglecting the batch effect (46). The batch effect is defined as a variability among different batches caused by several reasons, such as reagents and equipment made in batches, thus having systematic effects on the measurements; removal of the batch effect prior to merging different batches is essential (47). The present study used three different datasets; however, they were analysed separately. The dataset-dependent result was obtained, and then the overlapping data were identified. Thus, the batch effect should not have influenced the results of the present study.

Considering the promising bioinformatics results for KIF11 and the lack of previous studies confirming the association between KIF11 and SOC *in vivo* or *in vitro*, the role of KIF11 in SOC was further validated in clinical samples in the present study. KIF11 expression was significantly upregulated in tumour tissues compared with that in normal tissues. Furthermore, KIF11 expression was associated with a high histological grade, a high TNM stage and positive lymph node metastasis in univariate analysis, all of which are generally associated with unfavourable prognosis. GO annotation of gene sets demonstrated that KIF11 was involved in microtubule-associated pathways, such as microtubule motor activity. As microtubules serve a crucial role in SOC derived from the present study, it was speculated that elevated KIF11 may induce and promote SOC progression through its regulatory effect on microtubules. However, further studies are required to validate this assumption. Taken together, the results of the present study suggested that KIF11 may be a tumour-promoting gene in SOC.

To the best of our knowledge, the present study was the first to screen core genes for SOC via bioinformatics analysis and to investigate the association between KIF11 expression and the clinicopathological characteristics of patients with SOC. However, the present study was not without limitations. First, only two online databases with small sample sizes were used to investigate the core genes for SOC. Second, although clinical samples were assessed to determine the association between KIF11 expression and aggressive SOC behaviour, the present study lacked both *in vitro* and *in vivo* validation. Furthermore, follow-up data were not collected for the assessed samples, which may weaken the conclusions. Thus, prospective studies will aim to validate the role of KIF11 in SOC via *in vitro* and *in vivo* experiments. In addition, a study with a large sample size and complete follow-up data is required to further validate the results of the present study.

In conclusion, the present study identified potential core genes for SOC. Furthermore, KIF11 was validated as the representative core gene, and the results indicated that it may be a tumour-promoting gene. Thus, the results of the present study provided a potential biomarker and therapeutic target for SOC.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

The datasets used and/or analysed during the present study are available from the corresponding author upon reasonable request.

Authors' contributions

RZ and QZ designed the present study. JX and HC collected, analysed and interpreted the data. RZ and QZ drafted the initial manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The present study was approved by the Research Ethical Committee of The First Affiliated Hospital of Wenzhou Medical University (Wenzhou, China; approval no. 201902200012), and written informed consent was provided by all participants prior to the study. Patient information was anonymized and de-identified prior to analysis and all the patients agreed to publication.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

