Identification of Key Genes and Pathway for Ovarian Neoplasms Using the OVDM1 Cell Line Based on Bioinformatics Analysis

BCEF 1 Songna Yin
BF 1 Juan Du
CF 1 Jie Zhang
AEG 1 Xiang Zhang
AE 2 Ke Ma

Background: Ovarian neoplasms are the fifth most common cancer affecting the health of women, and they are the most lethal gynecologic malignancies; however, the etiology of ovarian neoplasms remains largely unknown. There is an urgent need to further broaden the understanding of the development mechanism of ovarian neoplasms through in vitro research using different cell lines.

Material/Methods: To screen the differentially expressed genes (DEGs) that may play critical roles in OVDM1 (an ovarian cancer cell line), the public microarray data (GSE70264) were downloaded and screened for DEGs. Then, Gene Ontology (GO) function analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were performed. To screen hub genes, the protein–protein interaction network was constructed. The expression level and survival analysis of hub genes in patients with ovarian neoplasms were also analyzed.

Results: There were 79 upregulated and 926 downregulated DEGs detected, and the biological processes of the GO analysis were enriched in extracellular matrix organization, extracellular structure organization, and chromosome segregation, whereas, the KEGG pathway analysis was enriched in cell cycle and cell adhesion molecules. The hub gene BIRC5, which might play a key role in ovarian neoplasms, was further screened.

Conclusions: The present study could deepen the understanding of the molecular mechanism of ovarian neoplasms using the OVDM1 cell line, which could be useful in developing clinical treatments of ovarian neoplasms.

MeSH Keywords: Gene Ontology • Ovarian Neoplasms • Protein Interaction Maps

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/915422
MOLECULAR BIOLOGY

Background

Ovarian neoplasms accounts for 2.5% of all malignancies in females, but 5% of female cancer deaths, due to low survival rates [1,2], which are mainly due to the late stage of diagnoses (stage III–IV) when tumors have metastasized to peritoneal and distant organs, at which point treatment is very difficult as the late-stage tumors become resistant to most cancer therapies [3–5]. According to the latest statistics from the World Health Organization, there were 65 538 new cases and 42 704 deaths caused by ovarian neoplasms in 2012 [6]. The diagnosis, prognosis, and identification of drug targets for ovarian neoplasms could benefit from comprehensive molecular profiling studies that reveal consequentially dysregulated genes and pathways underlying tumor progression and chemoresistance [7–10]. For example, the latest research has indicated that the inhibition of ATP11B expression could serve as a therapeutic strategy to overcome cisplatin resistance in ovarian neoplasms [11]. There was also a study that showed that the RNA binding protein SORB52 suppressed metastatic colonization of ovarian neoplasms by stabilizing tumor-suppressive immunomodulatory transcripts [12]. There have been other findings from studies that have demonstrated improved efficacy of using LPP-targeting siRNA in combination with cytotoxic drugs in high-grade serous ovarian neoplasms [13].

Olaparib was the first PARP (poly ADP ribose polymerase) inhibitor approved for maintenance treatment in patients with high-grade serous platinum-sensitive ovarian neoplasms [14]. There was another study that suggested that P2Y12 on platelets and ADP concentration at the interface between cancer cells and platelets could affect growth of primary ovarian neoplasms in mice, and inhibition of P2Y12 might be a new treatment option to consider for traditional surgery and chemotherapy for patients with ovarian neoplasms [15]. However, the underlying mechanism of the development of ovarian neoplasms has not yet been completely determined. Therefore, research of many more well-characterized cell lines derived from ovarian neoplasms are necessary due to the high level of genomic heterogeneity among and within ovarian tumors. The new ovarian neodiploid cell line, OVD1M, was derived from a highly aneuploid serous ovarian metastatic adenocarcinoma [3]. Hence, further study of the different molecular mechanisms of OVD1M and ovarian metastatic tumor tissue is necessary and will improve the understanding of the development of ovarian neoplasms.

In the present study, the array data of GSE70264 was downloaded from Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) to screen the differentially expressed genes (DEGs) between OVD1M and ovarian metastatic tumor tissue. Next, the DEGs were analyzed using a biological informatics approach. To identify the enriched biological functions and pathways of DEGs, the Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment of DEGs were performed. Furthermore, a protein interaction maps of DEGs was constructed; hub genes with a high degree of connectivity and the top 4 significant modules were chosen. Moreover, KEGG pathway enrichment of genes in the top 4 significant modules were also analyzed, and the expression level and survival analysis of hub genes were evaluated using the GEPIA online database (http://gepia.cancer-pku.cn/) [16]. These results may provide information for subsequent experimental studies of ovarian neoplasms using OVD1M, and contribute to the understanding of the molecular mechanisms underlying the development of ovarian neoplasms.

