

# Identification and Analysis of Genes Associated with Epithelial Ovarian Cancer by Integrated Bioinformatics Methods

**Ting Gui**

Peking Union Medical College Hospital

**Chenhe Yao**

Beijing Zhicheng Biomedical Technology Co, Ltd

**Binghan Jia**

Beijing Zhicheng Biomedical Technology Co, Ltd

**Keng Shen** (✉ [sk\\_shenkeng@sina.com](mailto:sk_shenkeng@sina.com))

Peking Union Medical College Hospital

---

## Research

**Keywords:** epithelial ovarian cancer (EOC), differentially expressed genes (DEGs), biomarker, prognosis, bioinformatical analysis

**Posted Date:** July 7th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-39371/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** Though considerable efforts have been made to improve the treatment of epithelial ovarian cancer (EOC), the prognosis of patients has remained poor. Identifying differentially expressed genes (DEGs) involved in EOC progression and exploiting them as novel biomarkers or therapeutic targets for EOC is highly valuable.

**Methods:** Overlapping DEGs were screened out from three independent gene expression omnibus (GEO) datasets and subjected to Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analyses. The protein-protein interactions (PPI) network of DEGs was constructed in the STRING database. The top 20 hub genes were selected using cytoHubba. The expression of hub genes was detected in GEPIA, Oncomine, and human protein atlas (HPA) databases. The relationship of hub genes with the pathological stage and the overall survival and progression-free survival in EOC patients was investigated using the cancer genome atlas data.

**Results:** A total of 306 DEGs were identified, including 265 up-regulated and 41 down-regulated. Through the PPI network analysis, the top 20 genes were screened out, among which 4 hub genes were selected after literature retrieval, including CDC45, CDCA5, KIF4A, ESPL1. The four genes were up-regulated in EOC tissues and the expression of these four genes decreased gradually with the continuous progression of EOC. Survival curves illustrated that patients with a lower level of CDCA5 and ESPL1 had better overall survival and progression-free survival.

**Conclusions:** Two hub genes, CDCA5 and ESPL1, identified as playing tumor-promotive roles, could be utilized as potential novel therapeutic targets for EOC treatment.

## Background

Ovarian cancer has the highest mortality in gynecologic cancers and most patients are diagnosed at advanced stages [1]. Many patients would still relapse even if they are treated with satisfied cytoreductive surgery (CRS) combined with standard platinum-based chemotherapy. The 5-year survival rate for patients with advanced ovarian cancer is about 30% [2]. Thus, investigating reliable and effective molecular markers and understanding essential genes involved in the biological process of ovarian cancer is urgently needed.

The Gene Expression Omnibus (GEO) database is a free online database that stores a multiple of high-throughput microarrays, chips, and next-generation sequence functional genomic data sets [3]. The GEO database could be utilized to screen out differentially expressed genes (DEGs), to explore molecular signals and correlations, and to analyze gene regulatory networks. However, due to high expenses and limited sample tissues, the analysis results of individual experiments may be biased and unreliable. Therefore, integrated analyses of multiple datasets may improve the accuracy and reliability of the analysis and hence produce a more comprehensive and well-rounded discovery of DEGs in a variety of cancers.

In this study, original data from microarray analysis conducted on epithelial ovarian cancer (EOC) samples were downloaded from the GEO database, and integrated analysis was implemented. A total of 20 overlapping DEGs were screened out by the intersection of three independent datasets. Then, gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis, and protein-protein interactions (PPI) were performed to evaluate the underlying molecular mechanisms involved in carcinogenesis and tumor progression. The hub genes that may play essential roles in EOC were identified, and the relative expression level of hub genes and their relationship with EOC patient survival were validated in multiple online databases.

## Materials And Methods

### Microarrays and bioinformatics analysis

The original CEL files of three independent GEO datasets (GSE119056, GSE54388, GSE66957) were downloaded and analyzed using R language. We utilized the affy package to perform the background correction and data normalization, including conversion of raw data formats, imputation of missing values and background correction. Then, the samples were subjected to differential expression analysis using the Limma package.  $p < 0.05$  and  $|\log FC| > 1$  were set as the threshold, and the genes that met the criteria were screened out as DEGs. The intersection of DEGs from three datasets was performed using the VennDiagram package in R language.

### Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis

KOBAS v 3.0 (<http://lobas.cbi.pku.edu.cn>) is a web server for gene/protein functional annotation and functional gene set enrichment [4]. The overlapping DEGs from three GEO datasets were subjected to GO and KEGG pathway analysis by this online tool.  $p < 0.05$  was considered as statistically significant.

### Identification of top modules and hub genes in a protein-protein interaction (PPI) network

Exploring the functional interactions between proteins is essential for understanding the molecular mechanisms of EOC. The Search Tool for the Retrieval of Interacting Genes (STRING v 11.0) is an online tool (<http://string-db.org>) designed to establish potential interactions among a good number of genes [5]. The overlapping DEGs were put into the software to construct the PPI network and were then visualized using Cytoscape software 3.7 (<http://www.cytoscape.org>) [6]. In addition, the Molecular Complex Detection (MCODE) in Cytoscape software (Cytoscape v 3.7.1) was applied to screen out top modules inside the PPI network with degree cut-off = 2, node score cut-off = 0.2, Max depth = 100 and k-score = 2. CytoHubba, a plugin in the Cytoscape software, was adopted to calculate the degree of each protein node. In our study, the top twenty genes were selected as hub genes.

### Validation of hub gene expression levels in multiple database

To validate the mRNA level of hub genes in EOC, we examined the relative expression of these genes in two databases, namely, GEO and GEPIA. GEO (Gene Expression Omnibus;

<http://www.ncbi.nlm.nih.gov/geo/>), an online platform providing microarray datasets and data-mining functions, could be applied to validate the expression of specific genes in multiple diseases including cancers, thereby facilitating the discovery of potential essential genes involved in disease development and progression [7]. GEPIA (Gene Expression Profiling Interactive Analysis; <http://gepia.cancer-pku.cn/>), a web-based tool based on the cancer genome atlas (TCGA) and GTEx data, could be utilized to conduct differential expression analysis, correlation analysis, patient survival analysis, similar gene detection and dimensionality reduction analysis [8]. In the current study, we detected the relative expression of three hub genes in these two databases with a threshold of  $p < 0.05$  and a fold change of 2.

### **Survival analysis and tumor stage/grade analysis of hub genes**

UCSC Xena v 1.0 (<http://xenabrowser.net/>) is an online database from which users could obtain functional genomic data sets to make correlations between genomic and/or phenotypic variables. In this study, we utilized this free online tool to detect whether the expression of hub genes was correlated with the overall survival and tumor stage/grade of EOC patients from TCGA samples. Patients were grouped into a relatively high expression group and a low expression group according to the median, and  $p < 0.05$  was considered as statistically significant.

### **Statistical analysis**

All the statistical analysis in this study was performed using SPSS 21.0 software. Comparisons between the two groups were performed using Student's two-tailed  $t$  test. Kaplan-Meier survival analysis was performed to compare EOC patient survival based on hub gene expression using log-rank test.  $p < 0.05$  was considered statistically significant.

