

Identification of Candidate Biomarkers Associated with Ovarian Cancer Progression and Prognosis by Integrated Microarray Analysis

Chen Xiaohong (✉ chenxh19@lzu.edu.cn)

Lanzhou Medical College: Lanzhou University <https://orcid.org/0000-0001-7810-8271>

Xin Jiachun

Lanzhou Medical College: Lanzhou University

Feng Zhijun

Lanzhou Medical College: Lanzhou University

Cai Hui

Gansu Provincial Hospital

Ma Shouye

Gansu Provincial Hospital

Wu Huifang

Gansu Provincial Hospital

Research Article

Keywords: ovarian cancer, bioinformatic analysis, differentially expressed genes, gene cluster, prognostic biomarkers

Posted Date: May 7th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-441360/v1>

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

[Read Full License](#)

Abstract

Ovarian cancer (OC) is one of the most common gynecological malignancies, with high mortality and few prognostic biomarkers. This study aimed to identify differentially expressed genes (DEGs) relevant to the prognosis of ovarian cancer. First, we performed differential analysis on GSE38666, GSE14407, GSE27651 and GSE18520 expression profile datasets from the Gene Expression Omnibus (GEO) database and obtained 638 DEGs using a Venn diagram tool. Second, we constructed a protein–protein interaction (PPI) network with the STRING database and used the Molecular Complex Detection (MCODE) plugin in Cytoscape software to perform gene cluster analyses for 269 upregulated and 367 downregulated DEGs, respectively. Based on the results of gene cluster analyses, we selected two gene clusters of upregulated DEGs with network scores of > 10 and two gene clusters of downregulated DEGs with scores of > 5 for further functional analysis with the clusterProfiler package. In addition, we analyzed the effect of genes from those clusters on the overall survival (OS) and progression-free survival (PFS) of OC patients based on Kaplan–Meier Plotter and PrognoScan databases. The gene clusters of upregulated DEGs are involved mainly in cell division and cell proliferation, and the genes of downregulated DEGs primarily participate in the TGF- β signaling pathway and various metabolic processes. Via cross-validation of survival analysis, we identified four hub genes (AURKA, CDCA5, CEP55 and UBE2C) that correlated with ovarian cancer OS and PFS, which derived from the upregulated DEGs. Patients with these four hub genes trended toward worse OS and PFS than patients with low expression. Unfortunately, no significantly prognosis-related genes were detected among downregulated clusters. The mRNA expression of these four genes were analysed using the Oncomine and GEPIA databases, and the result revealed that the mRNA expression of AURKA, CDCA5, CEP55 and UBE2C was greatly upregulated in OC tissues compare with normal tissues. The gene set/pathway primarily showed alterations in amplification and deep deletion, and correlation analysis indicated significant positive correlations between AURKA, CDCA5, CEP55 and UBE2C. Nevertheless, for clarifying the protein expression levels of the four prognostic genes, the staining of the immunohistochemistry (IHC) were obtained base on the Human Protein Atlas, and was showed the proteins levels of these four genes was significantly elevated in patients with OC. Our results revealed that these four hub genes may mediate ovarian cancer development and have valuable clinical potential as prognostic biomarkers for OC patients.

Introduction

Ovarian cancer (OC) is one of the most common gynecological malignancies, with high mortality and few prognostic biomarkers [1] [2]. These characteristics are due partially to the lack of observable symptoms, and no specific screening method is available for the early detection of OC. Cytoreductive surgery in combination with neoadjuvant chemotherapy is the primary treatment for ovarian cancer. The rapid development of targeted drugs has provided more opportunities for the treatment of ovarian cancer patients, but the overall survival of patients has not improved significantly [3, 4]. Therefore, it is necessary to explore potential genetic changes and molecular mechanisms involved in the occurrence and development of OC, which can help us identify more specific biomarkers for diagnosis and assessing

prognosis. Recently, interest has increased in applying gene expression profiling to help understand the pathogenesis of OC, develop biomarkers for early detection, and elucidate drug resistance mechanisms in OC [5]. Connor et al [6] found that Thy-1 was upregulated in OC and correlated obviously with poor prognosis. Han et al [7] reported that SOX30 can be used as a prognostic biomarker and chemoresistance indicator in OC. However, these studies focused mainly on single genes to expose the correlation between gene expression and OC, and determining whether differentially expressed genes (DEGs) have predictive and prognostic roles in OC remains difficult.

In the present study, we screened four gene expression datasets to identify the DEGs and to evaluate the associations between DEGs and the prognosis of ovarian cancer patients.

Through the analysis, we screened the four hub genes of AURKA, CDCA5, CEP55 and UBE2C can be used as effective biomarkers for diagnosis and prognosis assessment in OC patients.

Results

Preprocessing of microarray datasets and identification of DEGs

The four gene datasets contained a total of 48 normal ovarian tissues and 125 ovarian cancer tissues (**Table S1**), and differential gene analysis data were normalized (**Figure S1**). We initially identified 5,870, 4,040, 4,532 and 18,520 DEGs from GSE27651, GSE38666, GSE14407 and GSE18520, respectively, with the limma package of Rtools and the criteria $|\log_{2}FC| > 1$ and adjusted P value < 0.01 (**Figure 1A-D**). Then, all DEGs were divided into up- and down-regulated genes according to whether $|\log_{2}FC| > 0$ or $|\log_{2}FC| < 0$, respectively. A Venn diagram tool was applied to identify the DEGs in the sets of up- and downregulated DEGs. A total of 636 DEGs, including 269 upregulated and 367 downregulated DEGs, were identified in the OC samples compared to the normal ovarian tissue samples (**Figure 1E&F**). All of these DEGs are listed in **Table S2**.

PPI network construction and MCODE analysis

The PPI network of the 269 upregulated DEGs was constructed with that demonstrated 256 nodes and 773 edges were enriched 6 gene clusters (**Figure S2A, Table S3**). As well as, the PPI network of the downregulated DEGs contained 330 nodes and 94 edges and enriched 7 gene clusters using the MCODE plugin of Cytoscape software (**Figure S2B, Table S3**). Next, two gene clusters with network scores of > 10 were selected from the upregulated DEGs (**Figure 2A&B**), and two gene clusters with scores of > 5 from the downregulated gene clusters (**Figure 2C&D**) were selected as the most important molecular modules in the respective results and further analyzed for function and prognosis.

GO enrichment analysis showed that the two gene clusters of upregulated DEGs were primarily related to the biological processes (BP) of nuclear division and sister chromatid segregation in the cell cycle. The primary molecular functions included microtubule binding and various kinase regulator activities (**Figure**

3A&B). KEGG pathway analysis indicated genes enriched mainly in oocyte maturation, cell cycle, p53 signaling pathway, and cellular senescence pathway (**Figures 4A-F**).

The two gene clusters of downregulated DEGs in GO enrichment analysis indicated genes involved mainly in the endoplasmic reticulum lumen and mitochondrial outer membrane as cellular components, protein modification and regulation of various signal paths as biological processes, and hormone activity and alcohol dehydrogenase [NAD(P)+] activity as molecular functions (**Figure 5A&B**). No signaling pathways were significantly enriched for the first gene cluster ($P < 0.05$). The second gene cluster was mainly involved in many metabolic functions, such as tyrosine metabolism, fatty acid degradation and retinol metabolism (**Figure 6A-C**).

Survival analysis for identification of prognostic genes

The OS and PFS of patients were stratified by the expression of genes in the gene clusters of up- and down-regulated DEGs using the Kaplan–Meier Plotter database. Twenty-six DEGs from upregulated DEGs were identified to correlate with OS and PFS in ovarian cancer. The results were visualized through the forestplot package of Rtools (**Figures 7A-D**). However, only CHGB and MAOB genes from downregulated DEGs were relevant to prognosis in OC, and a higher expression of these genes predicted better OS and PFS (**Table S4**).

To further examine the prognostic potential of the twenty-eight genes from the up- and downregulated DEGs, we continued to reanalyze the survival differences through the PrognosScan database. Interestingly, two cohorts (GSE9891 and GSE17260) [8, 9], which included 278 samples and 110 samples at different stages of OC, showed that higher expression of AURKA, CDCA5, CEP55 and UBE2C was significantly associated with poorer prognosis (**Figures 8A-H**). Therefore, it is conceivable that high AURKA CDCA5, CEP55 and UBE2C expression is an independent risk factor and leads to a poor prognosis in OC patients.

Analyses the differential mRNA and protein expression of prognostic genes

We compared the transcriptional levels of these four genes in cancers with those in normal tissue samples using the ONCOMINE database (**Figure 9**). ONCOMINE analysis revealed that the mRNA expression of AURKA, CDCA5, CEP55 and UBE2C was upregulated in patients with OC (**Table 1**). The transcription levels of AURKA were significantly higher in patients with OC in four datasets [10-12]. In Yoshihara's dataset [10], AURKA was overexpressed in ovarian serous adenocarcinoma compared to normal samples, with a fold change of 14.459 and P-value of $5.17E-10$. In TCGA, AURKA was overexpressed in ovarian serous cystadenocarcinoma compared to normal samples, with a fold change of 6.504 and P-value of $6.53E-8$. In TCGA, Adib's and the Lu Ovarian dataset [11, 12], AURKA expression was overexpressed in ovarian serous cystadenocarcinoma compared to normal samples, with a fold change of 6.504 (P-value = $6.53E-8$), 2.287 (P-value = $4.20E-4$) and 1.506 (P-value = 0.002), respectively. The transcription levels of CDCA5 were significantly higher in serous cystadenocarcinoma than in normal tissue [10]. The Lu Ovarian dataset showed that CEP55 expression was significantly higher in different subtypes of OC [12], including serous adenocarcinoma, endometrioid adenocarcinoma and clear cell

adenocarcinoma, with fold changes of 2.457 (P-value = 9.44E-7), 2.033 (P-value = 1.45E-4) and 1.787 (P-value = 0.007), respectively. In Yoshihara's dataset [10], Lu's dataset [12] and TCGA, UBE2C was significantly overexpressed in ovarian serous adenocarcinoma, with fold changes of 12.955 (P-value = 5.73E-13), 2.358 (P-value = 1.39E-7) and 10.184 (P-value = 2.24E-7), respectively. UBE2C expression in ovarian carcinoma was significantly increased compared with that in the normal samples in Bonome's datasets [13].

