

# Identification of core genes for early diagnosis and the EMT modulation of ovarian serous cancer by bioinformatics perspective

Yanna Zhang<sup>1</sup>, Xun Wang<sup>1</sup>, Xiancheng Chen<sup>1</sup>

<sup>1</sup>State Key Laboratory of Biotherapy, Collaborative Innovation Center for Biotherapy, West China Hospital, Sichuan University, High Technological Development Zone, Chengdu 610041, Sichuan, People's Republic of China

**Correspondence to:** Xiancheng Chen; **email:** [chenwcrs@scu.edu.cn](mailto:chenwcrs@scu.edu.cn)

**Keywords:** ovarian serous cancer, early diagnosis, biomarkers, EMT

**Received:** August 28, 2020

**Accepted:** November 30, 2020

**Published:** January 25, 2021

**Copyright:** © 2021 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Ovarian serous carcinoma (OSC), as a common malignant tumor, poses a serious threat to women's health in that epithelial-mesenchymal transformation (EMT)-related modulation becomes heavily implicated in the invasion and progression of OSC. In this study, two core genes (BUB1B and NDC80) among the 16 hub genes have been identified to be involved in the molecular regulation of EMT and associated with the poor early survival of OSC at stages I+II. Through the Gene Regulatory Networks (GRN) analysis of 15 EMT regulators and core genes, it was revealed that TFAP2A and hsa-miR-655 could elaborately modulate EMT development of OSC. Next genetic variation analysis indicated that EMT regulator ELF3 would also serve as a crucial part in the occurrence and progression of OSC. Eventually, survival investigation suggested that TFAP2A, ELF3 and hsa-miR-655 were significantly associated with the overall survival of progressive OSC patients. Thus, combined with diversified bioinformatic analyses, BUB1B, NDC80, TFAP2A, ELF3 and hsa-miR-655 may act as the key biomarkers for early clinical diagnosis and prognosis evaluation of OSC patients as well as potential therapeutic target-points.

## INTRODUCTION

Ovarian cancer, about 90% originated from ovarian surface epithelium, is one of the most common malignant tumors in the female reproductive system, with a very low five-year survival rate and the highest mortality rate among all kinds of female tumors [1]. The molecular mechanisms in ovarian cancer development, recurrence and metastasis are complex and changeable, leading to insufficient innovation in early clinical diagnosis and treatment models [2]. As the most common pathological type among ovarian carcinomas, ovarian serous cancer (OSC) accounts for 80–95% of ovarian malignancies [3]. Although current therapeutic strategies for OSC have improved significantly, the 5-year survival rate of OSC is still much lower than other gynecological malignancies, with a relapsing rate of ~70% of patients [4]. Elucidating the molecular mechanism of OSC may help us understand the patho-

genesis and progress of OSC and identify new targets for effective treatment. However, relatively little is known about the molecular events leading to the development of this highly invasive disease [5].

Epithelial-mesenchymal transformation (EMT), one of the core biological process in the occurrence and development of epithelial ovarian cancer (EOC), has been considered as the crucial mechanism of ovarian surface epithelial cells participating in ovarian tumor growth, migration, invasion, metastasis and drug resistance formation [6]. Moreover, it is necessary to explore more meaningful biomarkers for early clinical diagnosis and EMT biological mechanism related to the pathogenesis, early prevention and treatment of OSC.

The main features of EMT are loss of epithelial phenotype and acquisition of stromal features [7], which makes epithelial cells lose intercellular connection,

reduce adhesion, and obtain mesenchymal characteristics [8]. The dissociation of cell connections during the EMT process is not a "collapse" caused by simple cytoskeletal changes, but a more fine-tuned way in that first is the dissociation of adhesive connections, then the cytoskeletal changes and the multi-step dissociation of tight connections process [9]. EMT plays an essential role in wound healing, stemness acquirement, tissue fibrosis and in OSC deterioration covering cancer invasiveness progression, distant metastasis, relapse and drug resistance developments [10]. Previous studies have shown that EMT transcriptional regulators, such as CDH1/E-cadherin, CDH2/N-cadherin, ZEB1, ZEB2, SNAI1, SNAI2, TWIST1 and TWIST2, are essential for promoting cell invasion, migration, proliferation and angiogenesis [11, 12]. During EMT conversion, tumor cells undergo obvious cytoskeletal reconstruction based on the expression of various transcription factors and activation of surface receptors related to EMT phenotype [13]. Dr. Ruby Huang's research indicated that AXL, when activated as an EMT regulator, could interact with other proteins, thus forming an intracellular signal to enhance the invasion, migration and proliferation of ovarian cancer cells as the crucial signal of EMT promoting cancer development [14]. ARK5/NUAK1 and HOXA10, as the regulatory factors in EMT cascade loop, were remarkably upregulated, when compared with adjacent normal tissues, thus enhancing invasiveness of ovarian cancer [15]. In the main signaling pathways closely related to EMT, BIRC5, CTNBN1 and relevant other proteins could also enhance the migration and proliferation of ovarian cancer cells with the expression of EMT markers [16, 17]. Meanwhile, PARP-1, also as a core EMT regulator, could play an important role in OSC progression [18, 19]. However, another EMT regulator ELF3, when up-regulated, could mediate the EMT signal molecules cascade to increase the expression of epithelial markers and decrease the mesenchymal markers in ovarian cancer [20]. Therefore, EMT is closely involved in the process of tumor development, invasion, metastasis and recurrence of OSC [13]. Nevertheless, the relation of expression profile of transcription factors and proteins associated with EMT to the diverse pathological features of OSC has not been comprehensively investigated.

