Comprehensive analyses reveal three-gene signatures in ovarian cancer

Chuang Li

Shengjing Hospital of China Medical University

Yuan Lyu

Shengjing Hospital of China Medical University

Caixia Liu (✉ liucx1716@163.com)

<https://orcid.org/0000-0001-7019-0266>

Research

Keywords: KEGG pathway, GO function, Bioinformatic analysis, Cancer cells

Posted Date: May 18th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-28741/v1>

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

[Read Full License](#)

Abstract

Background: Ovarian cancer is a common cancer that affects the quality of women's life. With the limitation of the early diagnosis of the disease, ovarian cancer has a high mortality rate worldwide. However, the molecular mechanisms underlying tumor invasion, proliferation, and metastasis in ovarian cancer remain unclear. We aimed to identify, using bioinformatics, important genes and pathways that may serve crucial roles in the prevention, diagnosis, and treatment of ovarian cancer.

Methods: Three microarray datasets (GSE14407, GSE36668, and GSE26712) were selected for whole-genome gene expression profiling, and differentially expressed genes were identified between normal and ovarian cancer tissues. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analyses were performed using DAVID. Additionally, a protein-protein interaction network was constructed to reveal possible interactions among the differently expressed genes. The prognostic values of the hub genes were investigated using Gene Expression Profiling Interactive Analysis (GEPIA) and the KM plotter database. Meanwhile, the mRNA expression analysis of the hub genes was performed using the GEPIA database.

Results: We obtained 247 upregulated and 530 downregulated differently expressed genes, and 52 hub genes in the significant gene modules. Enrichment analysis revealed that the hub genes were significantly ($P < 0.05$) associated with proliferation. Additionally, BIRC5, CXCL13, and PBK were revealed to be significantly associated with the clinical prognosis of patients with ovarian cancer. Immunohistochemical staining results obtained from the Human Protein Atlas revealed that BIRC5, PBK, and CXCL13 were highly expressed in ovarian cancer tissues.

Conclusion Three-gene signatures (BIRC5, CXCL13, and PBK) are associated with the occurrence, development, and prognosis of OC, and may therefore serve as biological markers of the disease.

Background

Ovarian cancer (OC) is the sixth most common meladies among women worldwide. However, even with advances in detection and therapeutic methods, the mortality rate in women is still high[1]. The most common cause of high mortality in OC is that patients experience nonspecific symptoms, if any, in the early stages of the disease that can result in delayed diagnosis. More than 75% of OC patients are diagnosed at an advanced stage, which hinders the implementation of effective treatment measures [2, 3]. Therefore, early screening and diagnosis is essential for OC patients.

To date, many genome-wide expression profiles have been acquired by gene-chip microarrays. We can identify new screening or treatment targets for different diseases using unbiased bioinformatic tools, and also understand potential biological functions and molecular mechanisms by analyzing gene expression data using pathway profiling software. A large number of datasets have been generated by gene-chip microarray in OC patients[4]; differently expressed genes (DEGs) can be analyzed by gene ontology (GO) function annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment

analyses to explore the possible mechanisms for disease development[5]. This prompted us to investigate potential biomarkers in OC and further validate the molecular mechanism through cellular experiments in the future.

Results

Identification of DEGs

We identified a total of 786 genes that met the above criteria, including 247 up-regulated genes and 539 down-regulated genes (**Supplementary Table 1**). Then, we plotted a heatmap to represent the expression of DEGs by pheatmap package in R(**Supplementary Fig. 1**).