Material and Methods

Microarray data

The gene expression profile of GSE70264 was downloaded from the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) database. The data were produced based on the GPL570 platform ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array). The GSE70264 dataset contained 9 samples, including 3 samples of OVD1M cell lines at late passage, 3 samples of OVD1M cell lines at early passage, and 3 samples of ovarian metastatic tumor tissues.

Identification of DEGs

The limma package [17] in Bioconductor was used to identify DEGs between OVD1M cell line samples and ovarian metastatic tumor tissue. The P-values of DEGs were calculated using a t-test in R with the limma package. P value <0.05 and |logFC| >2 were set as the cutoff criterion for statistically significant DEGs. Then, 1402 DEGs were found, including 462 upregulated genes and 940 downregulated genes. The DEGs heat map was generated in R with the heatmap package.

GO and KEGG pathway enrichment analysis of DEGs

GO analysis is a common useful method for annotating genes and gene products and for identifying molecular function (MF), biological process (BP) and cellular component (CC) attributes for high-throughput genome or transcriptome data [18,19]. KEGG is a collection of databases used for systematic analysis of gene functions and associating related gene sets with their pathways [20]. GO annotation (P<0.01, q<0.05) and KEGG pathway (P<0.05) enrichment analyses were conducted for DEGs in R with the clusterProfiler package [21].
Figure 1. Heatmap of DEGs. Red represents higher expression level, green represents a lower expression level and black represents that there is no differential expression among the genes. DEGs, differentially expressed genes.
**Integration of protein–protein interaction (PPI) network and module analysis**

The online Search Tool for the Retrieval of Interacting Genes (STRING) database (https://string-db.org/) [22] was used to evaluate the interactive relationships among DEGs regarding the predicted and experimental interactions of proteins. In this study, the PPI of DEGs were identified by calculating the combined score (threshold, score > 0.4). Then, the PPI network was visualized using Cytoscape software (http://cytoscape.org/). The plug-in Molecular Complex Detection (MCODE), an automated method for finding molecular complexes in large protein interaction networks, was used to detect significant modules of PPI network in Cytoscape. The criteria were set as follows: degree cutoff=2, node score cutoff=0.2, k-core=2 and max. depth=100. Top modules from the protein interaction maps using the MCODE method with a score of >6.0. Also, the top 20 genes with high degree of connectivity were selected as hub genes.

**The expression level and survival analysis of hub genes**

The Gene Expression Profiling Interactive Analysis (GEPIA) is a newly developed interactive web server for analyzing the RNA sequencing expression data of 9736 tumors and 8587 normal samples from the TCGA and the GTEx projects, using...
a standard processing pipeline. GEPIA provides customizable functions such as tumor/normal differential expression analysis, profiling according to cancer types or pathological stages, patient survival analysis, similar gene detection, correlation analysis and dimensionality reduction analysis. Then, the expression level between cancer and healthy people, and prognostic information of the 20 hub genes was detected using GEPIA (http://gepia.cancer-pku.cn/).

### Results

#### Identification of DEGs

Based on the cutoff criteria (adjust P value <0.05, |logFC| >2). A total of 1402 DEGs (479 upregulated and 926 downregulated) were identified after the analyses of GSE70264. DEGs expressions were illustrated by Heat map (Figure 1)
GO term and KEGG pathway enrichment analysis

In order to gain insights into the biological roles of the DEGs, the GO categories and KEGG pathway enrichment were conducted for DEGs in R with the clusterProfiler package. For biological processes (BP), the DEGs were enriched in extracellular matrix organization, extracellular structure organization and chromosome segregation (Figure 2A). The DEGs were particularly enriched in cell component (CC), including proteinaceous extracellular matrix and MHC class II protein complex (Figure 2B). For molecular function (MF), the DEGs were enriched in glycosaminoglycan binding and endopeptidase inhibitor activity (Figure 2C). The most significantly enriched pathways of DEGs were analyzed by KEGG analysis, including cell cycle and protein digestion and absorption. Moreover, the top 20 genes with degree >125 of connectivity were selected as hub genes, including TOP2A, CDK1, CCNB1, VEGFA, BIRC5, CCNA2, PCNA, CDC20, MAD2L1, BUB1, CDKN3, NDC80, CCNB2, AURKA, AURKB, BUB1B, KIF11, PBK, RRM2, and CENPA.