In this study, the obvious common differential expression genes (co-DEGs) from the gene expression data of 3 OSC datasets in the Gene Expression Omnibus (GEO) database were screened for Gene Ontology (GO) and Kyoto Encyclopedia of Gene and Genome (KEGG) functional enrichment analysis. Through the protein-protein interactions (PPIs) network co-expression interaction of co-DEGs, we carried out the integrated bioinformatics analysis to find those core genes with a significant hint for early clinical diagnosis on OSC.

Then, co-expression analyses for core genes and 15 EMT regulators involved in the multiple pathological features of ovarian cancer were conducted to detect the critical regulatory role of these genes and regulators in the occurrence and progression of OSC. Next, ovarian specific co-expression regulation analysis, KEGG function enrichment analysis, genetic variation, mutation count, overall survival status, and GRN analysis would be managed respectively. In summary, based on comprehensive bioinformatics analyses [21–23], this study would assist with exploring the potential biomarkers, elucidating the mechanisms underlying relevant pathophysiological events and finally exploiting effective and reliable targeted therapies for OSC.

## RESULTS

### DEGs identification with data normalization

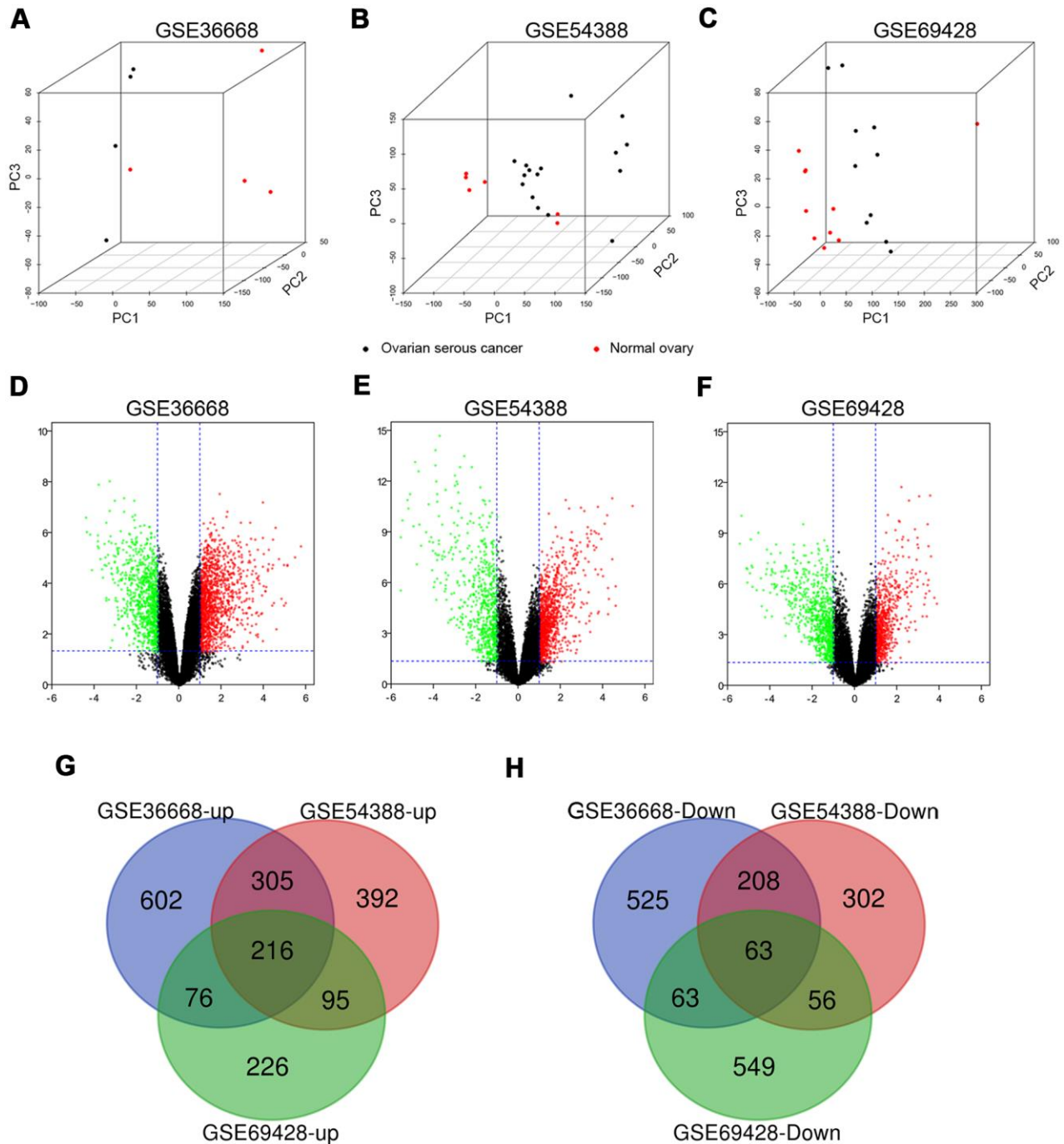
Three expression profiles (GSE36668, GSE54388, and GSE69428) were obtained from the GEO database, and the specific details were listed in Supplementary Table 1. These datasets, covering OSC tissues and normal ovary tissues, were both from patients with OSC, with GSE36668 including 4 OSC tissues and 4 normal ovary tissues, GSE54388 containing 16 OSC tissues and 6 normal ovary tissues, GSE69428 consisting of 10 OSC tissues