Using the GEPIA dataset, we further confirmed the differential mRNA expression of the four genes between OC and normal tissues. The results showed that mRNA expression levels of AURKA, CDCA5, CEP55 and UBE2C were significantly increased in OC tissues compared with normal ovarian tissues [(tumor sample: n = 426 vs. normal sample: n = 88) (**Figure 10**). Additionally, expression of the four genes was analyzed by ovarian cancer stage. However, in contrast to CDCA5, mRNA expression of AURKA, CEP55 and UBE2C was not significantly associated with FIGO stage of OC (**Figure S3**). The IHC staining and images were downloaded from the Human Protein Atlas, the proteins levels of these four prognostic genes was significantly elevated in tumor tissues compared with normal tissues (**Figure 11**).

Gene alteration and correlation analyses of prognostic genes

Mutations in AURKA, CDCA5, CEP55 and UBE2C retrieved in 1,680 cases from three TCGA datasets of ovarian serous carcinoma (606 cases from TCGA, Firehose Legacy; 585 cases from TCGA, PanCancer Atlas; and 489 cases from TCGA, Nature 2011) were analyzed using the cBioPortal database. Among the 3 OC datasets analyzed, the alterations of the four genes were mainly related to amplification and deep deletion, and ranges from 17.67% of 583 genes to 7.77% of 489 genes were identified for the gene sets submitted for analysis (**Figure 12A**). The percentages of AURKA, CDCA5, CEP55 and UBE2C gene alterations in OC were 6%, 2.1%, 1.3% and 4% (**Figure 12B**), respectively. Kaplan–Meier Plotter and log-rank test results indicated no significant difference in OS and PFS between the cases with alterations in one of the queried genes and those without alterations in any of the queried genes (P-values of 0.0851 and 0.213, respectively; **Figure 12 C&D**).

The GeneMANIA database was used for correlation analysis of AURKA, CDCA5, CEP55 and UBE2C at the gene level (**Figure 13A**). The 4 central nodes were surrounded by 20 nodes representing genes that were closely related to the family in terms of physical interactions, co-expression, predictions, co-localization, and genetic interactions. The results revealed relationships in physical interactions, co-expression and co-localization between AURKA, CDCA5, CEP55 and UBE2C. The same pathway was shared between AURKA and UBE2C. The predicted protein domains was shared among AURKA, CEP55 and UBE2C. However, no relationship in shared protein domains or genetic interactions were noted among the four genes. Further functional analysis revealed that these four genes mainly correlated with regulation of cell division, including mitotic cell cycle, mitotic cytokinesis, mitotic sister chromatid segregation mitosis, nuclear division and cell cycle G2/M phase transition (**Table S5**).

The Spearman correlations between AURKA, CDCA5, CEP55 and UBE2C in OC were determined by online analyses using the cBioPortal database (TCGA, PanCancer Atlas (**Figure 13B**)). Spearman's correlation

coefficient exceeding 0.30 indicated a good correlation. The results indicated a significant positive correlation between AURKA and CDCA5 ($R^2 = 0.674$, $P = 4.86e-41$), CEP55 ($R^2 = 0.564$, $P = 1.32e-26$), and UBE2C ($R^2 = 0.78$, $P = 7.36e-62$); CDCA5 and CEP- 55 ($R^2 = 0.616$, $P = 1.13e-32$) and UBE2C ($R^2 = 0.658$, $P = 1.30e-38$); and UBE2C and CEP55 ($R^2 = 0.512$, $P = 2.08e-21$) (**Figure S4**).

Immune infiltrate analysis of prognostic genes

Tumor inflammatory cell infiltration levels play an important role in cancer progression and patient survival. We analyzed whether AURKA, CDCA5, CEP55 and UBE2C expression correlated with inflammatory cell infiltration levels in OC. The results showed that the mRNA levels of the four genes correlated weakly with inflammatory cell infiltration (**Figure 14**). AURKA expression was related to macrophages cells, CD4+T cells and neutrophils, but no significant correlation with tumor purity or infiltrating levels of B cells, CD8+T cells and dendritic cells was noted. In addition to the negative correlation with CD8+T cells, CDCA5 expression was not significantly correlated with other immune inflammatory cells and tumor purity. CEP55 was negatively related to CD8+T cells and positively related to CD4+T cells. UBE2C was only associated with tumor purity and CD4+T cells. The results suggest that these four genes did not correlate closely with tumor immune inflammatory cell and tumor purity associated with the occurrence and development of OC. The detailed mechanisms require further study.

Discussion

OC is a gynecological malignancy with high mortality that is partially attributed to tumor heterogeneity, and its pathogenesis remains unclear. Recently, interest has been increasing in applying gene expression profiling to help understand the pathogenesis of OC, develop biomarkers for early detection, and elucidate drug resistance mechanisms in OC [14]. In the present study, we attempted to systematically explore the prognostic values, expression patterns, genetic alteration, correlation, and potential functions of different DEGs in OC. In total, 636 DEGs, including 269 upregulated and 367 downregulated DEGs, were identified between normal ovary tissue and OC tissue by screening 4 gene datasets. Then, based on the status of gene cluster analyses, we selected two gene clusters of upregulated DEGs with network scores of > 10 and two gene clusters of downregulated DEGs with scores of > 5 for functional and survival analyses. The gene clusters of upregulated DEGs were mainly involved in cell division and cell proliferation, and the genes of downregulated DEGs primarily participated in the TGF- β signaling pathway and various metabolic processes.

We focused on two gene clusters of up- and downregulated DEGs on prognosis analysis. The results showed the mRNA expression of AURKA, CDCA5, CEP55 and UBE2C was significantly increased and correlated with poorer prognosis in OC tissues compared with normal ovarian tissues through cross-validation. Additionally, the four gene sets/pathways primarily showed alterations in amplification and deep deletion, and the percentages of individual AURKA, CDCA5, CEP55 and UBE2C mRNA genes' alterations in OC were 6%, 2.1%, 1.3% and 4%, respectively. Correlation analysis indicated significant positive correlations between AURKA, CDCA5, CEP55 and UBE2C. Further functional analysis revealed

that these four genes mainly take part in regulation of cell division, including mitotic cell cycle, mitotic sister chromatid segregation mitosis and cell cycle G2/M phase transition, indicating that these genes play a vital role in the occurrence and development of OC.

Aurora kinase A (AURKA) is a member of a family of three serine-threonine kinases that are highly conserved throughout evolution and play important roles in cell cycle regulation [15–17]. AURKA has pleiotropic roles in centrosome maturation, spindle assembly, cytokinesis and kinetochores required for proper mitotic progression [15, 16]. AURKA expression levels and kinase activity are upregulated and relevant to poorer prognosis in many human cancers [18–20], indicating that AURKA might serve as a potential biomarker and target in the development of anticancer drugs. AURKA mediates RPS6KB1 phosphorylation at T389 to promote gastrointestinal cancer cell proliferation and survival upon inhibition of KRAS [21]. Knockdown of AURKA significantly inhibited OC cell-induced angiogenesis of endothelial cells compared to controls and also significantly inhibited cell proliferation, migration, and invasion of OC cells in vitro and vivo [22]. We found that increased AURKA RNA expression was closely associated with worse OS and PFS in OC patients, and the underlying mechanisms need to be further elucidated.

Cell-division cycle-associated 5 (CDCA5), also known as sororin, is thought to play a critical role in embryonic development, establishment or maintenance of cohesion and ensuring the accurate separation of sister chromatids during the S and G2/M phases of the cell cycle through interactions with cohesin and CDK1 [23, 24]. CDCA5 functions at the central region of the synaptonemal complex during mitosis in collaboration with SGO2-PP2A [25]. Concurrently, studies have indicated that the expression of CDCA5 correlates with tumorigenesis and development in several cancers, including nonsmall cell lung cancer, colorectal cancer, and esophageal squamous cell cancer [26–28]. However, the gene has not been reported in ovarian cancer and deserves further study.

Centrosomal protein 55 (CEP55) is dispensable for mouse embryonic development and adult tissue homeostasis [29]. CEP55 mediates the physical separation of daughter cells and completes cell division through binding GPPX3Y motifs of ALIX and TSG101 to block similar GPPX3Y motifs of ALIX and TSG101 from interacting and localizing to the midbody [30]. Recently, CEP55 has been studied in various tumor cancers, including those of the colon, liver, breast and lung. High levels of CEP55 improved migration and invasion and were closely related to poor prognosis in multiple tumor types [31, 32]. As well as its roles as a biomarker and actuator of tumorigenesis [33, 34], how CEP55 contributes to OC development remains unclear and it is a valuable avenue to explore in future studies.

Ubiquitin-conjugating enzyme 2C (UBE2C) is an integral component of the ubiquitin proteasome system and participates in destruction of the spindle assembly checkpoint and mitotic exit. UBE2C consists of a conserved core domain containing the catalytic Cys residue and an N-terminal extension. The UBE2C N-terminal extension regulates E3 enzyme activity as a part of an intrinsic inhibitory mechanism [35]. UBE2C mRNA and/or protein levels are aberrantly increased and correlated with poor clinical outcomes in many cancer types; UBE2C is considered a potential cancer biomarker [35–37]. Increased UBE2C expression was noted in oral squamous cell carcinoma (OSCC) compared to control cells and associated

with poor cell differentiation and lymph node invasion. Moreover, suppression of UBE2C expression decreased cell proliferation, migration/invasion, and colony formation and reduced expression of the cancer stem cell markers ALDH1/A2, CD44, CD166 and EpCAM [38]. In patients with HR+/HER2- breast cancer, high UBE2C expression was associated with significantly inferior survival with pN0 and pN1 tumors but not pN2/N3 tumors ($P < 0.05$). UBE2C depletion markedly inhibited cell proliferation and increased the cytotoxicity of tamoxifen by inducing apoptosis [39]. This study also revealed that UBE2C overexpression correlates with poor clinical outcomes in tumor tissue. However, few reports have addressed the correlation of UBE2C with these characteristics in OC, and the underlying mechanism of UBE2C in ovarian tumors requires further exploration.

Our analysis showed that high expression of AURKA, CDCA5, CEP55 and UBE2C genes correlates significantly with poor prognosis in OC patients. These four genes represent promising novel candidates and potential prognostic biomarkers for OC. However, the present study had certain limitations because the main sources of our clinical information were datasets from the GEO and TCGA databases, which are retrospective data. Therefore, the potential mechanisms of these genes should be elucidated further using other external datasets with complete clinical information and gene expression information to better understand the role these genes play in the pathogenesis and progression of OC. Moreover, the protein expression levels of the prognosis-related DEGs and their potential mechanisms in OC must be validated in further experimental studies.