Hub genes and module screening from PPI network

To further explore the relationships between DEGs at the protein level, the PPI networks of the DEGs were made based on the information in the STRING database, with a combined score >0.4 (Figure 3). The top four modules (modules 1, 2, 3, and 4) with score >10 were detected by MCODE in Cytoscape, the degree cutoff=2, node score cutoff=0.2, k-core=2, and max. depth=100 were set as the cutoff criteria (Figure 4). Further pathway analysis of genes in the top 4 module was performed by DAVID (Table 1). Enrichment analysis showed that the genes in module 1–4 were mainly implicated in cell cycle, Chemokine signaling pathway and protein digestion and absorption. Moreover, the top 20 genes with degree >125 of connectivity were selected as hub genes, including TOP2A, CDK1, CCNB1, VEGFA, BIRC5, CCNA2, PCNA, CDC20, MAD2L1, BUB1, CDKN3, NDC80, CCNB2, AURKA, AURKB, BUB1B, KIF11, PBK, RRM2, and CENPA.
Table 1. KEGG pathway analysis of differentially expressed genes in different modules (P<0.05, FDR <0.05).

| Module 1                      | Term                                      | P-value | FDR   | Genes                                                                 |
|-------------------------------|-------------------------------------------|---------|-------|----------------------------------------------------------------------|
| Cell cycle                    | 1.71E-16                                  | 1.89E-13|       | CDC6, CDK1, CDC20, CDC25C, MCM4, MUM, MCM6, CCNB1, CDC45, MAD2L1, CCNB2, PCNA, BUB1, BUB1B, CCNA2 |
| DNA replication               | 1.56E-08                                  | 1.35E-05|       | PRIM1, POLE2, PCNA, MCM4, MCM5, FEN1, MCM6                            |
| Progesterone-mediated oocyte maturation | 3.39E-06                              | 2.93E-03|       | CCNB1, CDK1, MAD2L1, CCNB2, BUB1, CDC25C, CCNA2                        |
| Oocyte meiosis                | 1.26E-05                                  | 1.09E-02|       | CDK1, MAD2L1, BUB1, FBXOS, AURKA, CDC20, CDC25C                        |
| Module 2                      | Chemokine signaling pathway               | 1.30E-09| 1.35E-06| ADCY3, PPBP, CCL20, CXCR4, AECV8, CXCL16, CXCL2, GNG4, CXCL10           |
| Module 3                      | Protein digestion and absorption          | 2.52E-12| 2.47E-09| COL18A1, COL13A1, COL3A1, COL22A1, COL1A2, COL15A1, COL12A1, COL24A1, COL11A1, COL10A1 |
| Inflammatory bowel disease (IBD) | 1.36E-06                               | 1.33E-03|       | HLA-DPA1, HLA-DMB, HLA-DOA, HLA-DM, TGBF2, HLA-DRA                   |
| Asthma                        | 2.00E-06                                  | 1.96E-03|       | HLA-DPA1, HLA-DMB, HLA-DOA, HLA-DM, HLA-DRA                          |
| Leishmaniasis                 | 2.28E-06                                  | 2.24E-03|       | HLA-DPA1, HLA-DMB, HLA-DOA, HLA-DM, TGBF2, HLA-DRA                   |
| Graft-versus-host disease     | 2.97E-06                                  | 2.91E-03|       | HLA-DPA1, HLA-DMB, HLA-DOA, HLA-DM, HLA-DRA                          |
| Allograft rejection           | 4.75E-06                                  | 4.65E-03|       | HLA-DPA1, HLA-DMB, HLA-DOA, HLA-DM, HLA-DRA                          |
| Rheumatoid arthritis         | 6.63E-06                                  | 6.50E-03|       | HLA-DPA1, HLA-DMB, HLA-DOA, HLA-DM, TGBF2, HLA-DRA                   |
| Type 1 diabetes mellitus      | 7.97E-06                                  | 7.81E-03|       | HLA-DPA1, HLA-DMB, HLA-DOA, HLA-DM, HLA-DRA                          |
| Intestinal immune network for | 1.26E-05                                  | 1.23E-02|       | HLA-DPA1, HLA-DMB                                                      |
| IgA production                |                                           |         |       |                                                                      |
| Amoebiasis                    | 1.65E-05                                  | 1.62E-02|       | COL3A1, COL1A2, COL24A1, COL11A1, TGBF2, FN1                           |
| Autoimmune thyroid disease    | 1.89E-05                                  | 1.85E-02|       | HLA-DPA1, HLA-DMB, HLA-DOA, HLA-DM, HLA-DRA                          |
| Staphylococcus aureus infection | 2.20E-05                             | 2.15E-02|       | HLA-DPA1, HLA-DMB, HLA-DOA, HLA-DM, HLA-DRA                          |
| Viral myocarditis             | 2.73E-05                                  | 2.67E-02|       | HLA-DPA1, HLA-DMB, HLA-DOA, HLA-DM, HLA-DRA                          |
| Toxoplasmosis                 | 2.78E-05                                  | 2.73E-02|       | HLA-DPA1, HLA-DMB, HLA-DOA, HLA-DM, TGBF2, HLA-DRA                   |

KEGG – Kyoto Encyclopedia of Genes and Genomes; FDR – false discovery rate.