GO function and KEGG pathway enrichment

We carried out GO function and KEGG pathway enrichment analyses with the criteria of $p < 0.05$ to explore the potential molecular mechanisms for DEGs using DAVID (Fig. 2). We presented the top ten significant terms for BP, CC, and MF by bubble plot based on p-values using the ggplot2.R package; while the top ten significant terms in the KEGG pathways were visualized using the GOplot.R package. The GO function enrichment analysis results showed that the DEGs were mainly enriched in the terms involving biological processes, including cellular protein modification processes, response to organic substances, regulation of signaling, regulation of cell communication, and positive regulation of metabolic processes (Fig. 2a). The cellular components were mainly involved in the extracellular region, membrane-bounded vesicles, extracellular exosomes, extracellular vesicles, and extracellular organelles (Fig. 2b). Regarding molecular function, the DEGs were significantly enriched in protein binding, protein homodimerization, calcium ion binding, heparin binding, and microtubule activities (Fig. 2c). The most enriched KEGG pathway terms for the DEGs were metabolic pathways, pathways in cancer, PI3K-Akt signaling pathway, proteoglycans in cancer, and ECM-receptor interaction (Fig. 2d).

PPI networks and modules

We uploaded the DEGs to the STRING website to construct the PPI network that were visualized using Cytoscape software. A total of 779 nodes with 1548 edges with scores ≥ 0.900 (the highest confidence) were selected to construct the PPI networks. The two significant modules were selected using the MOCODE plug-in. The first module consisted of 28 nodes and 356 edges (**Supplementary Fig. 2a**). In this module, 28 DEGs were up-regulated according to the gene expression profiling in the module. The second module consisted of 24 nodes and 144 edges (**Supplementary Fig. 2b**). In this module, 9 genes were up-regulated and 15 genes were down-regulated according to the gene expression profiling among the common hub genes in the module. For a more in-depth understanding of the common hub genes, we performed GO function and KEGG pathway enrichment analyses via the DAVID website. we found that the DEGs in Module 1 were mainly associated with the mitotic cell cycle process (BP), the microtubule cytoskeleton (CC), and heterocyclic compound binding (MF) (**Supplementary Fig. 3a**); whereas, DEGs in

Module 2 were mainly associated with the G-protein coupled receptor signaling pathway (BP), the extracellular region (CC), and receptor binding (MF) (**Supplementary Fig. 3b**).

Validation of DEG expression levels and survival analysis of common hub genes

The prognostic information involving PFS and OS was obtained from the KM plotter and GEPIA database respectively. We found that *BIRC5* (log-rank $p = 1.6e-02$), *PBK* (log-rank $p = 3.5e-02$), *CXCL 13* (log-rank $p = 9.4e-03$) were significantly correlated with the PFS of ovarian cancer patients based on the KM plotter. In order to determine the reliability of the prognosis values, we further analyzed the effect of the above genes on the overall survival of ovarian cancer patients using GEPIA. We found that only *PBK* (log-rank $p = 4.4e-02$), *BIRC5* (log-rank $p = 1.3e-02$), and *CXCL 13* (log-rank $p = 4.2e-04$) were correlated with the overall survival of ovarian cancer patients. Moreover, the mRNA expression levels of *PBK*, *BIRC5*, and *CXCL 13* were also higher in OV tissues compared to normal tissues based on GEPIA database analysis (Fig. 3a).

Real hub gene validation

We further compared the transcriptional levels of the real hub genes in cancer tissue with normal tissues using ONCOMINE databases. ONCOMINE analysis revealed that the mRNA expression levels of *BIRC5*, *CXCL 13*, *PBK* were significantly up-regulated in ovarian cancer tissues compared to normal tissues (Fig. 3b). Immunohistochemical staining data from HPA demonstrated the deregulated protein expression level of *PBK*, *CXCL 13*, and *BIRC5* (Fig. 3b). Therefore, we performed downstream analysis for *BIRC5*, *CXCL 13*, and *PBK*.