## **Results**

### **1. Identification of DEGs from GEO datasets analysis**

Raw data from three independent datasets (GSE119056, GSE54388, GSE66957) were downloaded from GEO and then subjected to differential expression analysis using R language.. Genes screened from the criteria set as  $p < 0.05$  plus  $|\log FC| > 1$  were plotted using R language to visualize the distribution of DEGs between EOC and compared normal controls from three datasets. The volcano plot of each gene expression profile data was shown in Fig. 1a-c. Red or blue dots represent significantly up-regulated or down-regulated genes, respectively. Afterwards, three hundred and six overlapping DEGs in total, including 265 up-regulated genes and 41 down-regulated genes, were found in EOCs compared with adjacent ovarian tissues under the criteria (Fig. 1d-1e).

### **2. GO analysis and KEGG analysis of the overlapping DEGs**

To gain a more in-depth understanding of the common DEGs from three datasets, GO analysis and KEGG pathway enrichment analysis were performed in KOBAS. The top 10 biological processes that these DEGs

involved in was presented in Fig. 2a, among which cell division, cell proliferation, adhesion, and response to drug were closely associated with cancer progression. Regarding cellular component, GO analysis results showed that the overlapping DEGs were mainly enriched in cytoplasm, nucleus, cell membrane, and extracellular exosome (Fig. 2b). It was well-established that extracellular exosome could participate in the malignant progression of cancer including EOC. For molecular function classification, the DEGs were significantly enriched in the following functions: protein binding, ATP binding, poly A RNA binding, and chromatin binding (Fig. 2c). The results from KEGG analysis showed that these DEGs were particularly enriched in pathways in cancer, cell cycle, and carbon metabolism (Fig. 2d). The above findings consistently indicated that these DEGs might modulate EOC proliferation and metastasis through multiple signaling pathways.

### **3. PPI network of common DEGs and hub gene identification**

To further explore the underlying association between DEGs, a PPI network was constructed by using the STRING database (Fig. 3a). Then, the top two modules inside the PPI network were identified with the MCODE application in Cytoscape (Fig. 3b-c). Overall, the number of nodes was 306, the number of edges was 1105, the average node degree was 7.22, the average local clustering coefficient was 0.424, and the PPI enrichment p-value was  $< 1.0e-16$ . For details, module 1 consisted of 30 nodes and 416 edges. For module 2, 12 nodes and 46 edges existed. Furthermore, cytoHubba was applied to screen out hub genes, and the top 20 genes were selected by using the MCC method (Fig. 3d). The top 20 hub genes included BUB1, CDK1, CCNB2, TPX2, KIF11, CDC45, CENPF, DLGAP5, CDCA5, UBE2C, TOP2A, ASPM, MELK, KIF4A, SPAG5, MKI67, CEP55, ESPL1, KIF14, and NEK2.

### **4. Literature retrieval of DEGs in Pubmed**

Literature retrieval of the 20 DEGs screened from the three datasets showed that there were 5 genes having only 1 or 2 research papers published in Pubmed until now, including CDC45, CDCA5, KIF4A, SPAG5, and ESPL1. Although there was only 1 research paper found focusing on the gene SPAG5, the correlation of SPAG5 with the ovarian cancer had been explored deeply in the paper. Thus, we selected the other four genes, CDC45, CDCA5, ESPL1, and KIF4A, as the hub genes in our subsequent research.

### **5. Validation of hub gene expression levels in multiple database**

We determined the transcriptional expression differences of hub genes between EOC tissues and normal tissues in datasets of GEO and GEPIA. As shown in Fig. 4 (GEO) and Fig. 5 (GEPIA), mRNA levels of four hub genes, CDC45, CDCA 5, ESPL1, and KIF4A, were significantly up-regulated in EOC samples compared with normal ovarian tissues.

## **6. Clinical Stage Analysis Of The Hub Genes**

The expression levels of the 4 hub genes at different stages were shown in Fig. 6. According to the result, it was easy to find that there were significant variations in the expression levels of CDC45 [ $\text{Pr}(> F) =$

0.000554], CDCA5 [Pr(> F) = 0.00668], KIF4A [Pr(> F) = 0.0217], and ESPL1 [Pr(> F) = 0.00966]. The overall trends indicated that the expression of these four genes decreased gradually with the continuous progression of OC.

## 7. Survival analysis of DEGs expression in patients with EOCs

To further investigate the prognostic values of DEGs in EOC patients, we conducted a survival assay based on the TCGA data downloaded from the UCSC Xena database. As suggested in Fig. 7(a-d), the relatively higher expression of CDCA5 and ESPL1 was associated with poor prognosis of EOC patients, coherent with higher expression in EOC tissues vs. lower expression in normal ovarian tissues, while the other two genes, CDC45 and KIF4A, had no statistical influence on patients' overall survival.

Furthermore, we also detected whether these genes were related to the progression-free survival of EOC patients, and survival curves illustrated that CDCA5 and ESPL1 notably affected the progression-free survival time of EOC patients (Fig. 7(e-h)). Evidently, patients with a lower level of CDCA5 and ESPL1 had better progression-free survival compared to patients with higher CDCA5 and ESPL1 expression.

From the analysis above, we concluded that CDCA5 and ESPL1 might be closely correlated with EOC overall and progression-free survival, implying the essential roles that these two genes might play in EOC progression.

## Discussion

Despite significant advances in EOC treatment, including surgery, chemotherapy, radiotherapy, and novel targeted agents, EOC had remained an intractable cancer over the past several decades. Therefore, uncovering the etiological and molecular mechanisms underlying EOC is of vital importance for cancer therapy and prevention. For many years, bioinformatics analysis has been playing crucial roles in cancer study, and it facilitates the understanding of carcinogenesis by integrating data at the genome level with systematic bioinformatics methods. Among the multiple bioinformatics strategies, DNA microarray gene expression profiling has been widely applied to explore DEGs involved in tumorigenesis, diagnosis, and therapeutic approaches [9].

In this study, we first screened DEGs from three independent GEO datasets, and implemented GO-KEGG pathways enrichment analysis. A PPI network was constructed in the STRING database and the top 20 hub genes were selected in Cytoscape. We then implemented literature retrieval of the 20 genes in Pubmed. Five genes were found having only one or two research papers published previously, including CDC45, CDCA5, KIF4A, ESPL1, and SPAG5. Although there was only one research paper found focusing on the gene SPAG5, the correlation of this gene with the ovarian cancer had been explored deeply in the paper. Therefore, we focused on the other four hub genes in our subsequent research. The relative expression of the four genes, CDC45, CDCA5, KIF4A, and ESPL1, was detected in Oncomine and GEPIA databases, suggesting that all the four hub genes were up-regulated in EOC tissues with statistical significance. Clinical stage analysis indicated that the expression of these four genes decreased

gradually with the continuous progression of OC. Survival curves illustrated that patients with a lower level of CDCA5 and ESPL1 had better overall survival and progression-free survival compared to patients with higher expression. Therefore, these two hub genes, CDCA5 and ESPL1, could be utilized as potential diagnostic indicators for EOC.