The expression level and overall survival of hub genes

The GEPIA was used to detect the hub genes expression level between cancer and healthy people, as showed in Figure 5, BIRC5 significantly increased expression levels in ovarian neoplasms patients. The prognostic information of the 20 hub genes was freely available in http://gepia.cancer-pku.cn/index.html. It was found that expression of BIRC5 (HR 0.73, P=0.013) was associated with worse overall survival for ovarian neoplasms patients (Figure 6).

Discussion

Ovarian neoplasms mortality has fallen more than 30% since the mid-1970s due to reduced morbidity and improved treatment. However, less than half of women survive more than 5 years after diagnosis because of the predominance of invasive high-grade serous carcinomas [1,24,25]. Many researchers have suggested a highly complex genomic landscape in ovarian neoplasms [26]. The current treatments available for ovarian neoplasms have certain limitations due to the high level of genomic heterogeneity among and within ovarian tumors.
Thus, gene targeted therapy appears to be particularly important, making research on many more well-characterized cell lines derived from ovarian neoplasms necessary. Bioinformatic analysis has become a powerful tool for mapping the molecular basis related to the development of tumors.

In our study, a total of 1402 DEGs (479 upregulated and 926 downregulated) were recognized by comparing OVDM1 and ovarian metastatic tumor tissue. GO analysis results showed that these DEGs were mainly enriched in extracellular matrix organization, extracellular structure organization, and chromosome segregation, which is consistent with the previous studies [27,28]. The KEGG pathways of DEGs included cell cycle and cell adhesion molecules. It has been demonstrated by Woopen et al. that there is a significant association of epithelial cell adhesion molecule (EpCAM) overexpression with a more favorable survival in epithelial ovarian neoplasms patients [29].

Then, the module analysis of the PPI network showed that the cell cycle, DNA replication, and progesterone-mediated oocyte maturation were the most significant pathways in module 1. There are studies that have shown that prexasertib, a cell cycle checkpoint kinases 1 and 2 inhibitor, lead to synergistic cytotoxic effects against BRCA wild type high-grade serous ovarian neoplasms cells by reducing Rad51 foci formation and inducing apoptosis. Also, the systematic analysis of DNA replication process in ovarian neoplasms could reveal information on some molecular mechanisms of genetic damage accumulation and might contribute to the pathogenesis of the ovarian neoplasms [30,31].

To identify the key candidate genes, the PPI network of DEGs was analyzed in Cytoscape and **TOP2A**, **CDK1**, **CCNB1**, **VEGFA**, **BIRC5**, **CCNA2**, **PCNA**, **CDC20**, **MAD2L1**, **BUB1**, **CDCN3**, **NDC80**, **CCNB2**, **AURKA**, **AURKB**, **BUB1B**, **KIF11**, **PBK**, **RRM2**, and **CENPA** were generated. Next, the expression of these 20 hub genes in ovarian neoplasms compared with the normal and overall survival was verified on the GEPIA website.

It was found that **BIRC5** was highly expressed in patients with ovarian neoplasms compared with normal people, and was associated with poor survival. In addition, **BIRC5** is a molecular inhibitor of cell apoptosis that is frequently overexpressed in malignant cells, and Kuo et al. have advocated that the anti-apoptotic protein **BIRC5** can maintain the survival of HIV-1-infected CD4+ T cells [32]. Zhao et al. found that disrupted **BIRC5** expression in 3 ovarian cell lines (SKOV3, OVCAR3, and UACC1598) significantly reduced cell growth and invasion and induced cell apoptosis. However, the mechanism is still unclear [33]. Detailed investigations of **BIRC5** are sorely needed.

**Conclusions**

Altogether, the DEGs that were identified, like **BIRC5**, as well as the cell cycle, DNA replication, and progesterone-mediated oocyte maturation, may count for the development of ovarian neoplasms. However, there is still a dearth of research on the exact mechanism of these pathways and **BIRC5** in ovarian neoplasms using OVDM1 cells. Thus, these results remind us that further studies should focus on these pathways and **BIRC5**.