Materials And Methods

Downloaded datasets and identification of differentially expressed genes (DEGs)

Four datasets associated with OC were identified and downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>): GSE38666 [40], GSE27651 [41], GSE14407 [42] and GSE18520 [43]. These datasets contain a total of 48 normal ovarian (OV) tissues and 125 OC tissues. The original GEO data were subjected to background correction and normalization [44] and then converted into expression measures using R software [45]. The DEGs were screened in each dataset with the limma package [46] in R software, and the selection criteria were defined as $|\log \text{fold change (FC)}| > 1$ and a P value of < 0.05 . The DEGs across the four datasets were identified by using a Venn diagram tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

Protein–protein Interaction (PPI) network construction and molecular complex detection (MCODE) analysis

The up- and down-regulated DEGs were used separately to construct a PPI using the STRING (<https://string-db.org/>) database with the following specifications: (1) human species; (2) meaning of network edges was set as evidence with a minimum required interaction score of 0.4; (3) active interaction sources of text mining, experiments, and databases; and (4) hiding of disconnected nodes in the network. Then, we downloaded the PPI data and further explored gene cluster status for up- and down-regulated DEGs using the MCODE and cytoHubba plugins in Cytoscape software (version 3.7.1,

<https://cytoscape.org/>), respectively. MCODE data were filtered according to the default parameters as follows: “Degree Cutoff = 2,” “Node Score Cutoff = 0.2,” “K-Core = 2” and “Max. Depth = 100” [47]. Based on the status of gene cluster analyses, we selected the top 2 gene clusters as the most important molecular modules in the respective results and further Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichments with the clusterProfiler package [48], and the results were visualized with the GOplot package [49]. An adjusted P-value <0.05 was established as a selection criterion.

Survival analysis for identification of prognostic genes

We investigated the effects of genes from the most important molecular modules on the overall survival (OS) and progression-free survival (PFS) of OC patients from the Kaplan–Meier (KM)Plotter database, and the results were visualized with the forestplot package of R [50]. The Kaplan–Meier Plotter database is an online database containing microarray gene expression data and survival information derived from GEO, EGA and The Cancer Genome Atlas (TCGA), which can be used to assess survival in 21 cancer types. The OS and PFS of patients with OC were determined by dividing the patient samples into two groups based on median expression (high vs. low expression) and evaluated with a hazard ratio with 95% confidence intervals and a log rank P-value.

The PrognoScan database [51] was used to cross-validate the prognosis-related genes that were obtained in KM Plotter database. The PrognoScan database is a large collection of publicly available cancer microarray datasets with clinical annotation as well as a tool for assessing the biological relationship between gene expression and prognosis, including OS and PFS. The threshold was adjusted to a Cox P-value <0.05.

Analyses the differential mRNA and protein expression of prognostic genes

To identify prognostic genes that play an important role in OC, we used the ONCOMINE database [52, 53] to analyze the mRNA levels of DEGs relevant to prognosis for OC patients. The gene expression array datasets of ONCOMINE (www.oncomine.org) compose a publicly accessible, online cancer microarray database that helps facilitate research from genome-wide expression analyses. The thresholds were restricted as follows: P-value = 0.001, fold-change = 1.5, a top 10% gene ranking and mRNA data.

To further determine the differential expression of prognostic genes, tumor and normal tissues were analyzed using the GEPIA database [54], which is an interactive web server that estimates mRNA expression data based on 9,736 tumors and 8,587 normal samples in TCGA and Genotype-tissue Expression (GTEx) dataset projects.

In addition, to clarify the protein expression levels of the hub genes, we downloaded the immunohistochemistry (IHC) staining and the IHC images from the Human Protein Atlas (HPA, <http://www.proteinatlas.org/>) [55] and explored whether there were statistically significant differences in the expression of hub genes between the normal and OC tissues, and showed some IHC images.

Alteration and correlation analyses of prognostic genes

The cBioPortal for Cancer Genomics (<http://www.cbioportal.org/>) is an open-source platform that provides information regarding the integrative analysis of complex cancer genomics and clinical profiles from more than 200 cancer studies in The Cancer Genome Atlas (TCGA)[56]. The cBioPortal contains an extensive variety of data, including DNA copy numbers, DNA methylation, mRNA and microRNA expression levels, and nonsynonymous mutations. In this study, the cBioPortal database was used to analyze genetic variations in prognostic genes. In addition, the database was used to evaluate correlations between prognostic genes.

GeneMANIA (<http://www.genemania.org>) provides a flexible web interface for deriving hypotheses based on gene functions [57]. GeneMANIA generates a list of genes with similar functions to the query gene and constructs an interactive functional-correlation network to illustrate relationships between genes and datasets. In this study, the database was utilized to construct a gene–gene interaction network for prognostic genes in terms of physical interactions, co-expression, predictions, co-localization, and genetic interaction as well as to evaluate their functions.

Immune infiltrate analysis of prognostic genes

TIMER is a comprehensive resource for systematic analysis of immune infiltrates across diverse cancer types [58, 59]. TIMER applies a previously published deconvolution statistical method to infer the abundance of tumor-infiltrating immune cells (TIICs) from gene expression profiles. The TIMER database includes 10,897 samples across 32 cancer types from TCGA to estimate the abundance of immune infiltrates. Correlations between the four prognostic gene expression and the infiltration levels of immune inflammatory cells, including B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, dendritic cells and tumor purity, were analyzed using the TIMER database [60].

Statistical analysis

Survival curves were constructed using the PrognosScan and Kaplan–Meier Plotter databases. The results generated in ONCOMINE are displayed with P-values, fold changes, and ranks. The results of Kaplan–Meier plots, PrognosScan, and GEPIA are shown with HR and P or Cox P-values from a log-rank test. The correlation of gene expression was evaluated by Spearman’s correlation and statistical significance. The strength of the correlation was determined using the following guide for the absolute value: 0.00–0.19, very weak; 0.20–0.39, weak; 0.40–0.59, moderate; 0.60–0.79, strong; 0.80–1.0, very strong.

Declarations

Acknowledgments

We are grateful for the contribution of all databases including GEO, GEPIA, Oncomine, Kaplan-Meier plotter, PrognosScan, cBioPortal data, STRING analysis and TIMER that provide free online tools and resources.

Authors' contributions

CH and CX designed the project. XJ and FZ contributed to the methods design. CH, MS and WH contributed to the data analysis and prepared the main manuscript. All authors reviewed the manuscript.

Funding

This work was supported by the following funding sources: 1. Key Laboratory of Molecular Diagnostics and Precision Medicine for Surgical Oncology in Gansu Province, Gansu Provincial Hospital, 24 West Donggang Road, Gansu 730000, China; 2. National Key Research and Development Program Task Statement (2018YFC1311506).

Availability of data and materials

The datasets generated and analysed during the current study are available from GEPIA, GEO, Oncomine, Kaplan-Meier plotter, PrognoScan, cBioPortal data, STRING analysis and TIMER that provide free online tools and resources.

Ethics approval and consent to participate

Not applicable. This article does not contain any studies with human participants or animals performed by any of the authors.

Competing interests

The authors declare that they have no competing interests.

Author details

¹The First Clinical Medical College, Lanzhou University, Lanzhou 730030, China; ²Department of Gynecology, Gansu Provincial Hospital, Lanzhou 730000, China; ³Department of General Surgery, The Second Clinical Medical College, Lanzhou University, Lanzhou 730030, China; ⁴Department of Surgery, Gansu Provincial Hospital, Lanzhou 730000, China

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(1):7–30.
2. Chen W, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016;66(2):115–32.
3. Oza AM, et al. Standard chemotherapy with or without bevacizumab for women with newly diagnosed ovarian cancer (ICON7): overall survival results of a phase 3 randomised trial. *Lancet Oncol.* 2015;16(8):928–36.
4. Kim JW, et al. Pazopanib Maintenance Therapy in East Asian Women With Advanced Epithelial Ovarian Cancer: Results From AGO-OVAR16 and an East Asian Study. *Int J Gynecol Cancer.*

- 2018;28(1):2–10.
5. Elzek MA, Rodland KD. Proteomics of ovarian cancer: functional insights and clinical applications. *Cancer Metastasis Rev.* 2015;34(1):83–96.
 6. Connor EV, et al. Thy-1 predicts poor prognosis and is associated with self-renewal in ovarian cancer. *Journal of ovarian research.* 2019;12(1):112.
 7. Han F, et al. SOX30 is a prognostic biomarker and chemotherapeutic indicator for advanced-stage ovarian cancer. *Endocrine-related Cancer.* 2019;26(3):303–19.
 8. Tothill RW, et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res.* 2008;14(16):5198–208.
 9. Yoshihara K, et al. Gene expression profile for predicting survival in advanced-stage serous ovarian cancer across two independent datasets. *PLoS One.* 2010;5(3):e9615.
 10. Yoshihara K, et al. Gene expression profiling of advanced-stage serous ovarian cancers distinguishes novel subclasses and implicates ZEB2 in tumor progression and prognosis. *Cancer Sci.* 2009;100(8):1421–8.
 11. Adib TR, et al. Predicting biomarkers for ovarian cancer using gene-expression microarrays. *Br J Cancer.* 2004;90(3):686–92.
 12. Lu KH, et al. Selection of potential markers for epithelial ovarian cancer with gene expression arrays and recursive descent partition analysis. *Clin Cancer Res.* 2004;10(10):3291–300.
 13. Bonome T, et al. A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. *Cancer Res.* 2008;68(13):5478–86.
 14. Elzek MA, Rodland KD. Proteomics of ovarian cancer: functional insights and clinical applications. *Cancer Metastasis Rev.* 2015;34(1):83–96.
 15. Fu J, et al. Roles of Aurora kinases in mitosis and tumorigenesis. *Mol Cancer Res.* 2007;5(1):1–10.
 16. Marumoto T, Zhang D, Saya H. Aurora-A - a guardian of poles. *Nat Rev Cancer.* 2005;5(1):42–50.
 17. Kollareddy M, et al. Aurora kinases: structure, functions and their association with cancer. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub.* 2008;152(1):27–33.
 18. Lee JW, et al. Combined Aurora Kinase A (AURKA) and WEE1 Inhibition Demonstrates Synergistic Antitumor Effect in Squamous Cell Carcinoma of the Head and Neck. *Clin Cancer Res.* 2019;25(11):3430–42.
 19. Guo M, et al. Increased AURKA promotes cell proliferation and predicts poor prognosis in bladder cancer. *BMC Syst Biol.* 2018;12(Suppl 7):118.
 20. Galusic D, et al. Higher AURKA and PLK1 expression are associated with inferior overall survival in patients with myelofibrosis. *Blood Cells Mol Dis.* 2020;81:102396.
 21. Wang-Bishop L, et al. Inhibition of AURKA Reduces Proliferation and Survival of Gastrointestinal Cancer Cells With Activated KRAS by Preventing Activation of RPS6KB1. *Gastroenterology.* 2019;156(3):662–75.e7.