- Siegel R, Naishadham D and Jemal A: Cancer statistics, 2013. *CA Cancer J Clin* 63: 11-30, 2013.
- Wright JD, Shah M, Mathew L, Burke WM, Culhane J, Goldman N, Schiff PB and Herzog TJ: Fertility preservation in young women with epithelial ovarian cancer. *Cancer* 115: 4118-4126, 2009.
- Jacobs I and Bast RC Jr: The CA 125 tumour-associated antigen: A review of the literature. *Hum Reprod* 4: 1-12, 1989.
- Wei SU, Li H and Zhang B: The diagnostic value of serum HE4 and CA-125 and ROMA index in ovarian cancer. *Biomed Rep* 5: 41-44, 2016.
- Ramalingam P: Morphologic, immunophenotypic, and molecular features of epithelial ovarian cancer. *Oncology (Williston Park)* 30: 166-176, 2016.
- Gaitskell K, Green J, Pirie K, Barnes I, Hermon C, Reeves GK and Beral V; Million Women Study Collaborators: Histological subtypes of ovarian cancer associated with parity and breastfeeding in the prospective Million Women Study. *Int J Cancer* 142: 281-289, 2018.
- Bassis ML: An embryologically derived classification of ovarian tumors. *JAMA* 174: 1316-1319, 1960.
- Sun YV and Hu YJ: Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. *Adv Genet* 93: 147-190, 2016.
- Mok SC, Bonome T, Vathipadikeal V, Bell A, Johnson ME, Wong KK, Park DC, Hao K, Yip DK, Donniger H, *et al*: A gene signature predictive for outcome in advanced ovarian cancer identifies a survival factor: Microfibril-associated glycoprotein 2. *Cancer Cell* 16: 521-532, 2009.
- Bowen NJ, Walker LD, Matyunina LV, Logani S, Totten KA, Benigno BB and McDonald JF: Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells. *BMC Med Genomics* 2: 71, 2009.
- Yoshihara K, Tajima A, Komata D, Yamamoto T, Kodama S, Fujiwara H, Suzuki M, Onishi Y, Hatae M, Sueyoshi K, *et al*: Gene expression profiling of advanced-stage serous ovarian cancers distinguishes novel subclasses and implicates ZEB2 in tumor progression and prognosis. *Cancer Sci* 100: 1421-1428, 2009.
- Alhamdoosh M, Ng M, Wilson NJ, Sheridan JM, Huynh H, Wilson MJ and Ritchie ME: Combining multiple tools outperforms individual methods in gene set enrichment analyses. *Bioinformatics* 33: 414-424, 2017.
- Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, *et al*: TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 44: e71, 2016.
- Shimizu Y, Kamoi S, Amada S, Hasumi K, Akiyama F and Silverberg SG: Toward the development of a universal grading system for ovarian epithelial carcinoma. I. Prognostic significance of histopathologic features—problems involved in the architectural grading system. *Gynecol Oncol* 70: 2-12, 1998.
- Wright JD, Matsuo K, Huang Y, Tergas AI, Hou JY, Khoury-Collado F, St Clair CM, Ananth CV, Neugut AI and Hershman DL: Prognostic performance of the 2018 international federation of gynecology and obstetrics cervical cancer staging guidelines. *Obstet Gynecol* 134: 49-57, 2019.
- Mazzei M, Vascellari M, Zanardello C, Melchioni E, Vannini S, Forzan M, Marchetti V, Albanese F and Abramo F: Quantitative real time polymerase chain reaction (qRT-PCR) and RNAscope in situ hybridization (RNA-ISH) as effective tools to diagnose feline herpesvirus-1-associated dermatitis. *Vet Dermatol* 30: 491-e147, 2019.
- R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2014.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, *et al*: Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* 5: R80, 2004.
- Sobin LH and Wittekind C (eds): TNM Classification of Malignant Tumours. Union for International Cancer Control (UICC). John Wiley & Sons, Oxford, 2009.
- Dorsett KA, Jones RB, Ankenbauer KE, Hjelmeland AB and Bellis SL: Sox2 promotes expression of the ST6Gal-I glycosyltransferase in ovarian cancer cells. *J Ovarian Res* 12: 93, 2019.
- Ma L, Zhang JH, Liu FR and Zhang X: Hypermethylation of promoter region of RASSF1A gene in ovarian malignant epithelial tumors. *Zhonghua Zhong Liu Za Zhi* 27: 657-659, 2005 (In Chinese).
- Blangy A, Lane HA, d'Hérin P, Harper M, Kress M and Nigg EA: Phosphorylation by p34cdc2 regulates spindle association of human Eg5, a kinesin-related motor essential for bipolar spindle formation in vivo. *Cell* 83: 1159-1169, 1995.
- Martens-de Kemp SR, Nagel R, Stigter-van Walsum M, van der Meulen IH, van Beusechem VW, Braakhuis BJ and Brakenhoff RH: Functional genetic screens identify genes essential for tumor cell survival in head and neck and lung cancer. *Clin Cancer Res* 19: 1994-2003, 2013.
- Venero M, Horbinski C, Crish JF, Jin X, Vasanji A, Major J, Burrows AC, Chang C, Prokop J, Wu Q, *et al*: The mitotic kinesin KIF11 is a driver of invasion, proliferation, and self-renewal in glioblastoma. *Sci Transl Med* 7: 304ra143, 2015.
- Imai T, Oue N, Sentani K, Sakamoto N, Oraoka N, Egi H, Hinoi T, Ohdan H, Yoshida K and Yasui W: KIF11 is required for spheroid formation by oesophageal and colorectal cancer cells. *Anticancer Res* 37: 47-55, 2017.
- Daigo K, Takano A, Thang PM, Yoshitake Y, Shinohara M, Tohnai I, Murakami Y, Maegawa J and Daigo Y: Characterization of KIF11 as a novel prognostic biomarker and therapeutic target for oral cancer. *Int J Oncol* 52: 155-165, 2018.
- Zhao WM, Seki A and Fang G: Cep55, a microtubule-bundling protein, associates with centralspindlin to control the midbody integrity and cell abscission during cytokinesis. *Mol Biol Cell* 17: 3881-3896, 2006.
- Wang Y, Jin T, Dai X and Xu J: Lentivirus-mediated knockdown of CEP55 suppresses cell proliferation of breast cancer cells. *Biosci Trends* 10: 67-73, 2016.
- Zhang W, Niu C, He W, Hou T, Sun X, Xu L and Zhang Y: Upregulation of centrosomal protein 55 is associated with unfavorable prognosis and tumor invasion in epithelial ovarian carcinoma. *Tumour Biol* 37: 6239-6254, 2016.
- Kurita K, Maeda M, Mansour MA, Kokuryo T, Uehara K, Yokoyama Y, Nagino M, Hamaguchi M and Senga T: TRIP13 is expressed in colorectal cancer and promotes cancer cell invasion. *Oncol Lett* 12: 5240-5246, 2016.
- Chudasama D, Bo V, Hall M, Anikin V, Jeyaneethi J, Gregory J, Pados G, Tucker A, Harvey A, Pink R, *et al*: Identification of cancer biomarkers of prognostic value using specific gene regulatory networks (GRN): A novel role of RAD51AP1 for ovarian and lung cancers. *Carcinogenesis* 39: 407-417, 2018.
- Roy D, Sheng GY, Herve S, Carvalho E, Mahanty A, Yuan S and Sun L: Interplay between cancer cell cycle and metabolism: Challenges, targets and therapeutic opportunities. *Biomed Pharmacother* 89: 288-296, 2017.
- Parker AL, Kavallaris M and McCarroll JA: Microtubules and their role in cellular stress in cancer. *Front Oncol* 4: 153, 2014.
- Turner N and Grose R: Fibroblast growth factor signalling: From development to cancer. *Nat Rev Cancer* 10: 116-129, 2010.
- Okada T, Murata K, Hirose R, Matsuda C, Komatsu T, Ikekita M, Nakawatari M, Nakayama F, Wakatsuki M, Ohno T, *et al*: Upregulated expression of FGF13/FHF2 mediates resistance to platinum drugs in cervical cancer cells. *Sci Rep* 3: 2899, 2013.
- Yu L, Toriseva M, Tuomala M, Seikkula H, Elo T, Tuomela J, Kallajoki M, Mirtti T, Taimen P, Boström PJ, *et al*: Increased expression of fibroblast growth factor 13 in prostate cancer is associated with shortened time to biochemical recurrence after radical prostatectomy. *Int J Cancer* 139: 140-152, 2016.
- Bublik DR, Bursac S, Sheffer M, Oršolić I, Shalit T, Tarcic O, Kotler E, Mouhadeb O, Hoffman Y, Fuchs G, *et al*: Regulatory module involving FGF13, miR-504, and p53 regulates ribosomal biogenesis and supports cancer cell survival. *Proc Natl Acad Sci USA* 114: E496-E505, 2017.
- Shang X, Lin X, Alvarez E, Manorek G and Howell SB: Tight junction proteins claudin-3 and claudin-4 control tumor growth and metastases. *Neoplasia* 14: 974-985, 2012.
- Jääskeläinen A, Soini Y, Jukkola-Vuorinen A, Auvinen P, Haapasaari KM and Karihtala P: High-level cytoplasmic claudin 3 expression is an independent predictor of poor survival in triple-negative breast cancer. *BMC Cancer* 18: 223, 2018.
- Ferrer-Torres D, Nancarrow DJ, Quirk R, Thomas DG, Nadal E, Lin J, Chang AC, Reddy RM, Orringer MB, Taylor JM, *et al*: Genomic similarity between gastroesophageal junction and esophageal Barrett's adenocarcinomas. *Oncotarget* 7: 54867-54882, 2016.

