Evaluation of core serous epithelial ovarian cancer genes as potential prognostic markers and indicators of the underlying molecular mechanisms using an integrated bioinformatics analysis

YU-BO ZHANG1, YUHAN JIANG2, JIAO WANG1, JING MA1 and SHIYU HAN3

1Department of Gynecology and Obstetrics, The Fourth Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang 150001; 2Department of Gynecology, The Affiliated Hospital of Jining Medical College, Jining, Shandong 272000; 3Department of Gynecology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang 150001, P.R. China

Received February 23, 2019; Accepted August 23, 2019

DOI: 10.3892/ol.2019.10884

Abstract. Ovarian cancer is a major cause of mortality in women. However, the molecular events underlying the pathogenesis of the disease are yet to be fully elucidated. In the present study, an integrated bioinformatics analysis was performed to identify core genes involved in serous epithelial ovarian cancer. A total of three expression datasets were downloaded from the Gene Expression Omnibus database, and included 46 serous epithelial ovarian cancer and 30 ovarian surface epithelium samples. The three datasets were merged, and batch normalization was performed. The normalized merged data were subsequently analyzed for differentially expressed genes (DEGs). In total, 2,212 DEGs were identified, including 1,300 upregulated and 912 downregulated genes. Gene Ontology analysis revealed that these DEGs were primarily involved in ‘regulation of cell cycle’, ‘mitosis’, ‘DNA packaging’ and ‘nucleosome assembly’. The main cellular components included ‘extracellular region part’, ‘chromosome’, ‘extracellular matrix’ and ‘condensed chromosome kinetochore’, whereas the molecular functions included ‘Calcium ion binding’, ‘polysaccharide binding’, ‘enzyme inhibitor activity’, ‘growth factor activity’, ‘cyclin-dependent protein kinase regulator activity’, ‘microtubule motor activity’ and ‘Wnt receptor activity’. Kyoto Encyclopedia of Genes and Genomes pathway analysis revealed that these DEGs were predominantly involved in ‘Wnt signaling pathway’, ‘pathways in cancer’, ‘PI3K-Akt signaling pathway’, ‘cell cycle’, ‘ECM-receptor interaction’, ‘p53 signaling pathway’ and ‘focal adhesion’. The 20 most significant DEGs were identified from the protein-protein interaction network, and Oncomine analysis of these core genes revealed that 13 were upregulated and two were downregulated in serous epithelial ovarian cancer. Survival analysis revealed that cyclin B1, polo like kinase 1, G protein subunit γ transducin 1 and G protein subunit γ 12 are key molecules that may be involved in the prognosis of serous epithelial ovarian cancer. These core genes may provide novel treatment targets, although their roles in the carcinogenesis and prognosis of serous epithelial ovarian cancer require further study.

Introduction

The mortality rate of ovarian cancer ranks as the highest among gynecological tumors in the western world, and its incidence is increasing on a yearly basis (1). This is due to a lack of specific symptoms, which impedes its early diagnosis and results in high recurrence rates following radical surgery and chemotherapy (1). Although treatment outcomes have greatly improved, the 5-year survival rate of patients with ovarian cancer remains low, at 46.5% in 2017 (2), whereas the survival rate of patients with distant metastases is worse (29%). Out of all cases, ~70% are diagnosed at an advanced stage, and have poor prognosis (3). The 5-year survival rate for patients with advanced ovarian cancer is only ~20%; however, if diagnosed early, this can increase to 85-90% (4). Among the different pathological types, serous epithelial ovarian cancer is the most common (5). Therefore, an early diagnosis of serous epithelial ovarian cancer may greatly improve prognosis.

At present, the standard method for the early diagnosis and monitoring of ovarian cancer is ultrasound examination combined with serum tumor marker detection (6). However, the specificity of this diagnostic method is low, and the 5-year survival rate after diagnosis using this approach is only 30% (7). The occurrence and development of tumors are associated with accumulated molecular genetic or genomic alterations (8). For instance, high-grade serous ovarian cancer cases frequently exhibit tumor protein p53 mutations and alterations in BRCA1/2 DNA repair associated and related homologous recombination genes, either by mutation, promoter

Correspondence to: Dr Shiyu Han, Department of Gynecology, The Fourth Affiliated Hospital of Harbin Medical University, 37 Yiyuan Road, Harbin, Heilongjiang 150001, P.R. China