22. Wang C, et al. Effect of AURKA Gene Expression Knockdown on Angiogenesis and Tumorigenesis of Human Ovarian Cancer Cell Lines. *Target Oncol.* 2016;11(6):771–81.
23. Ladurner R, et al. Sororin actively maintains sister chromatid cohesion. *Embo j.* 2016;35(6):635–53.
24. Schmitz J, et al. Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. *Curr Biol.* 2007;17(7):630–6.
25. Gómez R, et al. Sororin loads to the synaptonemal complex central region independently of meiotic cohesin complexes. *EMBO Rep.* 2016;17(5):695–707.
26. Chen H, 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.
27. Nguyen MH, 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.
28. Xu J, 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.
29. Tedeschi A, et al. Cep55 promotes cytokinesis of neural progenitors but is dispensable for most mammalian cell divisions. *Nat Commun.* 2020;11(1):1746.
30. Iwamori T, et al. TEX14 interacts with CEP55 to block cell abscission. *Mol Cell Biol.* 2010;30(9):2280–92.
31. Yin Y, et al. MiR-144 suppresses proliferation, invasion, and migration of breast cancer cells through inhibiting CEP55. *Cancer Biol Ther.* 2018;19(4):306–15.
32. Luo J, et al. MiR-195-5p Inhibits Proliferation and Induces Apoptosis of Non-Small Cell Lung Cancer Cells by Targeting CEP55. *Onco Targets Ther.* 2019;12:11465–74.
33. Wu M, et al. Identification of a Nine-Gene Signature and Establishment of a Prognostic Nomogram Predicting Overall Survival of Pancreatic Cancer. *Front Oncol.* 2019;9:996.
34. Jeffery J, et al. Beyond cytokinesis: the emerging roles of CEP55 in tumorigenesis. *Oncogene.* 2016;35(6):683–90.
35. Xie C, et al. Ubiquitin-conjugating enzyme E2C: a potential cancer biomarker. *Int J Biochem Cell Biol.* 2014;47:113–7.
36. Shen R, et al. The clinicopathological significance of ubiquitin-conjugating enzyme E2C, leucine-rich repeated-containing G protein-coupled receptor, WW domain-containing oxidoreductase, and vasculogenic mimicry in invasive breast carcinoma. *Medicine.* 2019;98(16):e15232.
37. Zhang J, et al. UBE2C Is a Potential Biomarker of Intestinal-Type Gastric Cancer With Chromosomal Instability. *Front Pharmacol.* 2018;9:847.
38. Liu PF, et al., *UBE2C is a Potential Biomarker for Tumorigenesis and Prognosis in Tongue Squamous Cell Carcinoma.* *Diagnostics (Basel),* 2020. 10(9).

39. Kim YJ, et al. UBE2C Overexpression Aggravates Patient Outcome by Promoting Estrogen-Dependent/Independent Cell Proliferation in Early Hormone Receptor-Positive and HER2-Negative Breast Cancer. *Front Oncol.* 2019;9:1574.
40. Lili LN, et al., *Molecular profiling predicts the existence of two functionally distinct classes of ovarian cancer stroma.* *Biomed Res Int*, 2013. **2013**: p. 846387.
41. King ER, et al. The anterior gradient homolog 3 (AGR3) gene is associated with differentiation and survival in ovarian cancer. *Am J Surg Pathol.* 2011;35(6):904–12.
42. Bowen NJ, et al. Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells. *BMC Med Genomics.* 2009;2:71.
43. Mok SC, et al. A gene signature predictive for outcome in advanced ovarian cancer identifies a survival factor: microfibril-associated glycoprotein 2. *Cancer Cell.* 2009;16(6):521–32.
44. Irizarry RA, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* 2003;4(2):249–64.
45. Gautier L, et al. affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics.* 2004;20(3):307–15.
46. Smyth GK. *Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray.* *Experiments**. *Statistical Applications in Genetics and Molecular Biology*, 2004. **3**(1).
47. Otasek D, et al. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol.* 2019;20(1):185.
48. Yu G, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics.* 2012;16(5):284–7.
49. Walter W, Sánchez-Cabo F, Ricote M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics.* 2015;31(17):2912–4.
50. Gyorffy B, Lánckzy A, Szállási Z. Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. *Endocr Relat Cancer.* 2012;19(2):197–208.
51. Mizuno H, et al. PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med Genomics.* 2009;2:18.
52. Rhodes DR, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia.* 2004;6(1):1–6.
53. Rhodes DR, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia.* 2007;9(2):166–80.
54. Tang Z, et al. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* 2019;47(W1):W556–w560.
55. Uhlén M, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.

56. Gao J, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):p11.
57. Warde-Farley D, et al., *The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function*. *Nucleic Acids Res*, 2010. **38**(Web Server issue): p. W214-20.
58. Li T, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res*. 2017;77(21):e108–10.
59. Li B, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol*. 2016;17(1):174.
60. *Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray..*

Tables

Table 1 is not available with this version.

Figures

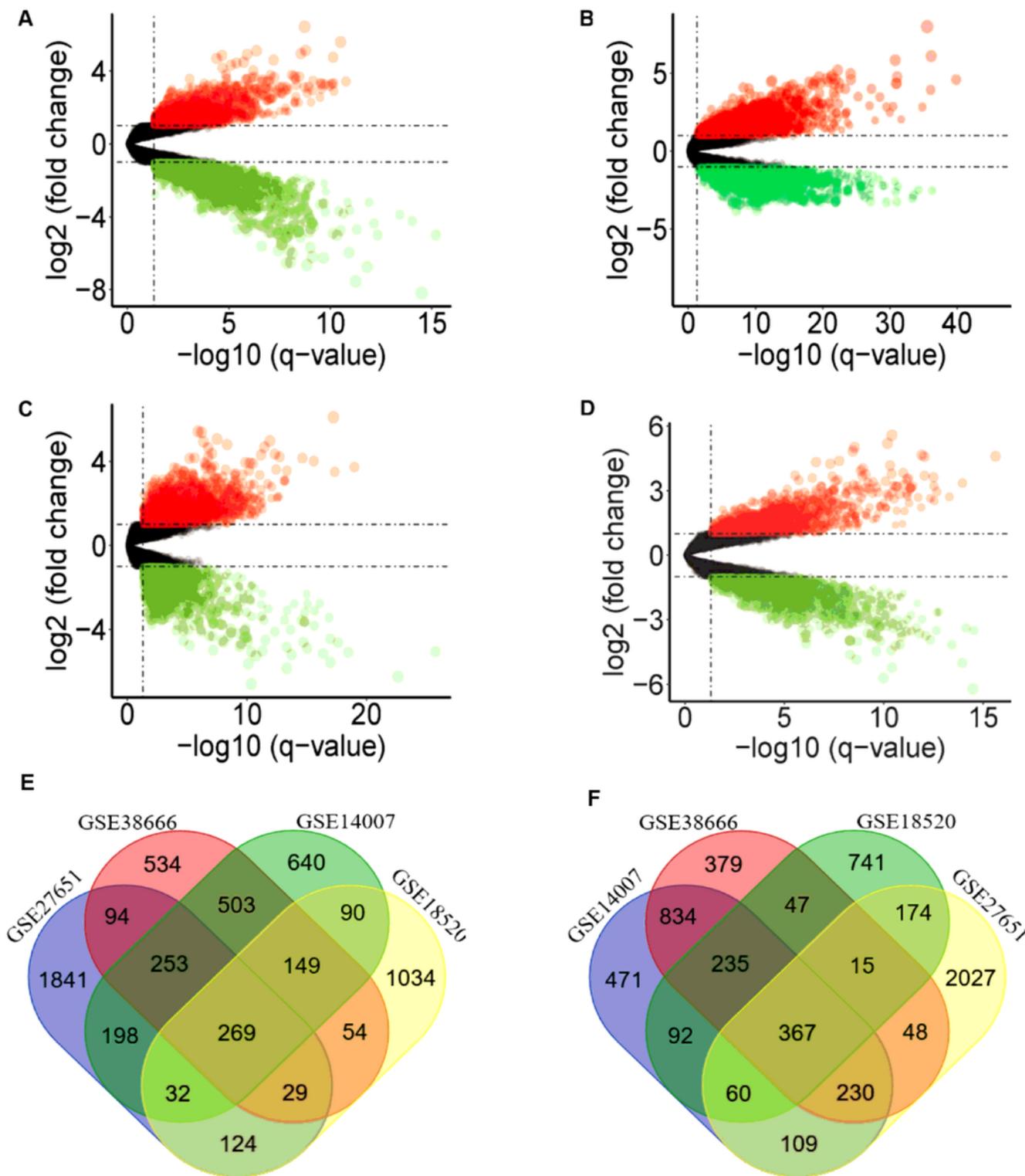


Figure 1

Initially identified 5,870, 4,040, 4,532 and 18,520 DEGs from GSE27651, GSE38666, GSE14407 and GSE18520, respectively, with the limma package of Rtools and the criteria $|\log_2\text{FC}| > 1$ and adjusted P value < 0.01 (Figure 1A-D). Then, all DEGs were divided into up- and down-regulated genes according to whether $|\log_2\text{FC}| > 0$ or $|\log_2\text{FC}| < 0$, respectively. A Venn diagram tool was applied to identify the DEGs in the sets of up- and downregulated DEGs. A total of 636 DEGs, including 269 upregulated and 367

downregulated DEGs, were identified in the OC samples compared to the normal ovarian tissue samples (Figure 1E&F). All of these DEGs are listed in Table S2.