41. Che J, Yue D, Zhang B, Zhang H, Huo Y, Gao L, Zhen H, Yang Y and Cao B: Claudin-3 inhibits lung squamous cell carcinoma cell epithelial-mesenchymal transition and invasion via suppression of the Wnt/ β -catenin signaling pathway. *Int J Med Sci* 15: 339-351, 2018.
42. Ahmad R, Kumar B, Chen Z, Chen X, Müller D, Lele SM, Washington MK, Batra SK, Dhawan P and Singh AB: Loss of claudin-3 expression induces IL6/gp130/Stat3 signaling to promote colon cancer malignancy by hyperactivating Wnt/ β -catenin signaling. *Oncogene* 36: 6592-6604, 2017.
43. Schneider MA, Christopoulos P, Muley T, Warth A, Klingmueller U, Thomas M, Herth FJ, Dienemann H, Mueller NS, Theis F, *et al*: AURKA, DLGAP5, TPX2, KIF11 and CKAP5: Five specific mitosis-associated genes correlate with poor prognosis for non-small cell lung cancer patients. *Int J Oncol* 50: 365-372, 2017.
44. Xu Z, Zhou Y, Cao Y, Dinh TLA, Wan J and Zhao M: Identification of candidate biomarkers and analysis of prognostic values in ovarian cancer by integrated bioinformatics analysis. *Med Oncol* 33: 130, 2016.
45. Guo Q, He Y, Sun L, Kong C, Cheng Y, Wang P and Zhang G: Identification of potential prognostic TF-associated lncRNAs for predicting survival in ovarian cancer. *J Cell Mol Med* 23: 1840-1851, 2019.
46. Reese SE, Archer KJ, Therneau TM, Atkinson EJ, Vachon CM, de Andrade M, Kocher JP and Eckel-Passow JE: A new statistic for identifying batch effects in high-throughput genomic data that uses guided principal component analysis. *Bioinformatics* 29: 2877-2883, 2013.
47. Danielsson F, James T, Gomez-Cabrero D and Huss M: Assessing the consistency of public human tissue RNA-seq data sets. *Brief Bioinform* 16: 941-949, 2015.