and 10 normal ovary tissues (Supplementary Table 1). We evaluated these datasets by Principal component analysis (PCA) after data normalization (Figure 1A–1C and Supplementary Videos 1–3). Then, the heatmaps of gene expression in GSE36668, GSE54388 and GSE69428 were shown in Supplementary Figure 2A–2C. After gene annotation, the DEGs were screened in each data series with  $\text{Log FC} \geq 1$  or  $\text{Log FC} \leq -1$  and  $p\text{-value} < 0.05$  as the criteria for selection. The GSE36668 dataset included 2058 DEGs, covering 1199 upregulated and 859 downregulated genes (Figure 1D); the GSE54388 dataset included 1637 DEGs consisting of 1008 upregulated and 629 downregulated genes (Figure 1E); the GSE69428 dataset included 1344 DEGs consisting of 613 upregulated and 731 downregulated genes (Figure 1F) as shown in Volcano plots. The details of significant DEGs from each dataset were displayed in Supplementary Table 2. Moreover, the overlap of co-DEGs in three datasets contained 279 genes, as Venn diagrams showed, consisting of 216 upregulated co-DEGs (Figure 1G) and 63 downregulated co-DEGs (Figure 1H) when compared with normal ovary samples.

### Functional annotation and PPIs network of co-DEGs

279 co-DEGs were subjected to GO enrichment for biological process (BP), molecular function (MF) and cellular component (CC) analyses according to criterion

of  $p$ -value<0.01. The BP analysis of co-DEGs was mainly focused on the cell division (GO:0051301), mitotic nuclear division (GO:0007067) and DNA replication (GO:0006260) (Supplementary Table 3 and

Figure 2A). For CC analysis, the co-DEGs were notably enriched in the midbody (GO:0030496), nucleoplasm (GO:0005654) and cytoplasm (GO:0005737) (Supplementary Table 3 and Figure 2B). Concerning the