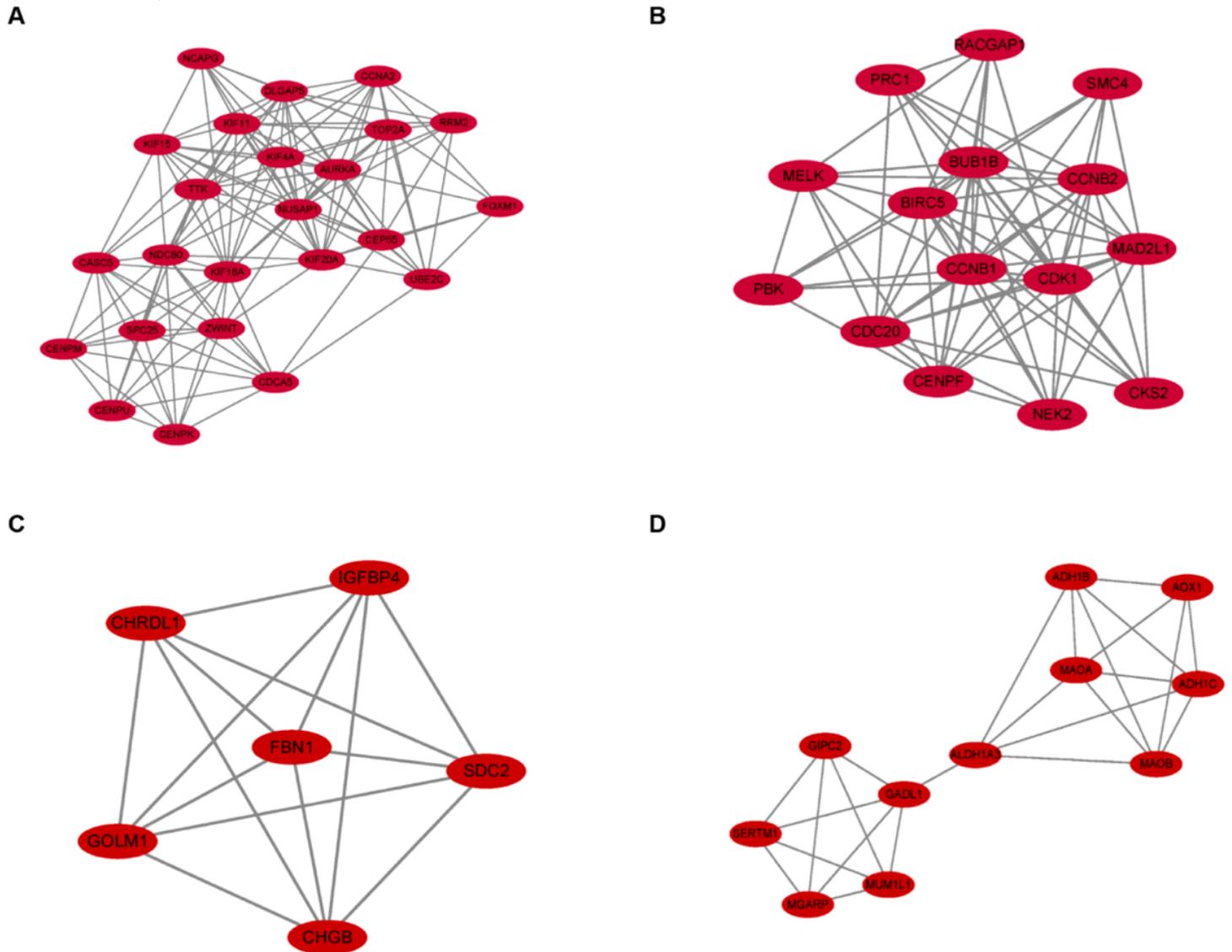


Figure 2

Two gene clusters with network scores of >10 were selected from the upregulated DEGs (Figure 2A&B), and two gene clusters with scores of >5 from the downregulated gene clusters (Figure 2C&D) were selected as the most important molecular modules in the respective results and further analyzed for function and prognosis.

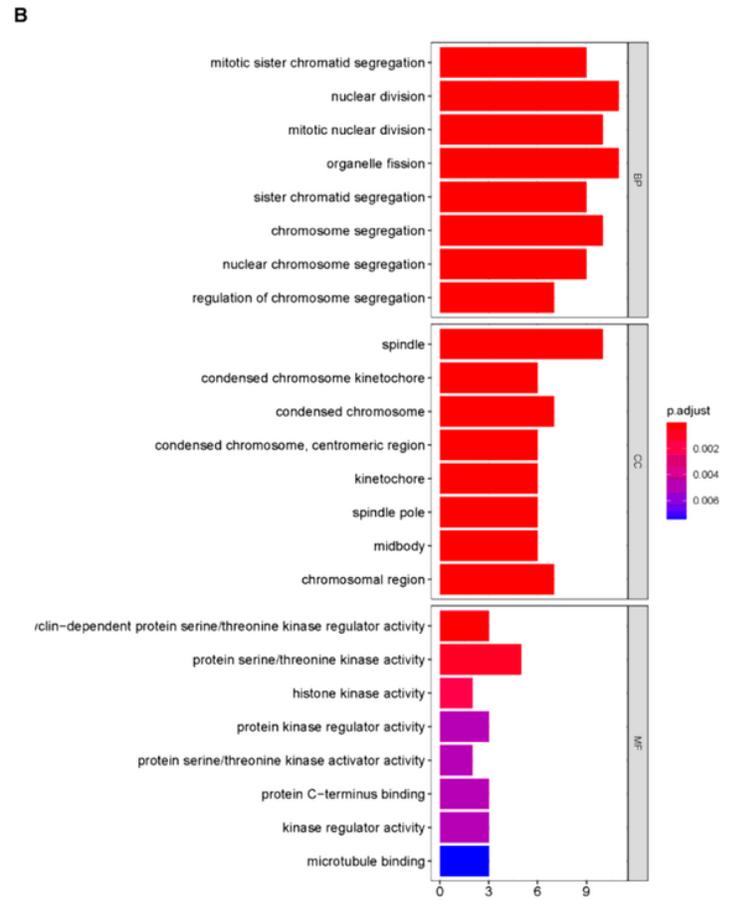
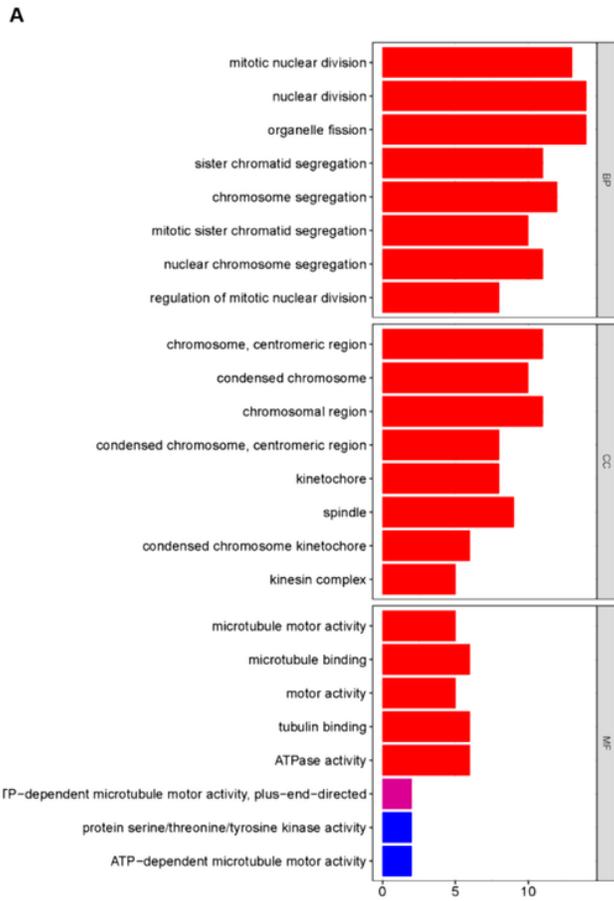


Figure 3

The primary molecular functions included microtubule binding and various kinase regulator activities (Figure 3A&B).

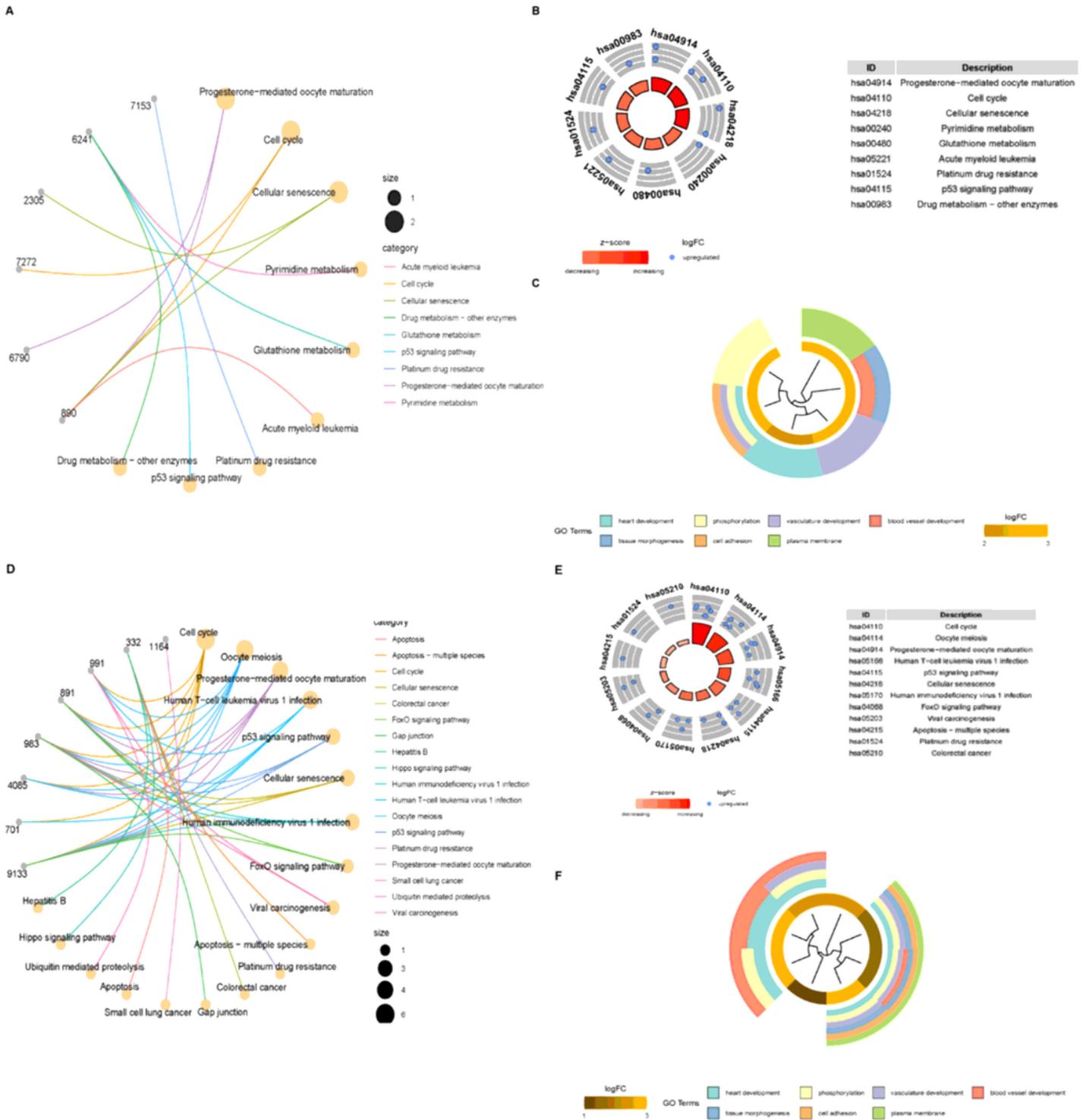


Figure 4

KEGG pathway analysis indicated genes enriched mainly in oocyte maturation, cell cycle, p53 signaling pathway, and cellular senescence pathway (Figures 4A-F).