Cell-division cycle-associated 5 (CDCA5), also known as sororin, is thought to play a critical role in ensuring the accurate separation of sister chromatids during the S and G2/M phases of the cell cycle through interactions with cohesin and cdk1 [10, 11]. CDCA5 has also been shown to interact with ERK as well as cyclin E1, a critical regulator of the G1/S mitotic checkpoint [10–12]. Recent studies have correlated the expression of CDCA5 with tumorigenesis and tissue invasion in several cancers.

Regarding lung cancer, several researches confirmed that CDCA5, exhibiting high specificity and sensitivity to distinguish malignant lesions from non-malignant tissues and associated with poor survival, could be identified as predictive biomarkers for tumorigenesis and poor prognosis of lung adenocarcinomas [13, 14]. In study performed by Nguyen et al, suppression of CDCA5 expression inhibited the growth of lung cancer cells; concordantly, induction of exogenous expression of CDCA5 conferred growth-promoting activity in mammalian cells. Their data suggested that transactivation of CDCA5 and its phosphorylation at Ser209 by ERK played an important role in lung cancer proliferation, and that the selective suppression of the ERK-CDCA5 pathway could be a promising strategy for cancer therapy [12].

In researches of hepatocellular carcinoma (HCC), CDCA5 was also found to be up-regulated in HCC cells, and related to poor prognosis [15]. CDCA5 participated the promotion of HCC cells proliferation, migration, and invasion, playing a tumor-promotive role and being a potential therapeutic target for patients with HCC [16, 17]. Besides, CDCA5 was found to be transcribed by E2F1, and could promote oncogenesis by enhancing cell proliferation and inhibiting apoptosis via the AKT pathway in HCC [18]. Another research found that increased CDCA5 expression was associated with increased tumor diameter and microvascular invasion in HCC [19]. Furthermore, silencing of CDCA5 inhibited cell proliferation and induced G2/M cycle arrest *in vitro*, and CDCA5 down-regulation in xenograft model impeded HCC growth *in vivo*. CDCA5 depletion decreased the levels of ERK 1/2 and AKT phosphorylation *in vitro* and *in vivo*. Taken together, these results indicated that CDCA5 might act as a novel prognostic biomarker and therapeutic target for HCC [20].

In addition, it has also been confirmed that CDCA5 was significantly upregulated in breast cancer, bladder cancer, oral squamous cell cancer, urinary tract carcinoma, head and neck squamous cell carcinoma, and esophageal squamous cell carcinoma, and the high expression of CDCA5 was closely related to pathological stages and poor prognosis of patients [21–26].

ESPL1, also known as extra spindle poles-like 1 protein or separin, plays a central role in chromosome segregation by cleaving the cohesin complex at the onset of anaphase [27], and altered ESPL1 activity is correlated with aneuploidy and cancer [28]. At present, the results on the roles of ESPL1 in cancers are conflicting.

ESPL1 expression has been found to be upregulated in a wide range of cancers and high expression of ESPL1 is associated with a loss of key tumor suppressor gene P53, which further contributes to the progression of mammary adenocarcinomas [29, 30]. The research conducted by Finetti et al reinforced that ESPL1 was a candidate oncogene in luminal B breast cancer, and the expression of ESPL1 might represent a promising therapeutic approach for the poor-prognosis tumors [31]. Genomic analysis of transitional cell carcinoma (TCC) by both whole-genome and whole-exome sequencing of 99 individuals with TCC found frequent alterations in ESPL1 [32]. Chen et al found that ESPL1 may be associated with bladder cancer development and recurrence [33]. In addition, Liu et al. identified 7 pivotal genes involved in endometrial cancer prognosis and constructed a prognostic gene signature, among which ESPL1 was one of the genes that were viewed as risky prognostic genes [34]. ESPL1 expression was found to be increased in endometrial cancer tissues, but the clinical significance and functional mechanism of ESPL1 in EC remains to be verified [34]. Nevertheless, it has also been reported that ESPL1 plays an opposite role in gastric adenocarcinoma. ESPL1 expression was negatively correlated with gastric adenocarcinoma pathologic stage progression, and the high expression of ESPL1 was significantly correlated with favorable outcomes [35]. Further work is required to resolve the conflicting roles of ESPL1 in cancer and determine its functions in cancers including the ovarian cancer.

There are several limitations in our study as follows. First, there is an urgent need for biological experiments to validate our results because our research is based on data analysis. Second, we lack the molecular mechanisms for these genes, and we will incorporate these for further exploration. In the future, we will further design experiments (including PCR, Western Blot, immunohistochemistry, etc.) based on specific mechanisms, conduct in-depth research, and improve the inadequacies.

## Conclusions

Our study provided a comprehensive bioinformatics analysis of DEGs, which may have the potential to serve as reliable molecular biomarkers for the diagnosis and prognosis of EOC. Two genes, CDCA5 and ESPL1, were validated to be up-regulated in EOC samples, and high expression of these two genes were related with poor overall survival and progression-free survival. CDCA5 and ESPL1 may act as efficient biomarkers and potential therapy targets in EOC. Further studies are merited to explore the biological functions of these genes and to clarify the underlying molecular mechanisms involved in the pathogenesis of EOC.

## Abbreviations

CRS

cytoreductive surgery; GEO:Gene Expression Omnibus; DEGs:differentially expressed genes; EOC:epithelial ovarian cancer; GO:gene ontology; KEGG:Kyoto encyclopedia of genes and genomes; PPI:protein-protein interactions; STRING:Search Tool for the Retrieval of Interacting Genes; MCODE:Molecular Complex Detection; GEPIA:Gene Expression Profiling Interactive Analysis; TCGA:tool based on the cancer genome atlas.

# Declarations

## Ethics approval and consent to participate

Not applicable.

## Content for publication

Not applicable.

## Availability of data and materials

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

## Competing interests

The authors declare that they have no competing interests.

## Funding

This study was supported by National Natural Science Foundation of China (No. 81402140).

## Authors' contributions

Conception and design: Ting Gui, Keng Shen

Acquisition of data: Ting Gui, Chenhe Yao

Analysis and interpretation of data: Ting Gui, Bingham Jia

Writing, review, and revision of the manuscript: Ting Gui, Keng Shen

## Acknowledgement

We thank the GEO, TCGA, DAVID, KEGG, STRING, GEPIA2, STITCH, Kaplan-Meier plotter databases for providing their platforms and the contributors for their valuable data sets.