E-mail: hsyhrbmu@gmail.com

Key words: ovarian cancer, Gene Expression Omnibus, multiple chips, bioinformatics analysis, differentially expressed genes
methylations or loss of heterozygosity (9). Therefore, it is important to investigate the molecular mechanisms underpinning the malignant behavior of serous epithelial ovarian cancer cells to develop more effective methods for early diagnosis, and to identify more reliable molecular markers that may be used either as novel therapeutic targets or to assess prognosis. Gene expression microarray analysis is an efficient and large-scale technique for obtaining genetic data (10). It has been widely used to explore gene expression profiles in numerous types of human cancer (11). Microarray data have become increasingly available in the public domain over the last few years, in platforms such as the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database. The large volume of data that has been published in these public databases, and the integration of multiple databases, allow for an exhaustive study of underlying molecular mechanisms. The integration and analysis of microarray data from several gene expression profiles may enable investigators to obtain more reliable molecular markers. However, since these data originate from different microarray products from a wide range of experiments using different reagents, which have also been performed by operators of varying proficiencies, a large degree of variability exists among datasets (12). This problem may be solved using batch normalization programs available in R software.

In the present study, we employed an integrated bioinformatics approach to identify potential molecular markers for the early detection and prognosis of serous epithelial ovarian cancer. Furthermore, the markers obtained may be targets for the development of novel therapies for serous epithelial ovarian cancer.

Materials and methods

Microarray data. The GEO database (www.ncbi.nlm.nih.gov/geo) is an international public repository that archives and distributes high-throughput gene expression data and other functional genomics datasets (13). The keywords ‘ovarian cancer gene expression’ were used to search the GEO database, and the CEL files of GSE14407 (14), GSE54388 (15) and GSE38666 (16) datasets were downloaded for subsequent analysis. The quality of the gene chips was detected by RNA degradation mapping (17). Only gene chips with a proper degradation slope in RNA degradation mapping were included in the subsequent analysis.

Data pre-treatment and identification of differentially expressed genes (DEGs). All data were processed using R software (www.r-project.org). The Affy package (version 3.9; www.bioconductor.org/packages/release/bioc/html/affy.html) was used to extract expression data from CEL files, and the Robust Multi-Array Average method in R was used to perform quartile data normalization of the three expression datasets (18). Following normalization, data from the three microarray datasets were merged to form a new gene expression profile. The sva package (version 3.9; bioconductor.org/packages/release/bioc/html/sva.html) in R was used to identify, estimate and remove unwanted sources of variation in high-throughput experiments to eliminate the batch effect (19). The DEGs between serous epithelial ovarian cancer and normal ovarian surface epithelial tissue from the three microarray datasets were analyzed using the Limma package (version 3.9; http://www.bioconductor.org/packages/release/bioc/html/limma.html). Values of log fold change (FC)>1.0 and adjusted P<0.05 were selected as the cut-off criteria for DEG selection.

GO and KEGG pathway enrichment analyses of DEGs. The DAVID database is an important online tool for gene function analysis. Gene Ontology (GO) analysis of the DEGs was performed using DAVID 6.7 (david.ncifcrf.gov). Kyoto Encyclopedia of Genes and Genomes (KEGG) is an online encyclopedia that assigns functions to genes and genomes at molecular and higher levels (20). KEGG pathway analysis of DEGs was performed using KEGG Orthology-Based Annotation System (KOBAS) 3.0 (kobas.cbi.pku.edu.cn), an online analysis tool. GO functional enrichment was assessed using the criteria of P<0.05 and false discovery rate (FDR) <0.05. P<0.05 was used to identify statistically KEGG pathways. Subsequently, the GOplot package (version 1.0.2; wencke.github.io) was used to construct the Chord diagram, and the clusterProfiler package (version 3.9; www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html) was used to create the bar plot.

Protein-protein interaction (PPI) networks. PPI networks may be used to understand normal cell function and to study disease pathogenesis (21). In the present study, the STRING database (string-db.org) was used to explore the PPIs of the DEGs, with a cut-off criterion set at an interaction score >0.99. PPI networks were constructed using Cytoscape software (version 3.6.1; https://cytoscape.org), which is a bioinformatics program for the visualization of molecular interaction networks. Each node in the PPI network represents a gene, protein or other molecule, and the connections between the nodes represent the interactions between these biomolecules. The most closely associated nodes may indicate core proteins or key genes with important physiological regulatory functions (22). Therefore, the interactions and pathway associations among proteins encoded by the DEGs in serous epithelial ovarian cancer were assessed in this manner.

Oncomine analysis of hub genes. Oncomine (www.oncomine.org) is a bioinformatics program designed to collect, standardize and analyze cancer transcriptome data. It integrates RNA- and DNA-sequencing data from various sources, including GEO, The Cancer Genome Atlas (TCGA) (https://cancergenome.nih.gov) and published literature (23). A meta-analysis of the selected hub genes in ovarian cancer compared with normal ovarian tissue was performed using Oncomine to compare these genes expression across different studies.