No signaling pathways were significantly enriched for the first gene cluster ($P < 0.05$). The second gene cluster was mainly involved in many metabolic functions, such as tyrosine metabolism, fatty acid degradation and retinol metabolism (Figure 6A-C).

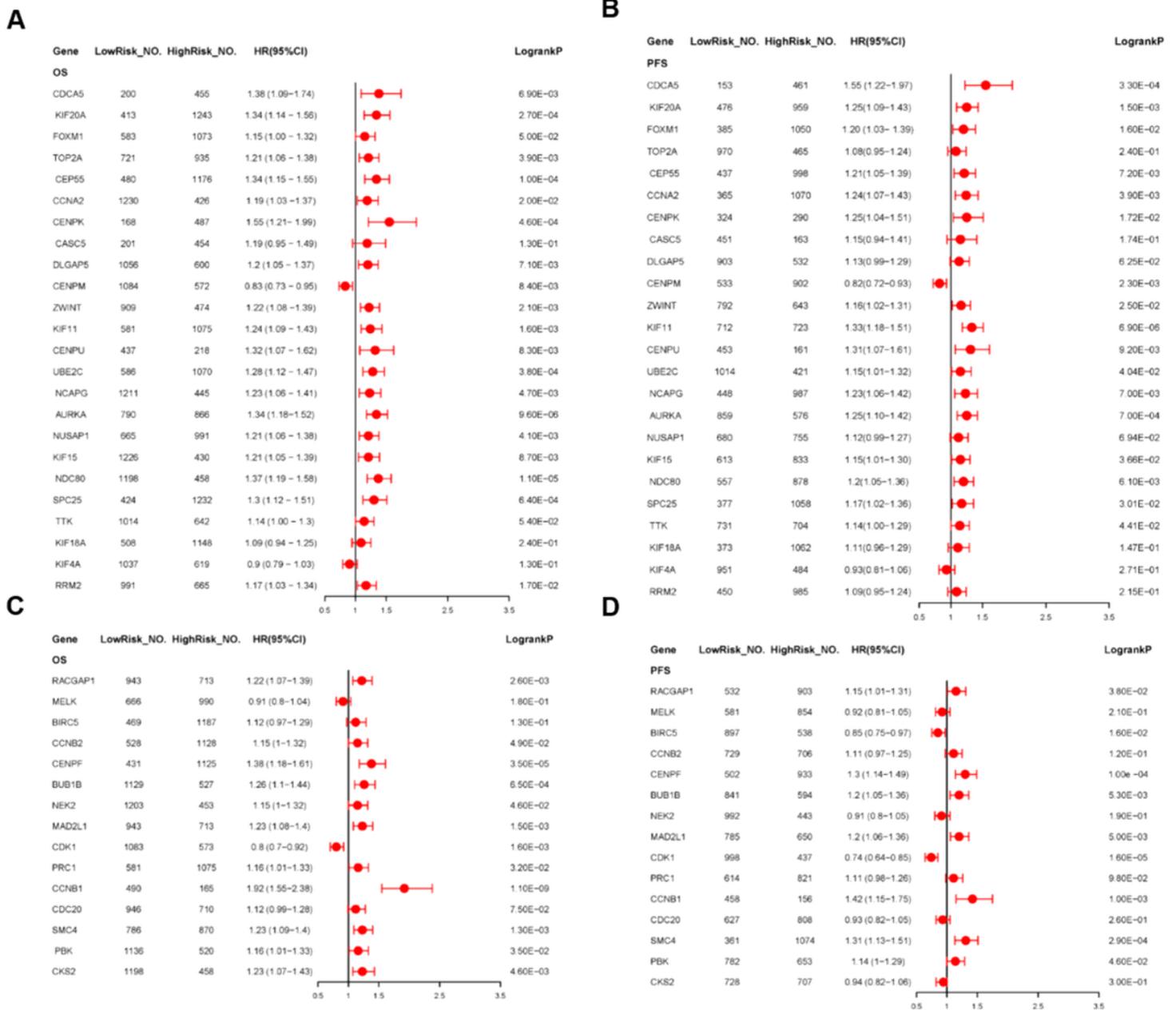
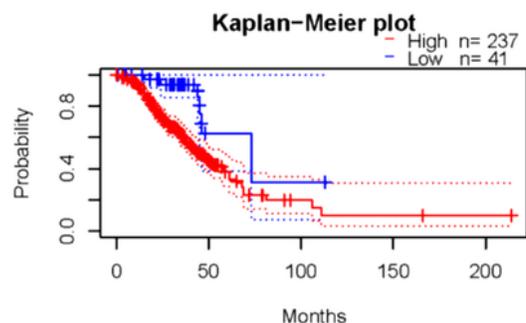


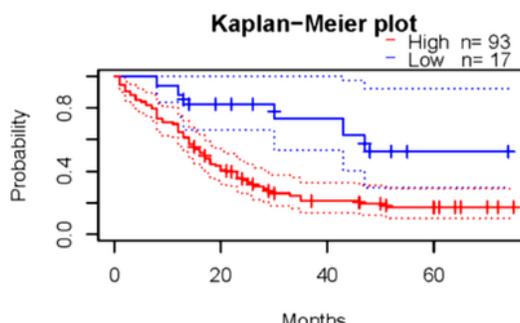
Figure 7

Twenty-six DEGs from upregulated DEGs were identified to correlate with OS and PFS in ovarian cancer. The results were visualized through the forestplot package of Rtools (Figures 7A-D).

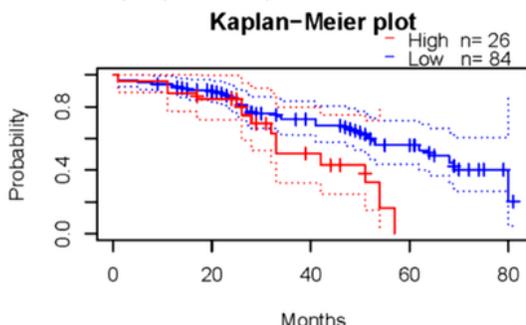
A GSE9891, OS, HR=1.27, Cox P-value=0.012251



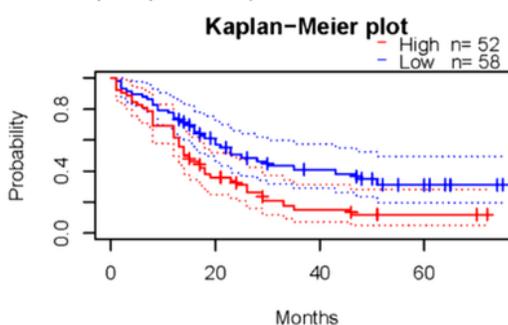
B GSE17260,PFS,HR=1.32,CoxP-value =0.026599



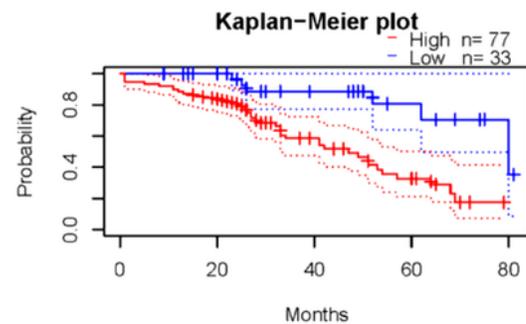
C GSE17260,OS,HR=1.55,CoxP-value =0.027802



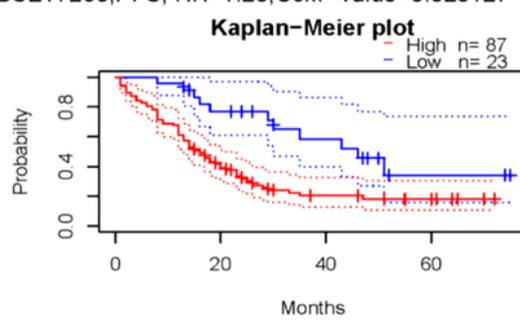
D GSE17260,PFS,HR=1.41,Cox P-value =0.015281



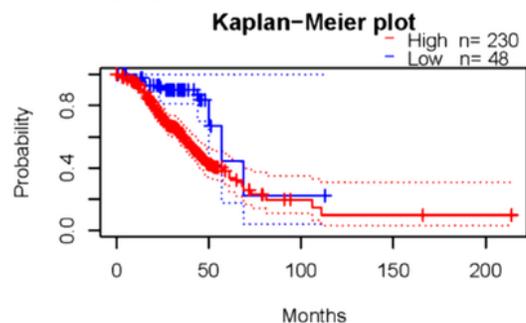
E GSE17260,OS, HR=1.55,CoxP-value=0.004303



F GSE17260,PFS, HR=1.28,CoxP-value=0.025127



G GSE17260,OS, HR=1.26,CoxP-value=0.008524



H GSE17260,PFS, HR=1.32,CoxP-value=0.017412

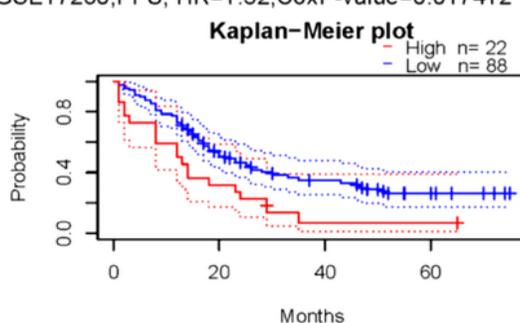


Figure 8

To further examine the prognostic potential of the twenty-eight genes from the up- and downregulated DEGs, we continued to reanalyze the survival differences through the Prognoscan database. Interestingly, two cohorts (GSE9891 and GSE17260) [8, 9], which included 278 samples and 110 samples at different stages of OC, showed that higher expression of AURKA, CDCA5, CEP55 and UBE2C was significantly

associated with poorer prognosis (Figures 8A-H). Therefore, it is conceivable that high AURKA CDCA5, CEP55 and UBE2C expression is an independent risk factor and leads to a poor prognosis in OC patients.