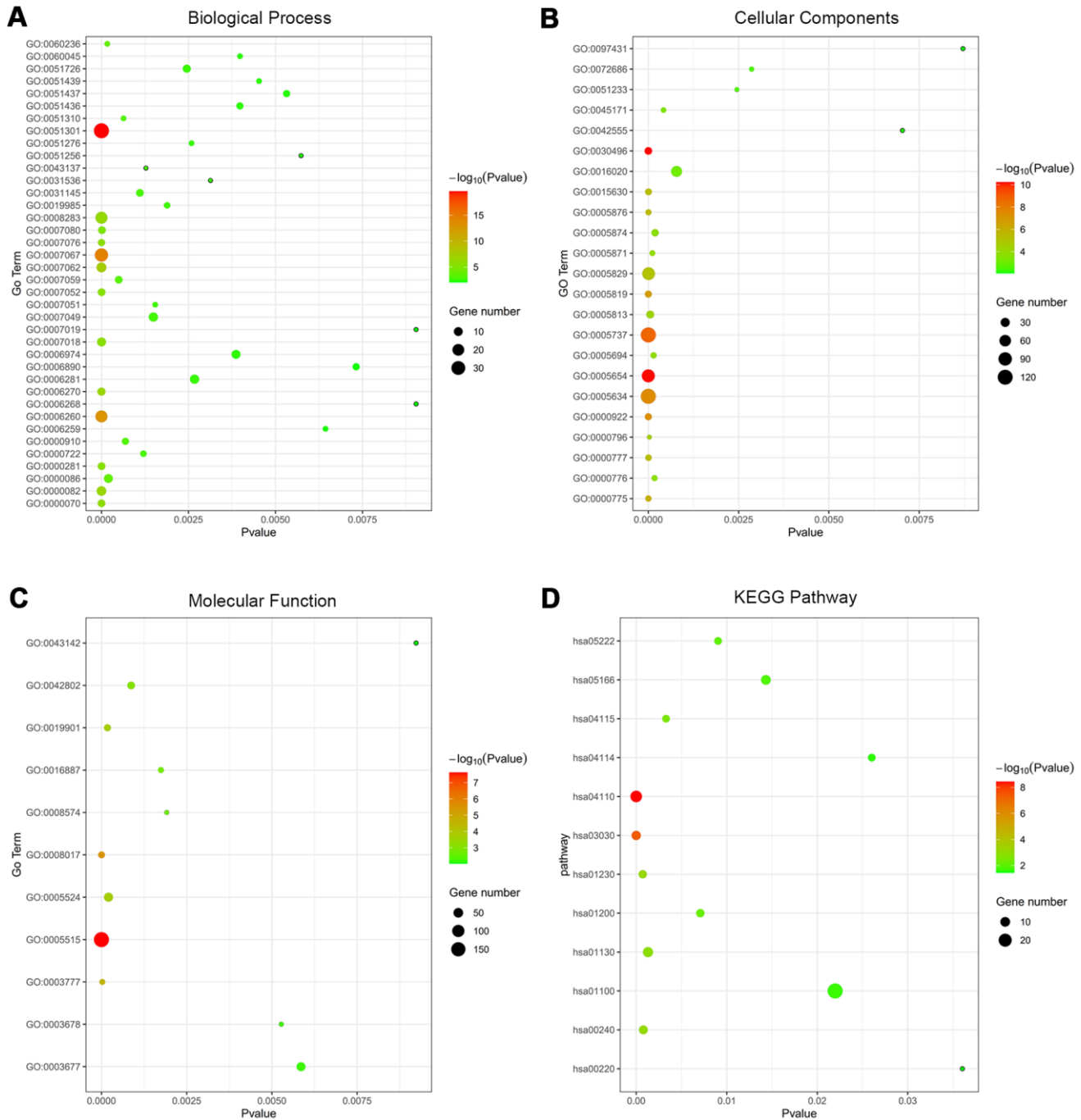


**Figure 1. The distribution of expression situation and DEGs identification among GSE36668, GSE54388 and GSE69428 after normalization.** (A–C) Whole transcriptomes were subjected to PCA on expressed genes to assess sample diversity and relatedness between OSC tissues (black dot) and normal ovary tissues (red dot). See also Supplementary Videos 1–3 (Supporting Information). (D, E) Volcano plots represent DEGs between OSC tissues and normal ovary tissues. Red dots indicate upregulation in DEG (LogFC ≥ 1,  $p$ -value < 0.05), and green dots indicate down regulation (LogFC ≤ -1,  $p$ -value < 0.05). Three-way Venn diagram based on whole transcriptomes represents the distribution of the up expressed genes (G) and the down expressed genes (H) among these datasets.

MF analysis, the co-DEGs were mostly enriched in protein binding (GO:0005515), microtubule binding (GO:0008017) and microtubule motor activity (GO:0003777) (Supplementary Table 3 and Figure 2C). In addition, we utilized the DAVID to categorize co-DEGs in the KEGG database. Subsequent results indicated that the co-DEGs were significantly involved in 12 signaling pathways, such as cell cycle (hsa04110),

DNA replication (hsa03030), and biosynthesis of amino acids (hsa01230) (Supplementary Table 3 and Figure 2D).

Then, PPIs network was constructed to explore the interaction relationship among co-DEGs in the pathogenesis of OSC. The interaction score  $\geq 0.09$  (high-confidence interaction score) for nodes was considered as a pronounced PPIs network (Figure 3). The backbone



**Figure 2. Functional annotation of co-DEGs between OSC tissues and normal ovary tissues using GO terms of BP, CC, MF and KEGG pathway.** Bubble map for Go terms of BP (A), CC (B), MF (C) and KEGG pathway (D).