Kaplan-Meier (KM) survival analysis. The KM estimate is a nonparametric statistic used to measure the percentage of patients living for a certain period of time following a specific treatment. The hub genes were analyzed using an online tool, KM Plotter (updated on 2/20/2019; kmplot.com/analysis), which was used to assess overall and progression-free survival of patients with serous epithelial ovarian cancer by the log-rank test. This tool was constructed using the gene expression and
survival data of 1,232 patients with serous epithelial ovarian cancer, which were downloaded from the GEO and TCGA databases (24).

**Results**

**Details of the datasets.** CEL files of the datasets GSE14407, GSE54388 and GSE38666 were downloaded from the GEO database. The platform used to generate data for all three datasets was the Human Genome U133 Plus 2.0 array (GPL570; HG-U133_Plus_2; Affymetrix; Thermo Fisher Scientific, Inc.). These datasets are stored in a public repository (doi.org/10.6084/m9.figshare.8148608.v2) and are easily obtained. The GSE14407 dataset included data from 12 healthy ovarian surface epithelial samples and 12 laser-capture microdissected serous ovarian cancer epithelial samples. The GSE54388 dataset included data from 6 human ovarian surface epithelial samples and 16 serous ovarian cancer epithelial samples. The GSE38666 dataset included data from 12 normal ovarian surface epithelial samples and 18 serous cancer epithelial samples. The characteristics of these datasets are shown in Table I. The data from the three microarray datasets were merged to form a novel gene expression profile, and the detailed results are available in a public repository (doi.org/10.6084/m9.figshare.8148617.v1). Gene chips of these datasets are stored in a public repository (doi.org/10.6084/m9.figshare.8148608.v2) and are easily obtained. The GSE14407 dataset included data from 12 healthy ovarian surface epithelial samples and 12 laser-capture microdissected serous ovarian cancer epithelial samples. The GSE54388 dataset included data from 6 human ovarian surface epithelial samples and 16 serous ovarian cancer epithelial samples. The GSE38666 dataset included data from 12 normal ovarian surface epithelial samples and 18 serous cancer epithelial samples. The characteristics of these datasets are shown in Table I. The data from the three microarray datasets were merged to form a novel gene expression profile, and the detailed results are available in a public repository (doi.org/10.6084/m9.figshare.8148617.v1). Gene chips of good quality from each dataset were selected for subsequent analysis. Additionally, the RNA degradation maps for the three datasets are shown in Fig. S1.

**Identification of DEGs.** The novel gene expression profile created by merging the three original microarray datasets was subsequently analyzed using the Limma package. According to the criteria of |log FC|>1.0 and adjusted P<0.05, 2,212 DEGs were identified, comprising 1,300 upregulated and 912 downregulated genes. The detailed results are shown in Table SII. The top 20 genes with the most connections, which represent the most important DEGs, are presented in Fig. 5B. Among the 20 closely associated genes, CDK1 exhibited the highest node degree of 106.

**Oncomine analysis of hub genes.** An Oncomine database analysis of cancer tissue compared with normal tissue was performed for the 20 core genes identified for serous epithelial ovarian cancer. These meta-analysis results revealed that cell division control protein 1 (CDC1), cyclin B1 (CCNB1), polo like kinase 1 (PLKI), cell division cycle 20 (CDC20), cyclin B2 (CCNB2), mitotic arrest deficient 2 like 1 (MAD2LI), cyclin A2 (CCNA2), histone cluster 1 H2B family member d (HIST1H2BD), centromere protein E (CENPE), BUB1 mitotic checkpoint serine/threonine kinase B (BUB1B), histone cluster 1 H2B family member h (HIST1H2BH), kinesin family member 2C (KIF2C) and aurora kinase A (AURKA) were upregulated, whereas G protein subunit γ 12 (NGN12) and G protein subunit γ 11 (NGN11) were downregulated, among the different datasets. The results of this analysis are shown in Fig. 6.

**KEGG pathway analysis.** The most significantly enriched pathways of the DEGs were identified using the KOBAS database. The results of this analysis are shown in Table V and Fig. 4. The signaling pathways of DEGs were predominantly enriched in 'Wnt signaling pathway', 'viral carcinogenesis', 'pathways in cancer', 'PI3K-Akt signaling pathway', 'cell cycle', 'extracellular matrix (ECM)-receptor interaction', 'p53 signaling pathway' and 'focal adhesion'.