Analysis Type by Cancer	Cancer vs. Normal		Cancer vs. Normal		Cancer vs. Normal		Cancer vs. Normal	
	AURKA		CDCA5		CEP55		UBE2C	
Bladder Cancer	6		2		4		6	
Brain and CNS Cancer	4	1	1	1	6		11	1
Breast Cancer	22	1	18		21		23	
Cervical Cancer	3		2		4		3	
Colorectal Cancer	19		12		16		19	
Esophageal Cancer	4		1		3		4	
Gastric Cancer	7		10		5		8	
Head and Neck Cancer	11		6		8		11	
Kidney Cancer		2			5		1	
Leukemia	1	9		2	1	2	1	4
Liver Cancer	4		2		4		5	
Lung Cancer	16		12		12		18	
Lymphoma	6	1	7		7		5	
Melanoma			1				1	
Myeloma								2
Other Cancer	4	2	3	2	7	6	11	
Ovarian Cancer	5		1		5		8	
Pancreatic Cancer	1	2	3	1	2		3	1
Prostate Cancer	3				1		2	
Sarcoma	10	1			9		11	
Significant Unique Analyses	125	19	81	6	118	8	149	8
Total Unique Analyses	340		305		387		356	

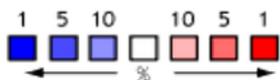


Figure 9

We compared the transcriptional levels of these four genes in cancers with those in normal tissue samples using the ONCOMINE database (Figure 9).

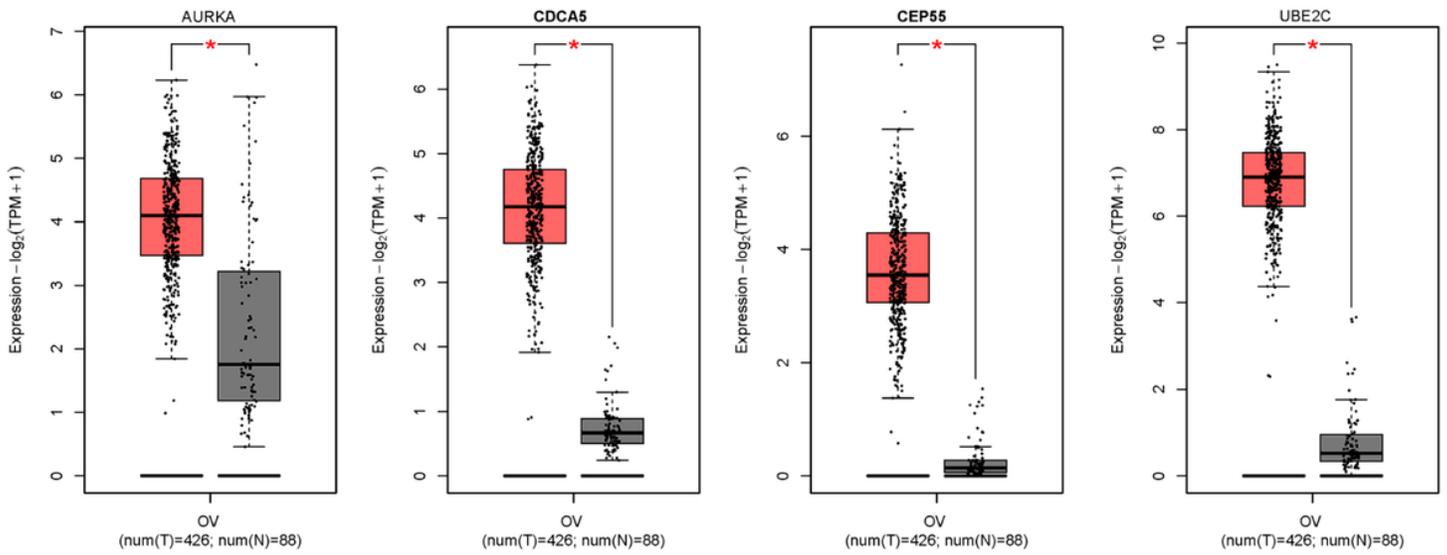
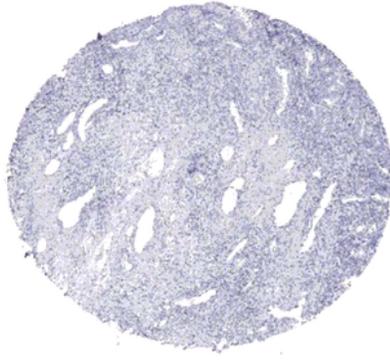


Figure 10

Using the GEPIA dataset, we further confirmed the differential mRNA expression of the four genes between OC and normal tissues. The results showed that mRNA expression levels of AURKA, CDCA5, CEP55 and UBE2C were significantly increased in OC tissues compared with normal ovarian tissues [(tumor sample: n = 426 vs. normal sample: n = 88) (Figure 10).

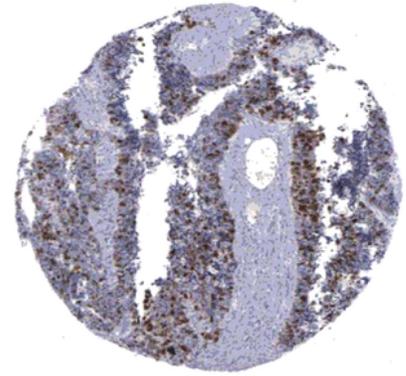
AURKA

Normal Tissue
Age: 33
Patient id: 2159
Staining: Not detected
Intensity: Negative
Quantity: None



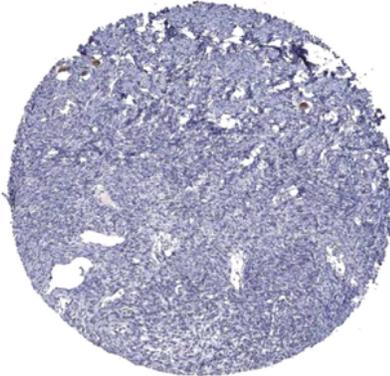
AURKA

Tumor Tissue
Age: 77
Patient id: 1115
Staining: High
Intensity: Strong
Quantity: 75%-25%



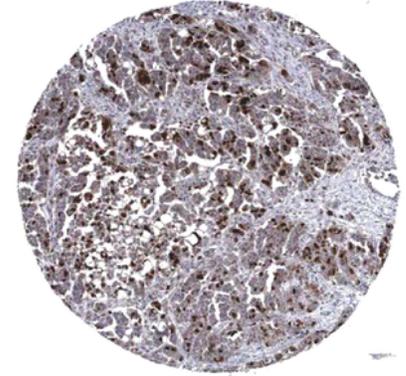
CDCA5

Normal Tissue
Age: 27
Patient id: 2004
Staining: Not detected
Intensity: Negative
Quantity: None



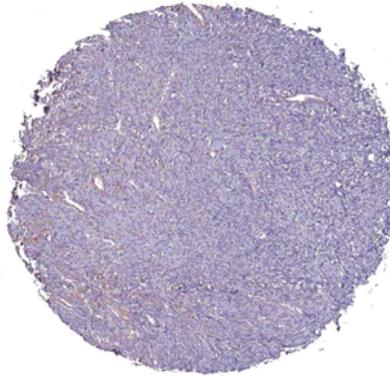
CDCA5

Tumor Tissue
Age: 55
Patient id: 3116
Staining: High
Intensity: Strong
Quantity: 75%-25%



CEP55

Normal Tissue
Age: 47
Patient id: 27113
Staining: Not detected
Intensity: Negative
Quantity: None



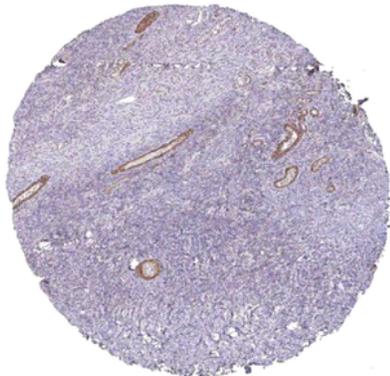
CEP55

Tumor Tissue
Age: 42
Patient id: 2118
Staining: Medium
Intensity: Moderate
Quantity: 75%-25%



UBE2C

Normal Tissue
Age: 35
Patient id: 4030
Staining: Not detected
Intensity: Negative
Quantity: None



UBE2C

Tumor Tissue
Age: 37
Patient id: 857
Staining: Medium
Intensity: Moderate
Quantity: 75%-25%



Figure 11

The IHC staining and images were downloaded from the Human Protein Atlas ,the proteins levels of these four prognostic genes was significantly elevated in tumor tissues compared with normal tissues (Figure 11).