## Authors' information

Ting Gui and Keng Shen:

1. Department of Obstetrics and Gynecology, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China.

Chenhe Yao and Bingham Jia:

2. Department of R&D Technology Center, Beijing Zhicheng Biomedical Technology Co, Ltd, Beijing, China.

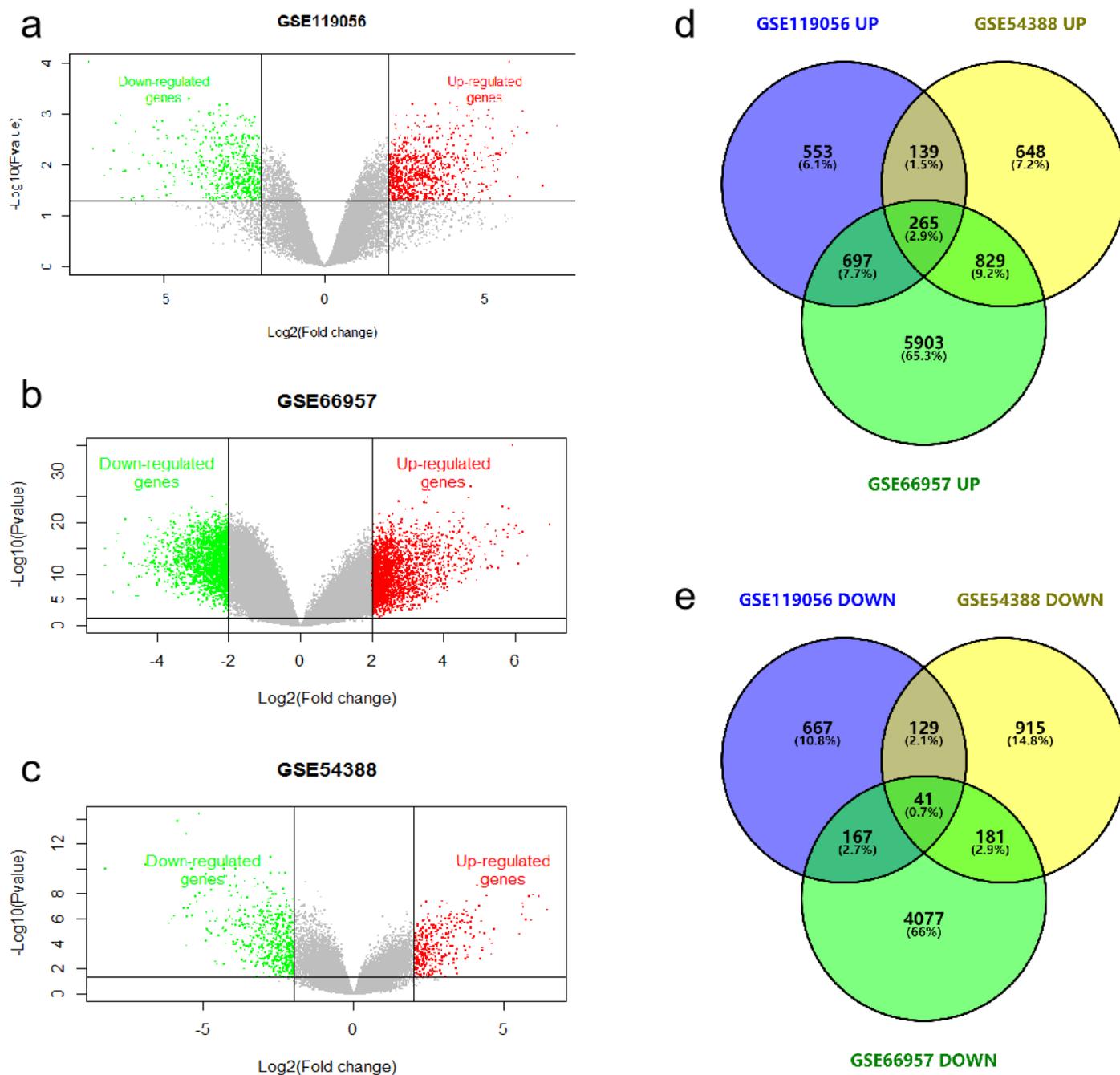
## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424.
2. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin.* 2019;69(5):363–85.
3. Clough E, Barrett T. The gene expression omnibus database. *Methods Mol Biol.* 2016;1418:93–110.
4. Wu J, Mao X, Cai T, Luo J, Wei L. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.* 2006;34:W720-4.
5. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* 2003;31(1):258–61.
6. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research.* 2003;13:2498–504.
7. Li X, Qu J, Yi H, Zhang P, Yi H, Wan X, et al. Integrated analysis of differential miRNA and mRNA expression profiles in human radioresistant and radiosensitive nasopharyngeal carcinoma cells. *PLoS One.* 2014;9(1):e87767.
8. Tang Z, Li C, Kang B, Gao GE, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017;45:W98–102.
9. Fan X, Shi L, Fang H, Cheng Y, Perkins R, Tong W. DNA microarrays are predictive of cancer prognosis: a re-evaluation. *Clin Cancer Res.* 2010;16(2):629–36.
10. Schmitz J, Watrin E, Lenart P, Mechtler K, Peters JM. Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesin in interphase. *Curr Biol.* 2007;17(7):630–6.
11. Borton MT, Rashid MS, Dreier MR, Taylor WR. Multiple levels of regulation of sororin by Cdk1 and Aurora. *J Cell Biochem.* 2016;117(2):351–60.
12. Nguyen MH, Koinuma J, Ueda K, Ito T, Tsuchiya E, Nakamura Y, et al. Phosphorylation and activation of cell division cycle associated 5 by mitogen-activated protein kinase play a crucial role in human lung carcinogenesis. *Cancer Res.* 2010;70(13):5337–47.
13. Yi M, Li T, Qin S, Yu S, Chu Q, Li A, et al. Identifying tumorigenesis and prognosis-related genes of lung adenocarcinoma: based on weighted gene coexpression network analysis. *Biomed Res Int.* 2020; 2020:4169691.
14. Wu Q, Zhang B, Sun Y, Xu R, Hu X, Ren S, et al. Identification of novel biomarkers and candidate small molecule drugs in non-small-cell lung cancer by interated microarray analysis. *Onco Targets Ther.* 2019;12:3545–63.