**PPI network construction.** All DEGs were screened using the STRING database to further investigate their properties and the interactions among them. The PPI network of DEGs, with a criterion of interaction score >0.99, was built using Cytoscape software, and the results are shown in Fig. 5A. To identify core genes, the number of connections were counted for each gene. The detailed results are shown in Table III. The top 20 genes with the most connections, which represent the most important DEGs, are presented in Fig. 5B. Among the 20 closely associated genes, CDK1 exhibited the highest node degree of 106.

**KM survival analysis.** Survival analysis of the 20 core genes for serous epithelial ovarian cancer was performed by constructing a KM curve using the KM Plotter package. This analysis revealed that high expression levels of CCNB1, GNG12 and G protein subunit g transducin 1 (GNFT1), and low expression levels of PLKI were associated with poor overall and progression-free survival in patients with serous ovarian cancer.

---

**Table I. Characteristics of the three datasets.**

<table>
<thead>
<tr>
<th>GSE accession number</th>
<th>GPL</th>
<th>Organism</th>
<th>Control samples, n</th>
<th>Cancer samples, n</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE14407</td>
<td>GPL570</td>
<td>Homo sapiens</td>
<td>12</td>
<td>12</td>
<td>USA</td>
</tr>
<tr>
<td>GSE54388</td>
<td>GPL570</td>
<td>Homo sapiens</td>
<td>6</td>
<td>16</td>
<td>USA</td>
</tr>
<tr>
<td>GSE38666</td>
<td>GPL570</td>
<td>Homo sapiens</td>
<td>12</td>
<td>18</td>
<td>USA</td>
</tr>
</tbody>
</table>

GPL, Gene Expression Omnibus platform.
cancer. In addition, high expression levels of AURKA, BUB1B, CDC20, CENPE and GNG11, and low expression levels of HIST1H2BN, were associated with poor overall survival, whereas high expression levels of adenylate cyclase 4 (ADCY4) and protein phosphatase 2 catalytic subunit α (PPP2CA) were associated with poor progression-free survival. The results for overall and progression-free survival analysis are shown in Fig. 7A and B, respectively.

Discussion

Ovarian cancer is the most prevalent gynecological cancer, and 75% of patients are diagnosed with advanced disease, of which only 20% survive for 5 years after diagnosis (25). The majority of patients with ovarian cancer are initially responsive to conventional chemotherapy, and enter clinical remission following initial treatment (26). However, tumor metastasis and recurrence occur in >70% of patients with ovarian cancer, despite treatment, and lead to mortality (27). Among the various types of ovarian cancer, serous epithelial ovarian cancer is the most common pathological type (5). Therefore, exploring the molecular mechanisms of serous epithelial ovarian cancer development is important to identify novel molecular markers and therapeutic targets. Identifying effective methods for preventing the progression of ovarian cancer is particularly important for improving the overall and progression-free survival of patients with serous epithelial ovarian cancer.

Previous research has suggested that molecular biomarkers may be used for the accurate diagnosis of cancer (28). These molecular markers may be more sensitive and specific than traditional screening methods, and they are easier to use (24). Microarray and high-throughput sequencing technologies, capable of detecting the expression levels of tens of millions of human genes, have been widely used to identify molecular biomarkers and potential targets for the diagnosis and treatment of cancer (29). Thus far, numerous basic research papers on the mechanisms of ovarian cancer have been published, but the 5-year survival rate of patients with ovarian cancer remains relatively low. Furthermore, no biomarkers for predicting the prognosis or monitoring the effectiveness of treatments have been identified, since the majority of studies have focused on simple genetic events or the results of a single experimental study (30). In the present study, three gene expression datasets from different experiments were combined and batch-corrected using the sva package. They were subsequently analyzed using R software and other bioinformatics tools. A total of 2,212 DEGs were identified in the present study using the Limma package. This included 1,300 upregulated and 912 downregulated genes. These were further divided into three groups through GO functional annotation, including molecular functions, biological processes and cellular components. The molecular functions included ‘Calcium ion binding’, ‘polysaccharide binding’, ‘enzyme inhibitor activity’, ‘growth factor activity’, ‘cyclin-dependent protein kinase regulator activity’, ‘microtubule motor activity’, ‘Wnt receptor activity’ and ‘protein kinase regulator activity’. The biological processes included ‘regulation of cell cycle’, ‘mitosis’, ‘DNA packaging’, ‘DNA replication’, and ‘Chromosome segregation’, whereas the cellular components
included ‘extracellular region part’, ‘chromosome’, ‘extracellular matrix’, ‘microtubule cytoskeleton’, ‘nucleosome’, ‘spindle’ and ‘condensed chromosome kinetochore’. The majority of these enrichment functions are associated with tumorigenesis and development. For instance, growth factor activity in various types of cancer is able to regulate cell