Mining genetic alterations and co-expressed genes

We analyzed *PBK*, *BIRC5*, and *CXCL 13* using cBioportal to explore the genomic alterations in ovarian cancer. Among the three ovarian cancer studies, alterations ranging from 6.54–14.92% were found for the gene sets submitted for analysis (Fig. 3c); the individual frequencies of alteration in *PBK*, *BIRC5*, and *CXCL 13* were shown in Fig. 3d respectively. In order to deepen our understanding on the potential function of *PBK*, *BIRC5*, and *CXCL 13*, we investigated the biological processes of the top 50 genes co-expressed with them. Regarding *PBK* and *BIRC5*, the mainly enriched biological processes were GO:0007067 (mitotic nuclear division), GO:0051301 (cell division), and GO:0007062 (sister chromatid cohesion) (**Supplementary Fig. 4a-b**). As for *CXCL 13*, the mainly enriched biological processes were GO:0006955 (immune response), GO:0042110 (T cell activation), GO:0031295 (T cell co-stimulation) (**Supplementary Fig. 4c**).

Results of GSEA and GSVA

In order to further understand the potential molecular mechanisms, we performed GSEA and GSVA to explore the pathways enriched in the hub gene high expression group. In the *BIRC5* high expression group, the results of GSEA revealed that the top three enriched KEGG pathways were DNA replication, RNA polymerase, and homologous recombination (Fig. 4a). In the *CXCL 13* high expression group, the results of GSEA revealed that the top three enriched KEGG pathways were primary immunodeficiency, intestinal

immune network for IgA production, and cytosolic DNA sensing pathway (Fig. 4b). As for the *PBK* high expression group, the results of GSEA revealed that the top three enriched KEGG pathways were citrate cycle (TCA cycle), RNA degradation, and oxidative phosphorylation (Fig. 4c). Furthermore, GSVA further revealed the significant association between *BIRC5*, *PBK*, *CXCL 13* and proliferative process (Fig. 4d-f).

miRNA-mRNA regulatory network

The network of predicted miRNA and *PBK*, *CXCL 13*, and *BIRC5* mRNA was shown in Supplementary Fig. 5. Notably, 62 miRNA targeted *BIRC5*, including hsa-miR-548ap-3p, hsa-miR-548aa, and hsa-miR-548t-3p; 53 miRNA targeted *CXCL 13*, including hsa-miR-1304-5p and hsa-miR-6124; 52 miRNA targeted *PBK*, including hsa-miR-548ap-3p, hsa-miR-548aa, hsa-miR-548t-3p, hsa-miR-1304-5p, and hsa-miR-6124(Supplementary Fig. 5).

Discussion

Ovarian cancer comprises a histologically and genetically broad range of tumors, including those of epithelial, sex cord-stromal and germ cell origin [5]. In this study, we identified potential gene targets for OC by comparing normal ovarian tissues and OC tissues using bioinformatic analyses on the basis of three profile datasets (GSE14407, GSE36668, and GSE26712). We identified 247 up-regulated and 539 down-regulated DEGs, which we then inspected with GO function and KEGG pathway enrichment analyses. The results of the enrichment analysis revealed that the genes were mainly enriched in the various types of cancer-related functions and pathways. In order to identify the protein interactions and key modules, we uploaded the DEGs to the STRING website. A PPI network that contained 779 nodes/genes and 1548 edges was constructed. The first key module was chosen from the PPI network: it contained 28 nodes/genes and 356 edges. The second key module contained 24 nodes/genes and 144 edges. In order to identify the prognosis value of the common hub genes from the above two gene modules, we performed the PFS and OS analyses using KM plotter and GEPIA database respectively; we confirmed the transcription level of the hub genes using the GEPIA database. Then, we identified *CXCR4*, *BIRC5*, *CXCL 13*, and *PBK* for the follow-up analysis; we also found that *CXCR4*, *BIRC5*, *CXCL 13*, and *PBK* were strongly positive in the ovarian cancer data set of the Oncomine database. However, *CXCR4* was excluded by HPA database analysis because of the lack of immunohistochemical staining data, so we performed the downstream analysis for *PBK*, *BIRC5*, and *CXCL 13*. Based on the cBioportal database, we found that the genomic alteration of the real hub genes ranged from 6.54–14.92% in the three ovarian cancer datasets. To clarify the biological role of *PBK*, *BIRC5*, and *CXCL 13*, we selected top 50 genes co-expressed with them and performed Go function analysis to explore the potential biological behavior related to these hub genes. Regarding the co-expressed genes of *PBK* and *BIRC5*, the biological processes were enriched in the mitotic nuclear division, cell division, and sister chromatid cohesion. Regarding co-expressed genes of *CXCL 13*, the biological processes were mainly enriched in the immune response, T cell activation, T cell costimulation. The results of GSEA and GSVA indicated that many cell cycle-related pathways were enriched in the high-expression groups of the real hub genes, suggesting their contribution