15. Bai L, Ren Y, Cui T. Overexpression of CDCA5, KIF4A, TPX2, and FOXM1 coregulated cell cycle and promoted hepatocellular carcinoma development. *J Comput Biol* 2019 Oct 9. doi:10.1089/cmb.2019.0254. [Epub ahead of print].
16. Zhang X, Yan Z, Wang L, Zhang S, Gao M. STAT1-induced upregulation of lncRNA RHPN1-AS1 predicts a poor prognosis of hepatocellular carcinoma and contributes to tumor progression via the miR-485/CDCA5 axis. *J Cell Biochem* 2020 Feb 17. doi:10.1002/jcb.29689. [Epub ahead of print].
17. Wan Z, Zhang X, Luo Y, Zhao B. Identification of hepatocellular carcinoma-related potential genes and pathways through bioinformatics-based analyses. *Genet Test Mol Biomarkers*. 2019;23(11):766–77.
18. Chen H, Chen J, Zhao L, Song W, Xuan Z, Chen J, et al. CDCA5, transcribed by E2F1, promotes oncogenesis by enhancing cell proliferation and inhibiting apoptosis via the AKT pathway in hepatocellular carcinoma. *J Cancer*. 2019;10(8):1846–54.
19. Tian Y, Wu J, Chagas C, Du Y, Lyu H, He Y, et al. CDCA5 overexpression is an indicator of poor prognosis in patients with hepatocellular carcinoma (HCC). *BMC Cancer*. 2018;18(1):1187.
20. Wang J, Xia C, Pu M, Dai B, Yang X, Shang R, et al. Silencing of CDCA5 inhibits cancer progression and serves as a prognostic biomarker for hepatocellular carcinoma. *Oncol Rep*. 2018;40(4):1875–84.
21. Chang IW, Lin VC, He HL, Hsu CT, Li CC, Wu WJ, et al. CDCA5 overexpression is an indicator of poor prognosis in patients with urothelial carcinomas of the upper urinary tract and urinary bladder. *Am J Trans Res*. 2015;7(4):710–22.
22. Tokuzen N, Nakashiro K, Tanaka H, Iwamoto K, Hamakawa H. Therapeutic potential of targeting cell division cycle associated 5 for oral squamous cell carcinoma. *Oncotarget*. 2016;7(3):2343–53.
23. Fu G, Xu Z, Chen X, Pan H, Wang Y, Jin B. CDCA5 functions as a tumor promoter in bladder cancer by dysregulating mitochondria-mediated apoptosis, cell cycle regulation and PI3K/AKT/mTOR pathway activation. *J Cancer*. 2020;11(9):2408–20.
24. Qadir F, Lalli A, Dar HH, Hwang S, Aldehlawi H, Ma H, et al. Clinical correlation of opposing molecular signatures in head and neck squamous cell carcinoma. *BMC Cancer*. 2019;19(1):830.
25. Xu J, Zhu C, Yu Y, Wu W, Cao J, Li Z, et al. Systematic cancer-testis gene expression analysis identified CDCA5 as a potential therapeutic target in esophageal squamous cell carcinoma. *EBioMedicine*. 2019;46:54–65.
26. Zhou Q, Ren J, Hou J, Wang G, Ju L, Xiao Y, et al. Co-expression network analysis identified candidate biomarkers in association with progression and prognosis of breast cancer. *J Cancer Res Clin Oncol*. 2019;145(9):2383–96.
27. Solomon DA, Kim JS, Waldman T. Cohesin gene mutations in tumorigenesis: from discovery to clinical significance. *BMB Rep*. 2014;47(6):299–310.
28. Kumar R. Separase: function beyond cohesion cleavage and an emerging oncogene. *J Cell Biochem*. 2017;118(6):1283–99.
29. Wen DY, Lin P, Pang YY, Chen G, He Y, Dang YW, et al. Expression of the long intergenic non-protein coding RNA 665 (LINC00665) gene and the cell cycle in hepatocellular carcinoma using the cancer

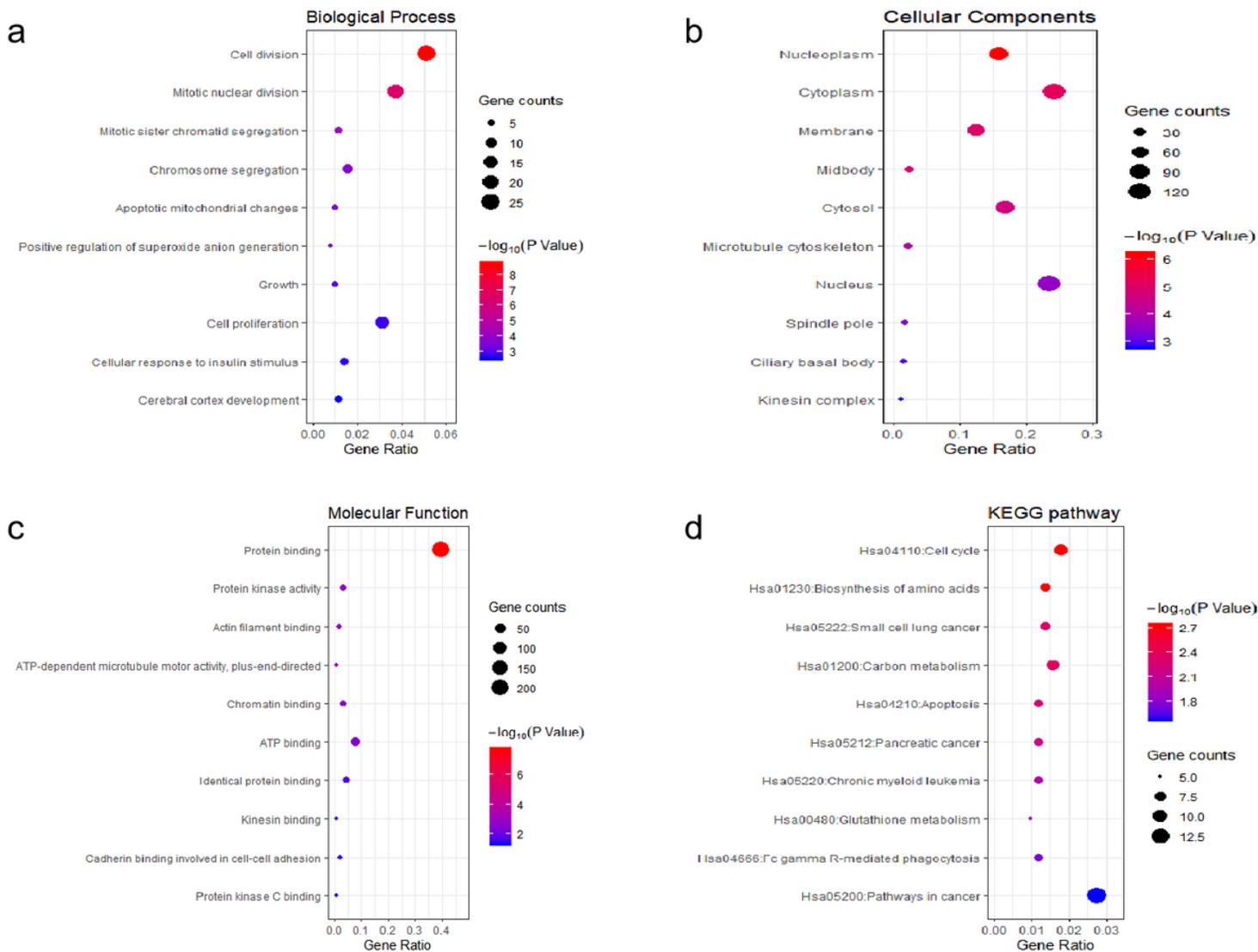
- genome atlas, the gene expression omnibus, and quantitative real-time polymerase chain reaction. *Med Sci Monit.* 2018;24:2786–808.
30. Mukherjee M, Ge G, Zhang N, Edwards DG, Sumazin P, Sharan SK, et al. MMTV-Espl1 transgenic mice develop aneuploid, estrogen receptor alpha (ER $\alpha$ )-positive mammary adenocarcinomas. *Oncogene.* 2014;3348:5511–22.
  31. Finetti P, Guille A, Adelaide J, Birnbaum D, Chaffanet M, Bertucci F. ESPL1 is a candidate oncogene of luminal B breast cancers. *Breast Cancer Res Treat.* 2014;147(1):51–9.
  32. Guo G, Sun X, Chen C, Wu S, Huang P, Li Z, et al. Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nat Genet.* 2013;45(12):1459–63.
  33. Chen Q, Hu J, Deng J, Fu B, Guo J. Bioinformatics analysis identified key molecular changes in bladder cancer development and recurrence. *BioMed Res Int.* 2019; 2019: 3917982.
  34. Liu L, Lin J, He H. Identification of potential crucial genes associated with the pathogenesis and prognosis of endometrial cancer. *Front Genet.* 2019;10:373.
  35. Wang D, Zhu H, Guo M, Fan X, Hu S, Yan K, et al. Expression and prognostic value of cell-cycle-associated genes in gastric adenocarcinoma. *BMC Gastroenterol.* 2018;18(1):81.