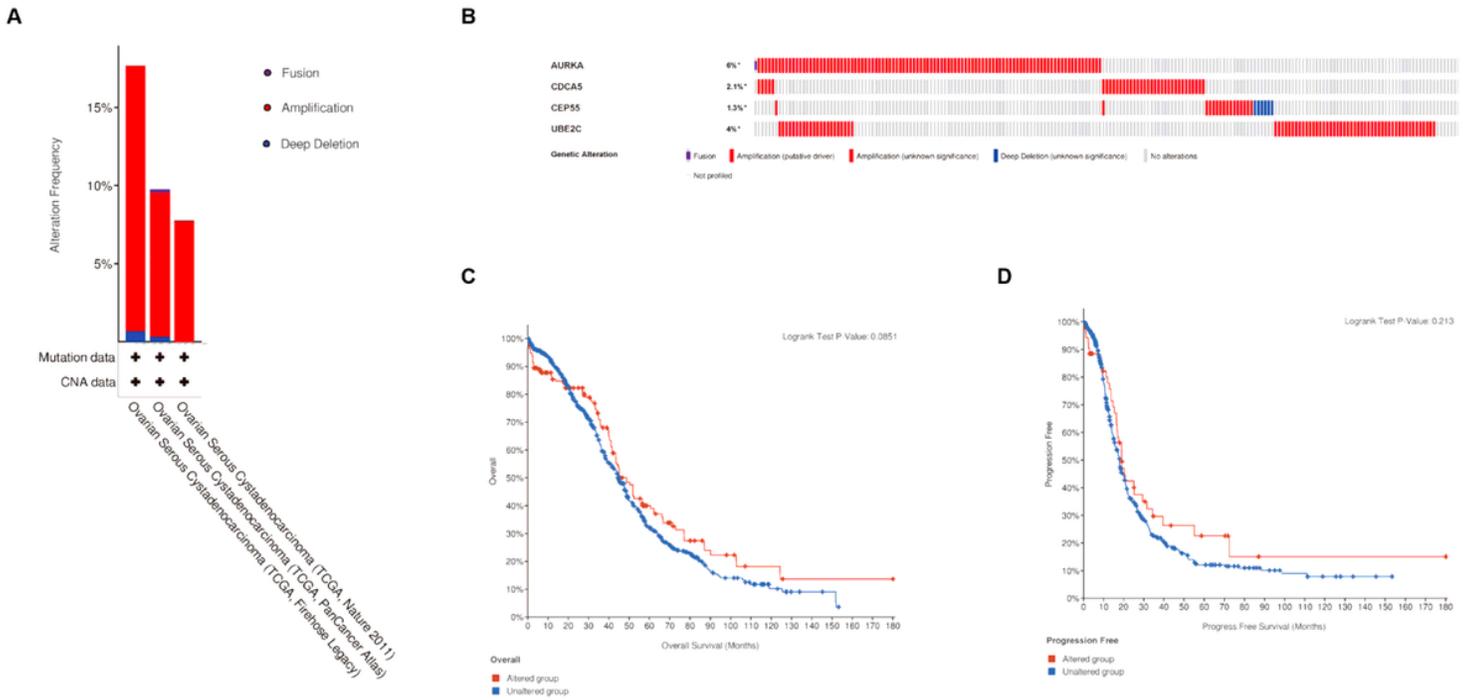


Figure 12

Mutations in AURKA, CDCA5, CEP55 and UBE2C retrieved in 1,680 cases from three TCGA datasets of ovarian serous carcinoma (606 cases from TCGA, Firehose Legacy; 585 cases from TCGA, PanCancer Atlas; and 489 cases from TCGA, Nature 2011) were analyzed using the cBioPortal database. Among the 3 OC datasets analyzed, the alterations of the four genes were mainly related to amplification and deep deletion, and ranges from 17.67% of 583 genes to 7.77% of 489 genes were identified for the gene sets submitted for analysis (Figure 12A). The percentages of AURKA, CDCA5, CEP55 and UBE2C gene alterations in OC were 6%, 2.1%, 1.3% and 4% (Figure 12B), respectively. Kaplan–Meier Plotter and log-rank test results indicated no significant difference in OS and PFS between the cases with alterations in one of the queried genes and those without alterations in any of the queried genes (P-values of 0.0851 and 0.213, respectively; Figure 12 C&D).

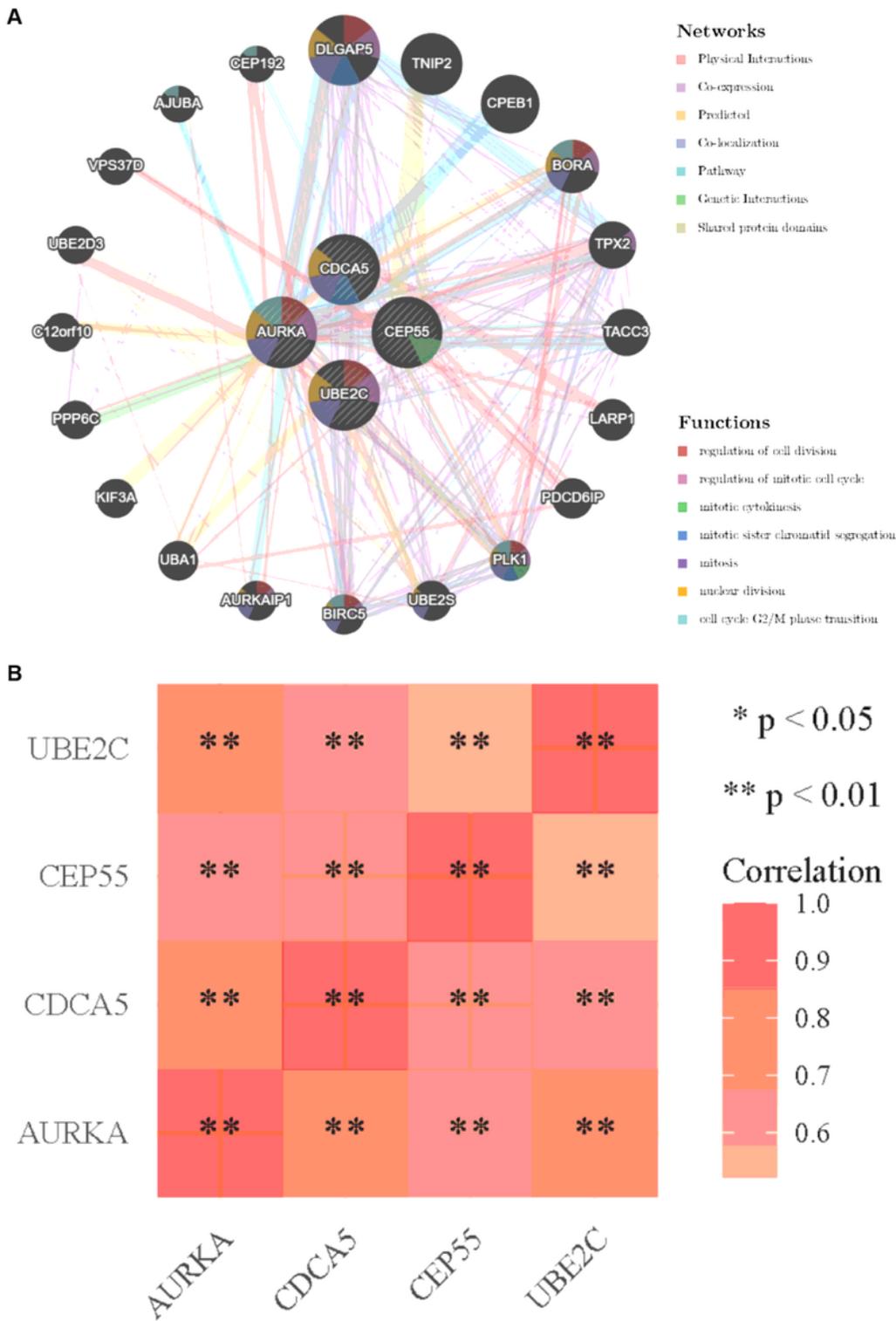


Figure 13

The GeneMANIA database was used for correlation analysis of AURKA, CDCA5, CEP55 and UBE2C at the gene level (Figure 13A). The Spearman correlations between AURKA, CDCA5, CEP55 and UBE2C in OC were determined by online analyses using the cBioPortal database (TCGA, PanCancer Atlas (Figure 13B)).

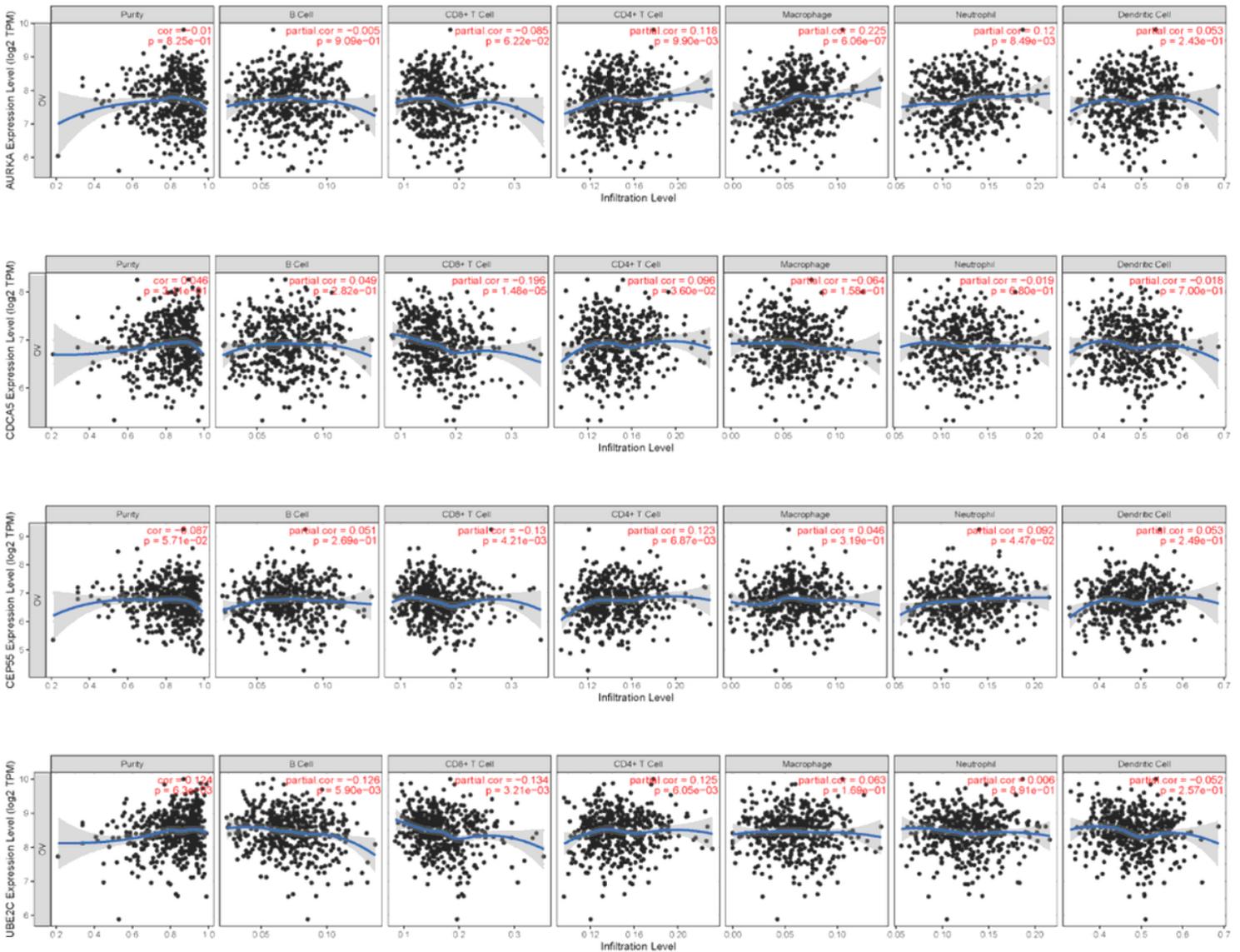


Figure 14

Tumor inflammatory cell infiltration levels play an important role in cancer progression and patient survival. We analyzed whether AURKA, CDCA5, CEP55 and UBE2C expression correlated with inflammatory cell infiltration levels in OC. The results showed that the mRNA levels of the four genes correlated weakly with inflammatory cell infiltration (Figure 14).

Supplementary Files

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

- [TableS1.docx](#)
- [TableS2.docx](#)
- [TableS3.docx](#)
- [TableS4.docx](#)

- TableS5.docx
- FIGUERS1.tif
- FIGURES2.tif
- FIGURES3.tif
- FIUGRES4.tif