to ovarian carcinoma cell proliferation. Moreover, our results also revealed that *CXCL 13* could be related to the immune regulation in ovarian cancer microenvironment.

In total, we performed comprehensive bioinformatic analysis to identify three gene signatures that were related to the survival and prognosis of ovarian cancer. *BIRC5* encodes the protein survivin, which is a member of the apoptosis inhibitor family. Survivin is highly expressed in a variety of cancer tissues compared to normal tissues. It was reported that expression level of survivin (*BIRC5*) gene is significantly increased in primary prostate cancers and metastases, but there is no difference between recurrent and non-recurrent prostate cancer[7]. The chemokine ligand 13 (*CXCL13*) and its chemokine receptor 5 (*CXCR5*) play fundamental roles in inflammatory, infectious and immune responses. And the *CXCL13/CXCR5* signaling axis is a crucial role in the cancer cell growth, survival, invasion, and metastatic dissemination[8]. Wei et al. reported that intra tumoral *CXCL13* expression can be used as an independent prognostic marker for patients after gastric cancer resection, and that the expression level of *CXCL13* may appear as a predictive biomarker of response to postoperative chemotherapy in these patients[9]. PDZ-binding kinase (PBK) promotes malignant tumor progression in various cancers such as colorectal and oral cancer. Su et al. found that PBK/TOPK expression detected by the immunohistochemical staining was significantly associated with prognosis in colorectal cancer patients based on a retrospective analysis[10]. Chang et al. found that PBK was highly expressed in oral cancer patients, and that high PBK/TOPK expression could be a favorable prognosis marker for these patients[11]. Our study further identified that PBK could be a potential biomarker in OC; however, the results need to be further validated by molecular and cellular experiments.

Conclusion

In conclusion, we identified three potential therapeutic targets for ovarian cancer: *BIRC5*, *CXCL 13*, and *PBK*. These genes have been previously reported to be involved in many pathways associated with tumorigenesis, but the link with ovarian cancer have not been completely identified. All of the candidate DEGs that are ovarian cancer-related genes should be further confirmed via molecular biological experiments. Therefore, we will further investigate the biological function of *PBK*, *BIRC5*, and *CXCL 13* in ovarian cancer by cytology experiments.

Methods

Microarray data preprocessing and analysis

We downloaded the microarray files GSE14407, GSE36668, and GSE26712 from the public GEO database[12, 13](<https://www.ncbi.nlm.nih.gov/geo/>) which included 12 OC tissues and 12 normal ovary tissues, 4 OC tissues and 4 normal ovary tissues, and 185 OC tissues and 10 normal ovary tissues, respectively. Because there is no gene symbol in the list of GSE14407, GSE36668, and GSE26712 datasets, we converted the probe names to the gene symbol according to the Whole Human Genome Microarray 4 × 44K G4112F (Probe Name Version). However, a large number of probes can correspond to

the same gene symbol, so we used the p-value as the criterion for the downstream differential gene expression analysis. The probes with the most significant p-values were chosen to correspond to DEGs(Fig. 1).