## Figures



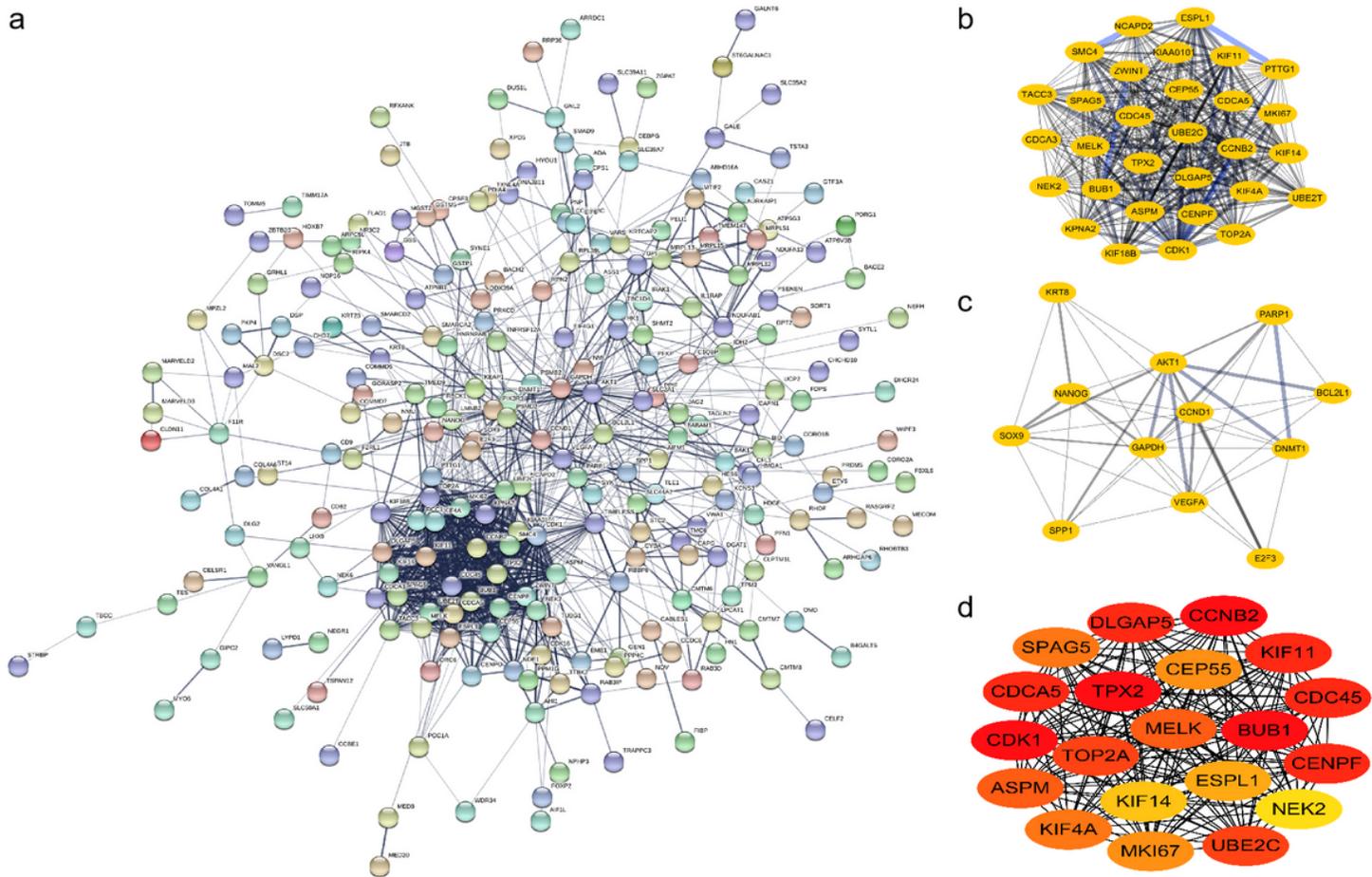
**Figure 1**

Identification of differentially expressed genes (DEGs) in three gene expression omnibus (GEO) datasets. (a) Volcano plot of DEGs in GSE119056. (b) Volcano plot of DEGs in GSE54388. (c) Volcano plot of DEGs in GSE66957. Red, blue and gray color represent relatively high, low and equal expression of genes in the corresponding group, respectively. (d) Venn diagram of 265 overlapping up-regulated genes from intersection of three independent GEO datasets.  $p < 0.05$  and  $|\log_2 \text{FC}| > 1$  were set as the threshold. (e) Venn diagram of 41 overlapping down-regulated genes from intersection of three independent GEO datasets.  $p < 0.05$  and  $|\log_2 \text{FC}| > 2$  were set as the threshold.