Figure 2. GO enrichment analysis of DEGs in ovarian cancer. (A) GO enrichment with FDR <0.05 in ovarian cancer and (B) GO analysis with a P-value <0.05. DEGs were divided into three functional groups: Biological processes, cellular components and molecular functions. GO, Gene Ontology; DEG, differentially expressed gene; FDR, false discovery rate.
proliferation, differentiation and apoptosis, thus affecting the ability of cells to self-renew, migrate, senesce or undergo apoptosis (31). Cyclin-dependent kinases (CDKs/cyclins) form a family of heterodimeric kinases that serve important roles in regulating cell cycle progression, transcription and other major biological processes (32). Alterations in CDK activity affect the proliferation of cancer cells, and abnormal activities of these proteins have been reported in various types of human cancer, such as pancreatic cancer (32,33). Wnt signaling regulates an evolutionarily conserved pathway that serves an important role in numerous cellular activities, including cell proliferation, calcium homeostasis and cellular polarity (34). Wnt receptor activity is upregulated in a variety of cancer types, such as colorectal and gastric cancer (34-36). Microtubules are dynamic structures that are involved in cell movement, intracellular trafficking and mitosis (37).

Table II. Enrichment of biological processes.

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
<th>Gene count, n</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0000280</td>
<td>Nuclear division</td>
<td>63</td>
<td>1.52x10^-16</td>
</tr>
<tr>
<td>GO:0007067</td>
<td>Mitosis</td>
<td>63</td>
<td>1.52x10^-16</td>
</tr>
<tr>
<td>GO:0048285</td>
<td>Organelle fission</td>
<td>63</td>
<td>1.65x10^-15</td>
</tr>
<tr>
<td>GO:0051301</td>
<td>Cell division</td>
<td>71</td>
<td>3.59x10^-14</td>
</tr>
<tr>
<td>GO:0007059</td>
<td>Chromosome segregation</td>
<td>27</td>
<td>5.61x10^-9</td>
</tr>
<tr>
<td>GO:0051726</td>
<td>Regulation of cell cycle</td>
<td>62</td>
<td>6.58x10^-8</td>
</tr>
<tr>
<td>GO:0006323</td>
<td>DNA packaging</td>
<td>31</td>
<td>1.24x10^-7</td>
</tr>
<tr>
<td>GO:0007051</td>
<td>Spindle organization</td>
<td>18</td>
<td>1.71x10^-7</td>
</tr>
<tr>
<td>GO:0065004</td>
<td>Protein-DNA complex assembly</td>
<td>24</td>
<td>4.68x10^-6</td>
</tr>
<tr>
<td>GO:0007017</td>
<td>Microtubule-based process</td>
<td>44</td>
<td>4.44x10^-5</td>
</tr>
<tr>
<td>GO:0006260</td>
<td>DNA replication</td>
<td>35</td>
<td>9.68x10^-3</td>
</tr>
<tr>
<td>GO:0007155</td>
<td>Cell adhesion</td>
<td>89</td>
<td>0.001049</td>
</tr>
<tr>
<td>GO:0042127</td>
<td>Regulation of cell proliferation</td>
<td>98</td>
<td>0.001138</td>
</tr>
<tr>
<td>GO:0016477</td>
<td>Cell migration</td>
<td>42</td>
<td>0.001186</td>
</tr>
<tr>
<td>GO:0006259</td>
<td>DNA metabolic process</td>
<td>67</td>
<td>0.001746</td>
</tr>
<tr>
<td>GO:0001525</td>
<td>Angiogenesis</td>
<td>26</td>
<td>0.001804</td>
</tr>
<tr>
<td>GO:0008283</td>
<td>Cell proliferation</td>
<td>59</td>
<td>0.002059</td>
</tr>
</tbody>
</table>

GO, Gene Ontology.

Table III. Enrichment of cellular components.