Data processing of DEGs

We merged the GSE14407, GSE36668, and GSE26712 datasets and performed batch normalization using the surrogate variable analysis (sva) and limma packages[14–17]. Then we further used the limma R package to identify the DEGs in the normal ovarian tissue and paired tumor tissue. Genes that met the following criteria were chosen as DEGs[18]: $|\log_2\text{-fold change}| \geq 1.0$ and adjusted p-values < 0.05 (moderated t-statistics corrected by Benjamini and Hochberg's method).

GO function and KEGG pathway enrichment analyses

We performed GO function and KEGG pathway enrichment analyses for the DEGs. GO function can reveal the possible mechanisms involving these DEGs in the three aspects of biological function, including the biological process (BP), molecular function (MF), and cellular component (CC)[19]. KEGG pathway enrichment analysis can provide the known signal metabolic pathway in which these DEGs are involved. We used the DAVID website to perform the GO function and KEGG pathway enrichment analyses for DEGs according to the p-value < 0.05 criteria. The DAVID website (<https://david.ncifcrf.gov/>) is a web-accessible program that provides comprehensive function enrichment tools to understand the biological meanings for genes.

PPI networks and module analysis

We used the Search Tool for the Retrieval of Interacting Genes (STRING) to find the key proteins and construct PPI networks. The STRING database contains more than 24.6 million proteins and 2 billion interactions from 5,090 organisms[20]. We uploaded the DEGs to the STRING database and set the interaction score ≥ 0.9 (the highest confidence) as the threshold. We used the Cytoscape software to visualize the PPI networks[21, 22]. We investigated the key modules that were constructed by Molecular Complex Detection (MCODE) in Cytoscape software. The analysis parameters k-score ≥ 8 and $p < 0.05$ were set as the significant thresholds. We performed the GO function and KEGG pathway enrichment analyses for the DEGs in the key modules.

Validation of hub genes

In order to validate the reliability of the hub genes, we uploaded said genes into the Gene Expression Profiling Interactive Analysis (GEPIA) database[23]. We identified gene expression with $|\log \text{fold change}| > 1.0$ and p-value < 0.001 as different expression genes. In this study, the effect of the hub genes on overall survival (OS) and progression-free survival (PFS) was evaluated combining the GEPIA and Kaplan–Meier plotter databases, respectively. The OS and PFS were determined by splitting ovarian cancer patients into two groups based on the median expression of hub genes (high vs. low expression) and assessing with log-rank test.

Oncomine database and The Human Protein Atlas database analyses

Oncomine database (<http://www.oncomine.org>)[24] is a publicly accessible, online cancer gene information mining platform for the purpose of cancer genome expression analysis. To date, the Oncomine database has collected 715 datasets and 86,733 cancer and normal tissue sample data. This database was used to identify the different gene expression and perform clinical and pathological analyses for various types of cancer. In this study, to investigate the correlation between the expression of the hub genes and tumorigenesis, we evaluated the expression level of the hub genes in common cancer types using the Oncomine database. We further analyzed the protein expression and distribution of the hub genes in ovarian cancer tissues and normal tissues by The Human Protein Atlas (HPA) database.

Exploring cancer genomic data in TCGA

The cBioPortal for Cancer Genomics (<http://cbioportal.org>)[25] is a cancer genome database that helps facilitate research from multidimensional cancer genomic data analysis. cBioportal comprises a collection of more than 200 studies, including all TCGA projects and published studies from the literature. In this research, we explored the genetic alterations and co-expression of hub genes. In order to investigate the potential biological functions of real hub genes, we further performed GO function enrichment analysis for the top 50 co-expressed genes respectively.

Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA)

We utilized the GSEA software to perform GSEA analysis of the real hub genes with merged dataset (GSE14407, GSE26712, and GSE36668). Then, we used the “GSVA” R package to further explore the pathways most significantly correlated with these hub genes[26]. In this study, we divided the merged datasets (GSE14407, GSE36668, and GSE26712) into two groups according to the median expression level of real hub genes. Terms with normalized enrichment score ≥ 1 and false discovery rate < 0.25 were identified in the GSEA. We considered terms with a p value ≤ 0.001 statistically significant in the GSVA.