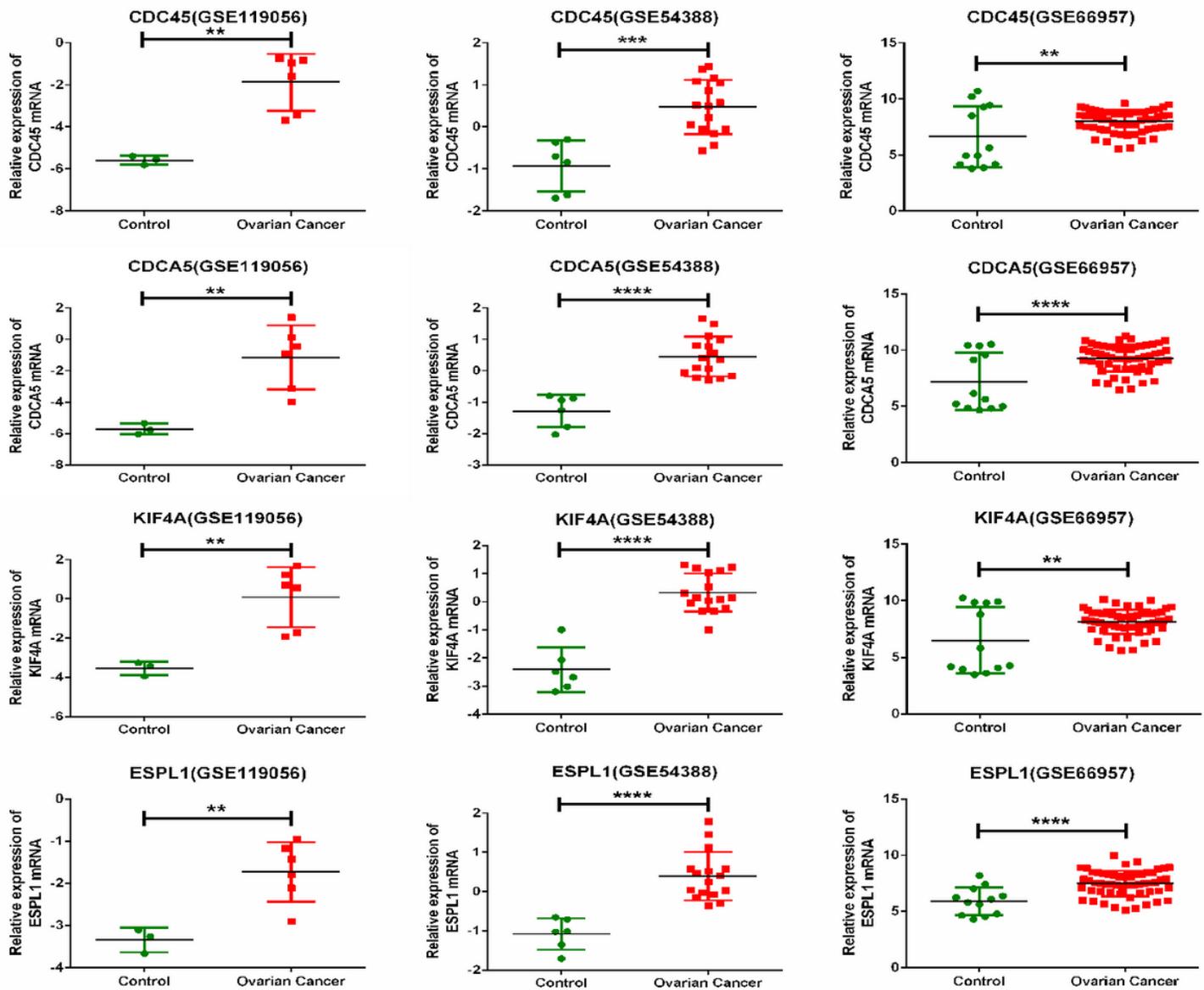
**Figure 2**

GO analysis and KEGG pathway analysis of the overlapping DEGs. (a) Top 10 of biological process. (b) Top 10 of cellular component. (c) Top 10 of molecular function. (d) Top 10 of KEGG pathway enrichment.



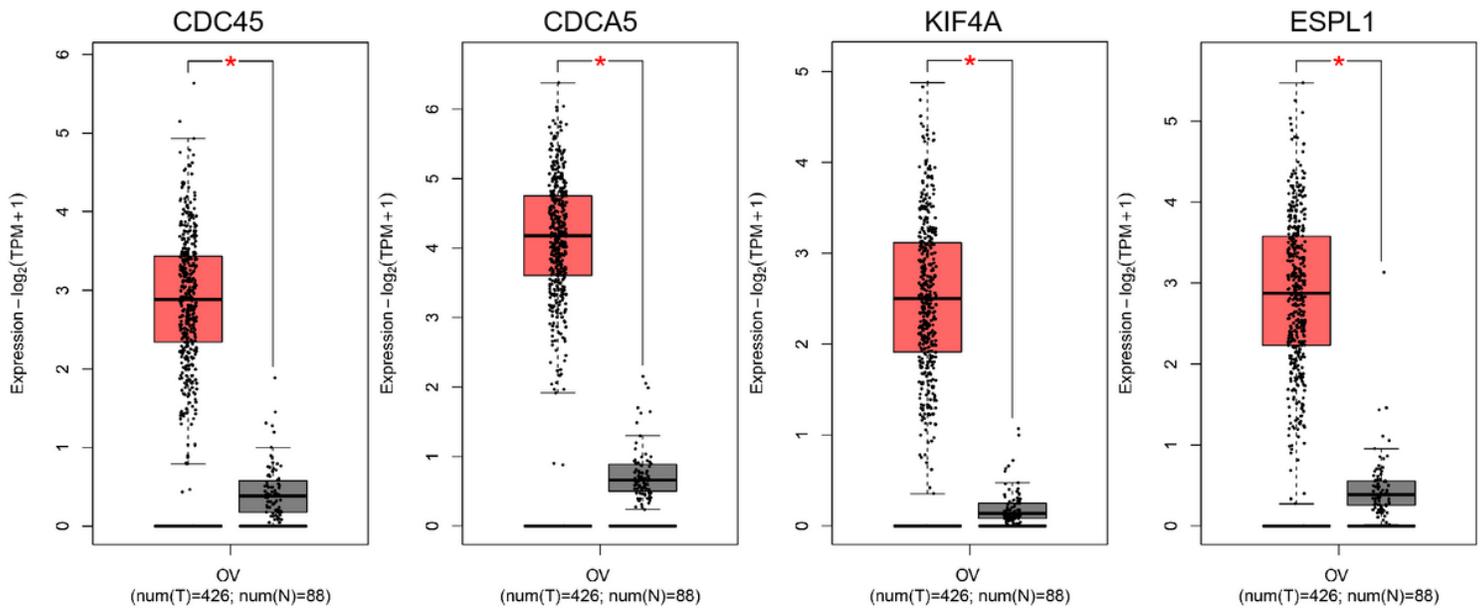
**Figure 3**

Protein-protein interactions (PPI) network, module analysis, and hub gene identification. (a) PPI network of overlapping differentially expressed genes was constructed in STRING database. (b-c) Top two modules screened using Molecular Complex Detection (MCODE) in Cytoscape software. (d) Top twenty hub genes selected by the CytoHubba in Cytoscape based on the degree of each protein node.