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
<th>Gene count, n</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0044421</td>
<td>Extracellular region part</td>
<td>144</td>
<td>3.98x10^-10</td>
</tr>
<tr>
<td>GO:0000793</td>
<td>Condensed chromosome</td>
<td>34</td>
<td>2.21x10^-8</td>
</tr>
<tr>
<td>GO:0005819</td>
<td>Spindle</td>
<td>36</td>
<td>5.96x10^-8</td>
</tr>
<tr>
<td>GO:0031012</td>
<td>Extracellular matrix</td>
<td>63</td>
<td>7.92x10^-8</td>
</tr>
<tr>
<td>GO:0005578</td>
<td>Proteinaceous extracellular matrix</td>
<td>59</td>
<td>1.55x10^-7</td>
</tr>
<tr>
<td>GO:0000777</td>
<td>Condensed chromosome kinetochore</td>
<td>20</td>
<td>3.54x10^-7</td>
</tr>
<tr>
<td>GO:0005694</td>
<td>Chromosome</td>
<td>75</td>
<td>4.68x10^-7</td>
</tr>
<tr>
<td>GO:0000779</td>
<td>Condensed chromosome</td>
<td>21</td>
<td>7.25x10^-7</td>
</tr>
<tr>
<td>GO:0015630</td>
<td>Microtubule cytoskeleton</td>
<td>83</td>
<td>2.51x10^-6</td>
</tr>
<tr>
<td>GO:0005876</td>
<td>Spindle microtubule</td>
<td>11</td>
<td>1.34x10^-4</td>
</tr>
<tr>
<td>GO:0005615</td>
<td>Extracellular space</td>
<td>90</td>
<td>2.11x10^-4</td>
</tr>
<tr>
<td>GO:0005815</td>
<td>Microtubule organizing center</td>
<td>41</td>
<td>2.91x10^-4</td>
</tr>
<tr>
<td>GO:0031262</td>
<td>Ndc80 complex</td>
<td>4</td>
<td>0.002697</td>
</tr>
<tr>
<td>GO:0000786</td>
<td>Nucleosome</td>
<td>14</td>
<td>0.003458</td>
</tr>
<tr>
<td>GO:0032993</td>
<td>Protein-DNA complex</td>
<td>15</td>
<td>0.020501</td>
</tr>
</tbody>
</table>

GO, Gene Ontology; Ndc80, NDC80 kinetochore complex component.
Table IV. Enrichment of molecular function.

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
<th>Gene count, n</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0004857</td>
<td>Enzyme inhibitor activity</td>
<td>43</td>
<td>1.20x10⁻⁴</td>
</tr>
<tr>
<td>GO:0001871</td>
<td>Pattern binding</td>
<td>28</td>
<td>2.80x10⁻⁴</td>
</tr>
<tr>
<td>GO:0030247</td>
<td>Polysaccharide binding</td>
<td>28</td>
<td>2.80x10⁻⁴</td>
</tr>
<tr>
<td>GO:0030246</td>
<td>Carbohydrate binding</td>
<td>51</td>
<td>3.27x10⁻⁴</td>
</tr>
<tr>
<td>GO:0005509</td>
<td>Calcium ion binding</td>
<td>109</td>
<td>4.52x10⁻⁴</td>
</tr>
<tr>
<td>GO:0005539</td>
<td>Glycosaminoglycan binding</td>
<td>25</td>
<td>8.32x10⁻⁴</td>
</tr>
<tr>
<td>GO:0008083</td>
<td>Growth factor activity</td>
<td>24</td>
<td>0.01087</td>
</tr>
<tr>
<td>GO:0016538</td>
<td>Cyclin-dependent protein kinase regulator activity</td>
<td>6</td>
<td>0.01207</td>
</tr>
<tr>
<td>GO:0003777</td>
<td>Microtubule motor activity</td>
<td>14</td>
<td>0.01373</td>
</tr>
<tr>
<td>GO:0005201</td>
<td>Extracellular matrix structural constituent</td>
<td>15</td>
<td>0.01470</td>
</tr>
<tr>
<td>GO:0004859</td>
<td>Phospholipase inhibitor activity</td>
<td>5</td>
<td>0.01543</td>
</tr>
<tr>
<td>GO:0051287</td>
<td>NAD or NADH binding</td>
<td>10</td>
<td>0.01735</td>
</tr>
<tr>
<td>GO:0046915</td>
<td>Transition metal ion transmembrane transporter activity</td>
<td>7</td>
<td>0.02473</td>
</tr>
<tr>
<td>GO:0042813</td>
<td>Wnt receptor activity</td>
<td>4</td>
<td>0.03605</td>
</tr>
<tr>
<td>GO:0019887</td>
<td>Protein kinase regulator activity</td>
<td>13</td>
<td>0.04381</td>
</tr>
</tbody>
</table>

GO, Gene Ontology.

Figure 3. Distribution of a number of differentially expressed genes in ovarian cancer for cancer-associated GO enrichment. GO, Gene Ontology; FC, fold change.
Alterations in microtubule activity have been reported in a range of cancer types, such as breast and non-small cell lung cancer (37). These alterations have been associated with poor prognosis and chemotherapy resistance in solid and
Figure 5. PPI network of DEGs. (A) PPI network of DEGs with the criterion of interaction score >0.99. (B) Top 20 genes with the most connection nodes. The x-axis represents the number of connections. PPI, protein-protein interaction; DEG, differentially expressed gene.
hematological types of cancer (37). Nucleosome assembly following DNA replication, DNA repair and gene transcription is critical for the maintenance of genome stability and epigenetic information (38). Alterations or mutations that affect nucleosome assembly have also been implicated in certain types of cancer, such as cervical cancer (38,39).