Construction of the miRNA-mRNA network

The miRDB database (<http://www.mirdb.org/>) [27, 28] is an online platform for miRNA target prediction and functional annotation. Targets in the miRDB were based on the bioinformatic tool MirTarget that predicts miRNA targets with machine learning methods. In this study, we used miRDB to predict the miRNA associated with our hub gene targets. Next, the miRNA-mRNA regulatory networks were visualized via Cytoscape.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the GEO database[<https://www.ncbi.nlm.nih.gov/geo/>].

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable

Authors' contributions

All authors participated in data processing and analysis.

Acknowledgments

Not applicable.

Abbreviations

OC: ovarian cancer; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; DEGs: differently expressed genes; GSEA: Gene set enrichment analysis; GSVA: gene set variation analysis; OS: overall survival; PFS: progression-free survival

References

- [1]Deb B, Uddin A, Chakraborty S. miRNAs and ovarian cancer: An overview. J Cell Physiol. 2018; 233: 3846-3854. doi: 10.1002/jcp.26095.
- [2]Elzek MA, Rodland KD. Proteomics of ovarian cancer: functional insights and clinical applications. Cancer Metastasis Rev. 2015; 34: 83-96. doi: 10.1007/s10555-014-9547-8.
- [3]Emmings E, Mullany S, Chang Z, Landen CN, Jr Linder S, Bazzaro M. Targeting Mitochondria for Treatment of Chemoresistant Ovarian Cancer. Int J Mol Sci. 2019; 20: 229. doi: 10.3390/ijms20010229.

- [4]Tassi RA, Gambino A, Ardighieri L, Bignotti E, Todeschini P, Romani C, et al. FXYD5 (Dysadherin) upregulation predicts shorter survival and reveals platinum resistance in high-grade serous ovarian cancer patients. *Br J Cancer*. 2019; 121: 584-592. doi: 10.1038/s41416-019-0553-z.
- [5]Mou T, Zhu D, Wei X, Li T, Zheng D, Pu J, et al. Identification and interaction analysis of key genes and microRNAs in hepatocellular carcinoma by bioinformatics analysis. *World J Surg Oncol*. 2017;15: 63. doi: 10.1186/s12957-017-1127-2.
- [6]Karnezis AN, Cho KR, Gilks CB, Pearce CL, Huntsman DG. The disparate origins of ovarian cancers: pathogenesis and prevention strategies. *Nat Rev Cancer*. 2017; 17: 65-74. doi: 10.1038/nrc.2016.113.
- [7]Hennigs JK, Minner S, Tennstedt P, Löser R, Huland H, Klose H, et al. Subcellular Compartmentalization of Survivin is Associated with Biological Aggressiveness and Prognosis in Prostate Cancer. *Sci Rep*. 2020; 10: 3250. doi: 10.1038/s41598-020-60064-9.
- [8]Hussain M, Adah D, Tariq M, Lu Y, Zhang J, Liu J. CXCL13/CXCR5 signaling axis in cancer. *Life Sci*. 2019; 227: 175-186. doi: 10.1016/j.lfs.2019.04.053.
- [9]Wei Y, Lin C, Li H, Xu Z, Wang J, Li R, et al. CXCL13 expression is prognostic and predictive for postoperative adjuvant chemotherapy benefit in patients with gastric cancer. *Cancer Immunol Immunother*. 2018; 67: 261-269. doi: 10.1007/s00262-017-2083-y.
- [10]Su TC, Chen CY, Tsai WC, Hsu HT, Yen HH, Sung WW, et al. Cytoplasmic, nuclear, and total PBK/TOPK expression is associated with prognosis in colorectal cancer patients: A retrospective analysis based on immunohistochemistry stain of tissue microarrays. *PLoS One*. 2018; 13: e0204866. doi: 10.1371/journal.pone.0204866.
- [11]Chang CF, Chen SL, Sung WW, Hsieh MJ, Hsu HT, Chen LH, et al. PBK/TOPK Expression Predicts Prognosis in Oral Cancer. *Int J Mol Sci*. 2016; 17: 1007. doi: 10.3390/ijms17071007.
- [12]Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002; 30: 207-10. doi: 10.1093/nar/30.1.207.
- [13]Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, et al. NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res*. 2009; 37: D885-890. doi: 10.1093/nar/gkn764.
- [14]Smyth GK, Speed T. Normalization of cDNA microarray data. *Methods*. 2003; 31: 265-73. doi: 10.1016/s1046-2023(03)00155-5.
- [15]Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004; 3: Article3. doi: 10.2202/1544-6115.1027.