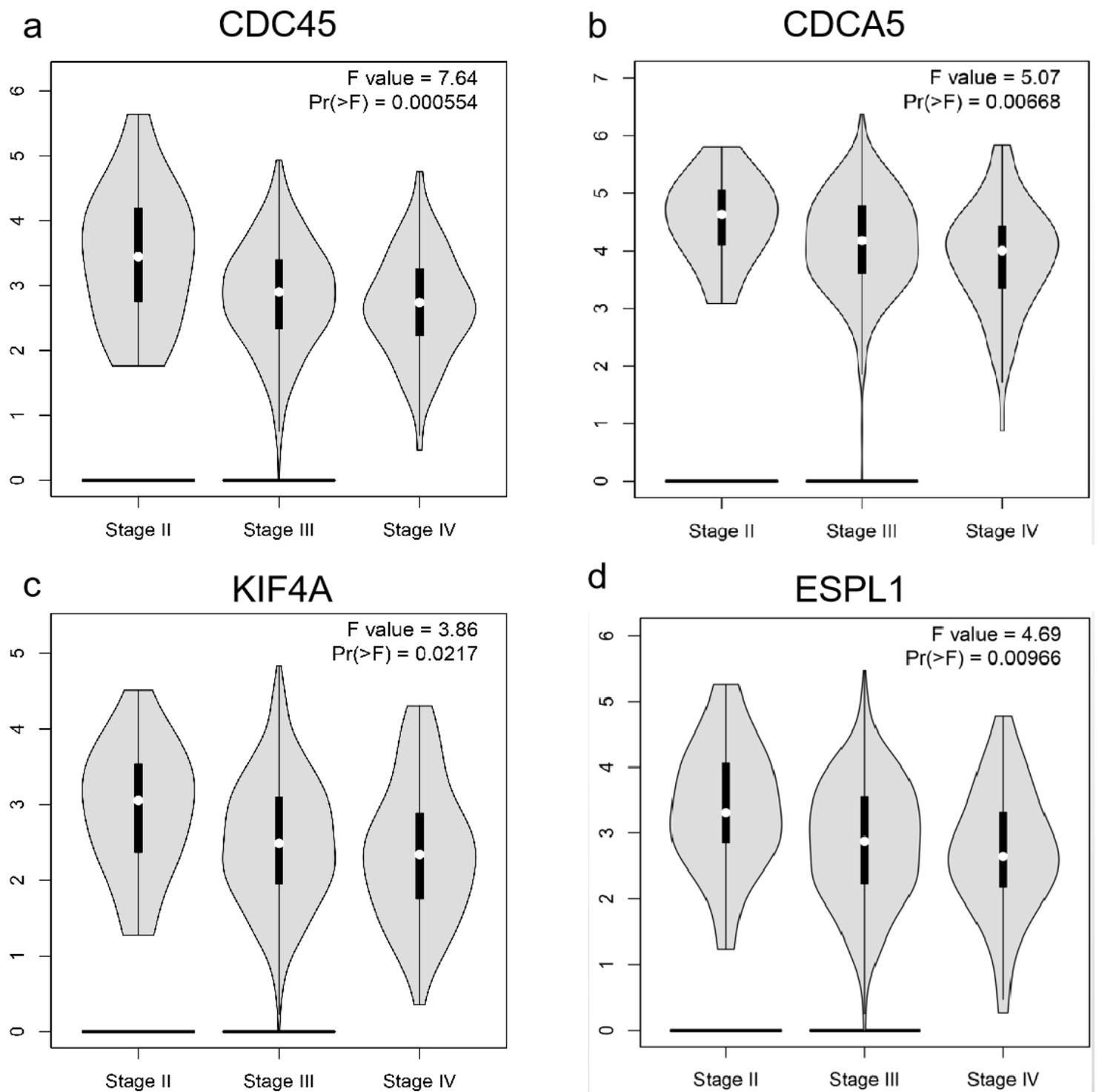
**Figure 4**

Validation of hub gene expression levels in GEO The expression of CDC45, CDCA5, KIF4A, and ESPL1 in EOC tissues were significantly elevated compared with adjacent ovarian tissues. (a) CDC45; (b) CDCA5; (c) KIF4A; (d) ESPL1; \* $p < 0.05$ .



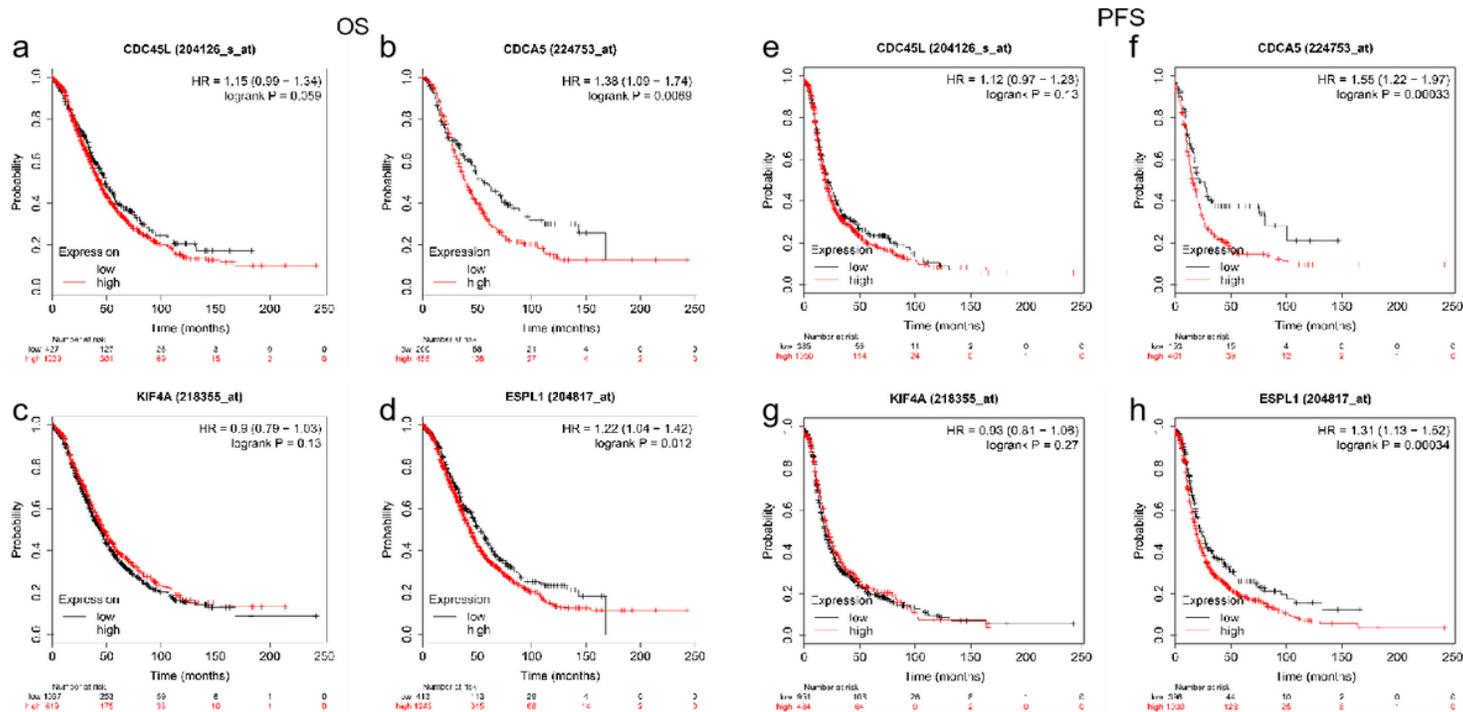
**Figure 5**

Validation of hub gene expression levels in GEPIA The expression of CDC45, CDCA5, KIF4A, and ESPL1 in EOC tissues were significantly elevated compared with adjacent ovarian tissues. (a) CDC45; (b) CDCA5; (c) KIF4A; (d) ESPL1; \*p < 0.05.



**Figure 6**

The expression level of hub genes in EOC tissues at different stages. To further verify the expression level of the hub genes in EOC tissues at different stages, the hub genes were analyzed by the GEPIA2 online database. ANOVA was performed to assess the statistical significance of the variations.  $Pr(>F) < 0.05$  was considered statistically significant. According to the result, there were significant variations in the expression levels of CDC45, CDCA5, KIF4A, and ESPL1 (a-d). The overall trends indicated that the expression of these four genes decreased gradually with the continuous progression of OC.



**Figure 7**

Overall survival and progression-free survival curves of DEGs in patients with EOCs from TCGA database. (a) The relatively higher expression of CDC45 had no statistical influence on patients' overall survival. (b) The relatively higher expression of CDCA5 was associated with poor prognosis of EOC patients. (c) The relatively higher expression of KIF4A had no statistical influence on patients' overall survival. (d) The relatively higher expression of ESPL1 was associated with poor prognosis of EOC patients. (e) The relatively higher expression of CDC45 had no statistical influence on patients' progression-free survival. (f) The relatively higher expression of CDCA5 was associated with poor prognosis of EOC patients. (g) The relatively higher expression of KIF4A had no statistical influence on patients' progression-free survival. (h) The relatively higher expression of ESPL1 was associated with poor prognosis of EOC patients.