In addition, the enriched KEGG pathways of DEGs identified in the present study included the ‘cell cycle’, ‘pathways in cancer’, ‘PI3K-Akt signaling pathway’, ‘Wnt signaling pathway’, ‘ECM-receptor interaction’, ‘mTOR signaling pathway’ and ‘focal adhesion’. The significance of the PI3K-Akt signaling pathway in ovarian cancer has been

Figure 6. Oncomine analysis of core serous epithelial ovarian cancer genes in cancer vs. normal tissue across multiple datasets.
Figure 7. Kaplan-Meier survival curve analysis of the core genes. (A) Overall survival and (B) progression-free survival analyses are shown. HR, hazard ratio.
reported previously (40). In a copy number analysis on 93 primary ovarian tumors using array comparative genomic hybridization, Huang et al (40) identified that the PK3-Akt signaling pathway was the most frequently altered cancer-associated signaling pathway. The Wnt/β-catenin signaling pathway regulates a variety of fundamental cellular functions, including proliferation, polarity, adhesion and motility during development, differentiation and adult tissue homeostasis (41). Furthermore, the Wnt/β-catenin signaling pathway has been demonstrated to be essential for the growth and progression of ovarian cancer (42). Bodnar et al (43) demonstrated that activation of the Wnt/β-catenin signaling pathway may facilitate the proliferation and differentiation of ovarian cancer cells, inhibit apoptosis and promote ovarian cancer growth (43). The mTOR signaling pathway regulates several major physiological processes, including protein synthesis, macromolecular biosynthesis, cytoskeleton remodeling, angiogenesis, survival, metabolism, autophagy and response to stress (44). Due to its pivotal role in cell growth and differentiation, its dysregulation is associated with pathological conditions, including tumor transformation and progression in breast, gastrointestinal, liver and prostate cancer (45). The detection of components of these signaling pathways, and their expression levels, may help predict the occurrence and development of serous epithelial ovarian cancer, and provide potential therapeutic targets.

In the present study, 20 closely associated genes were identified by constructing a PPI network of proteins encoded by DEGs. Oncomine analysis further revealed that the following 15 genes were core serous epithelial ovarian cancer-associated genes among the different datasets: CDC1, CCNB1, PLK1, CDC20, CCNB2, MAD2L1, CCNA2, HIST1H2BD, CENPE, BUB1B, HIST1H2BH, KIF2C, AURKA, GNG12 and GNG11. Among the 20 closely associated genes, CDK1 exhibited the highest node degree of 106. CDK1 is an important cell cycle-regulating protein that serves key roles in the cell cycle G2/M-phase regulation network (46). Upregulated protein expression levels of CDK1 have been detected in numerous human malignant tumor tissues, and have been found to be closely associated with the malignant prognosis (47,48). Yang et al (49) demonstrated that high expression levels of cytoplasmic CDK1 promote the growth of epithelial ovarian cancer cells, indicating a poor overall survival rate (49). Therefore, CDK1 is expected to be an effective therapeutic target for epithelial ovarian cancer by disrupting the ovarian cancer cell cycle. Survival analysis identified CCNB1, PLK1, GNG12 and GNG11 as being associated with the overall and progression-free survival of patients with serous epithelial ovarian cancer. In addition, high expression levels of AURKA, BUB1B, CDC20, CENPE and GNG11, and low expression levels of HIST1H2BN, were associated with poor overall survival of serous epithelial ovarian cancer, whereas high expression levels of ADCY4 and PPP2CA were associated with poor progression-free survival.