- [16]Dessau RB, Pipper CB. ["R"-project for statistical computing]. *Ugeskr Laeger*. 2008;170:328-30.
- [17]Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012; 28: 882-3. doi: 10.1093/bioinformatics/bts034.
- [18]Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med*. 1990; 9: 811-8. doi: 10.1002/sim.4780090710.
- [19]Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25: 25-29. doi: 10.1038/75556.
- [20]Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013; 41: D808-15. doi: 10.1093/nar/gks1094.
- [21]Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol*. 2011; 696:291-303. doi: 10.1007/978-1-60761-987-1_18.
- [22]Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 2011; 27: 431-2. doi: 10.1093/bioinformatics/btq675.
- [23]Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017; 45: W98-W102. doi: 10.1093/nar/gkx247.
- [24]Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004; 6: 1-6. doi: 10.1016/s1476-5586(04)80047-2.
- [25]Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012; 2: 401-404. doi: 10.1158/2159-8290.CD-12-0095.
- [26]Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013; 14: 7. doi: 10.1186/1471-2105-14-7.
- [27]Liu W, Wang X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol*. 2019; 20: 18. doi: 10.1186/s13059-019-1629-z.
- [28]Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res*. 2020; 48: D127-D131. doi: 10.1093/nar/gkz757.

Figures

Figure 1

Flowchart of the study. 3 public ovarian cancer datasets were included, and we performed comprehensive analysis to identify the gene signatures in ovarian cancer.

Figure 2

The top10 terms of enrichment analysis. (a) Bubble plot depicting the GO terms of biological process involved DEGs; (b) Bubble plot depicting the GO terms of cellular component involved DEGs; (c) Bubble plot depicting the GO terms of molecular function involved DEGs; (d) Chord plot depicting the relationship between genes and KEGG pathways.

Figure 3

Validation of hub genes and Alteration frequency of BIRC5, CXCL13, PBK. (a) Gene expression of common hub genes of OV database based on GEPIA database; (b) The graph shows the numbers of datasets with up-regulated expression (red) or down-regulated expression (blue) of the BIRC5, PBK, CXCL13 and the corresponding immunochemical staining; (c) Summary of alterations in combined BIRC5, PBK, CXCL13; (d) OncoPrint visual summary of alteration of BIRC5, PBK, CXCL13

a

b

c

Figure 4

The results of GSEA. (a-c) Top 3 gene sets enriched in the high-expression group of single hub genes (a: BIRC5, b: CXCL13; c: PBK).

Figure 5

The results of GSVA. GSVA-derived clustering heatmaps of differentially expressed pathways for single hub genes (a: BIRC5, b: CXCL13; c: PBK).

Supplementary Files

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

- [SupplementaryFigure1.pdf](#)

- [SupplementaryFigure5.pdf](#)
- [SupplementaryFigure2.pdf](#)
- [SupplementaryFigure4.pdf](#)
- [SupplementaryFigure3.pdf](#)
- [SupplementaryTable1.doc](#)