CCNB1 is a mitotic cyclin, due to its crucial role in modulating G2/M-phase progression in the cell cycle (50). It has been demonstrated to be involved in cell growth, differentiation, apoptosis and metastasis in numerous types of cancer such as lung cancer (51-53). Previous studies have indicated that CCNB1 is associated with malignancy, and upregulation of CCNB1 has been identified as a marker of poor prognosis in patients with non-small cell lung cancer, head and neck squamous cell carcinoma, breast cancer and hepatocellular carcinoma (54-57). Therefore, CCNB1 has the potential to also be a molecular marker of ovarian cancer prognosis. PLK1 is a member of the polo subfamily of serine/threonine protein kinases (collectively referred to as PLKs), which serve key roles in a variety of cellular processes, including cell cycle progression, differentiation and survival (58). Overexpression of PLK1 in breast cancer cells is able to initiate transcriptional programs required for mitosis by phosphorylating the transcription factor forkerhead box M1, overriding the DNA damage checkpoint, contributing to the induction of invasiveness by phosphorylating vimentin and impairing mitotic integrity, which lead to aneuploidy and are associated with tumor formation (59). PLK1 is upregulated in various types of human cancer, including glioma, thyroid cancer, head and neck squamous cell carcinoma, melanoma, and colorectal, esophageal, ovarian, breast and prostate cancer (60). Weichert et al (61) reported that PLK1 is frequently upregulated in malignant epithelial ovarian tumors, and that this upregulation is associated with mitosis and poor prognosis in patients (61). However, a recently published study revealed that overexpression of PLK1 could act as a tumor suppressor by disrupting mitotic progression and cytokinesis in vitro and in vivo, and an increase in PLK1 levels in patients with breast cancer was associated with an improved prognosis (62). In the present study, high expression levels of PLK1 were associated with an improvement in overall and progression-free survival of patients with serous epithelial ovarian cancer. However, further research is required to explore the association between PLK1 and survival in such patients.

GNG12 is a member of the G-protein family, corresponding to the G-protein γ12 subunit (63). Larson et al (64) revealed that GNG12 is a negative regulator of the response to lipopolysaccharide, and may be a critical factor in the overall inflammatory signaling cascade (64). Proteomic analysis has demonstrated that GNG12 regulates cell growth and casein synthesis by activating the Leu-mediated mTOR complex 1 signaling pathway (65). However, at present, a limited number of studies have been published regarding GNG12, and therefore further studies are required to determine its role in cancer. CENPE is a kinesin motor protein found in kinetochore protein complexes, whose motility is required for medium-term correct chromosomal alignment (66). Balamuth et al (67) reported that CENPE may be a novel target for neuroblastoma. In addition, CENPE has been revealed to be upregulated in invasive breast tumors compared with normal breast tissue (68). BUB1B exerts an important role in spindle assembly checkpoint signaling and the stable attachment of kinetochore and spindle microtubules (69-71). Therefore, the disruption of BUB1B function often leads to abnormal mitosis. A growing body of evidence suggests that BUB1B serves a key role in several types of cancer, including breast, stomach, colorectal and prostate cancer (72-75).

AURKA, a member of the serine/threonine kinase family, is localized on centrosomes and mitotic spindles, where it mediates mitotic progression and chromosomal stability (76). The AURKA gene is upregulated in numerous types of malignancies, including bladder, breast, colon, liver, ovarian, pancreatic,
gastric and esophageal cancer (77). Several previously published studies have revealed that upregulation of AURKA in clinical head and neck squamous cell carcinoma (HNSCC) specimens is associated with invasion, advanced stage and poor prognosis (78,79). Mignogna et al (80) revealed that AURKA may be used to predict resistance to platinum-based chemotherapy, and as a prognostic factor in ovarian cancer. Therefore, AURKA warrants further investigation in prospective clinical trials, and may have prognostic and therapeutic value in ovarian cancer.

In conclusion, the present study integrated multiple microarray datasets from the NCBI GEO database into one dataset, which was subsequently subjected to bioinformatics analysis. DEGs were identified, GO and KEGG analyses were performed and a PPI network of DEGs in serous epithelial ovarian cancer was constructed. DEGs were revealed to be mainly enriched in pathways associated with tumor formation and development, such as ‘Wnt signaling pathway’, ‘PI3K-Akt signaling pathway’, ‘pathways in cancer’ and ‘mTORC signaling pathway’, which provide a theoretical basis for studying the biological processes of serous ovarian cancer. In addition, the Oncomine database was used to validate the identified candidate genes across multiple databases. Finally, the effect of these genes on survival rate was investigated. Overall, the results obtained in the present study enhanced the understanding of the pathogenesis of serous epithelial ovarian cancer and provided novel avenues for investigating the potential molecular mechanisms.

The present study had important clinical implications for the early diagnosis, prognosis and development of more precise molecular therapies of ovarian cancer, although further studies are required to validate the identified candidate genes.

Acknowledgements
Not applicable.

Funding
No funding was received.

Availability of data and materials
The datasets generated and/or analyzed during the current study are available in the [FIGSHARE] repository (https://figshare.com/authors/Yubo_Zhang/6712286).

Authors’ contributions
YBZ designed the study, analyzed the data, and wrote the manuscript. YJ collected the data and drafted the manuscript. JW and JM designed the study and analyzed the data. SH designed the study and revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Patient consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.