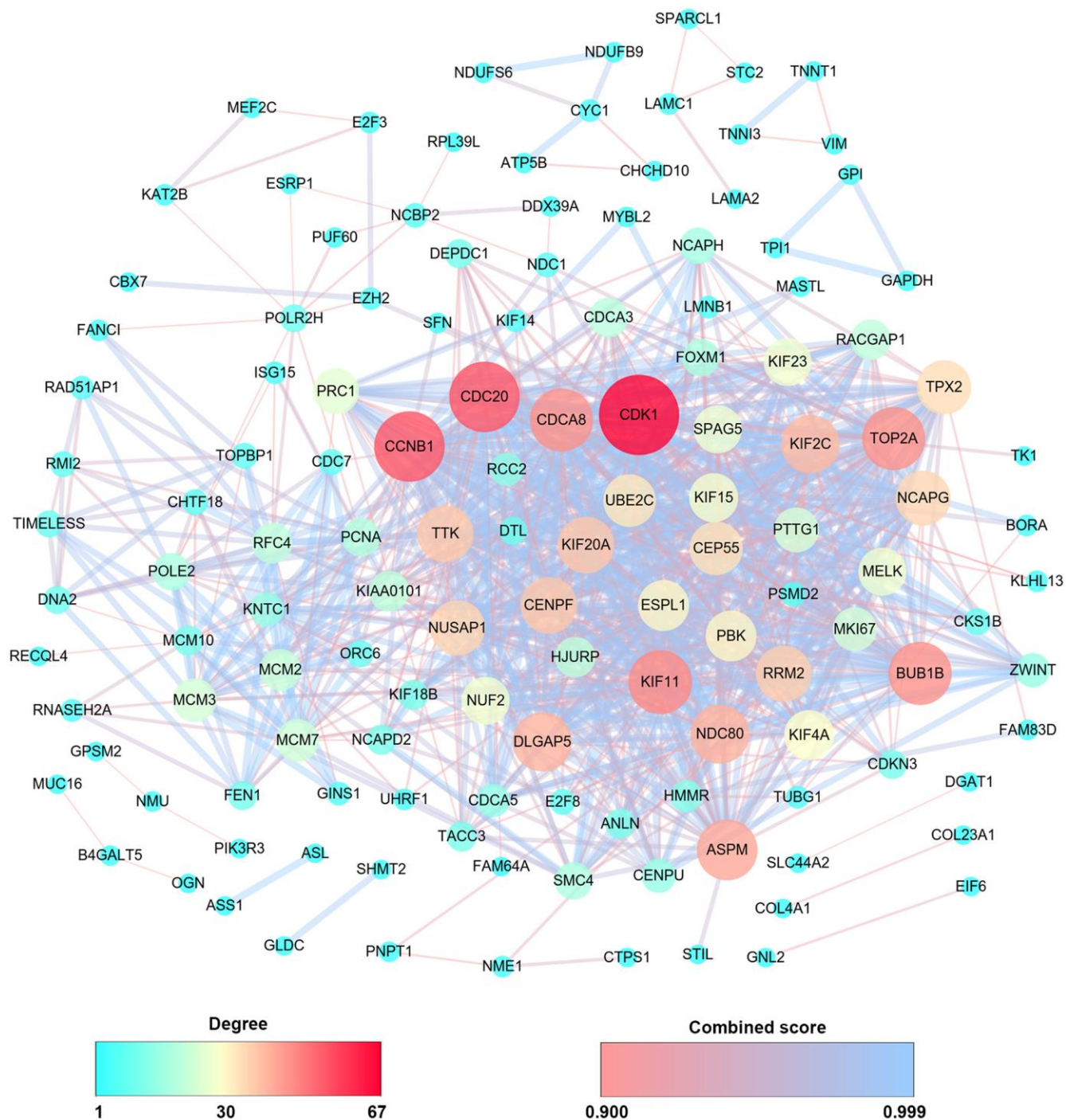


network of co-DEGs consists of 134 nodes with an estimated clustering coefficient of 0.671 and PPIs enrichment ( $p$ -value $<1.0e-16$ ). Moreover, the topological parameters of co-DEGs PPIs network were displayed in Table 1, including the Avg. clustering coefficient (Figure 4A), closeness centrality (Figure 4B), betweenness centrality (Figure 4C), shortest path length distribution

(Figure 4D), the distribution of the node degree (Figure 4E) and topological coefficients (Figure 4F).

### Core genes identification

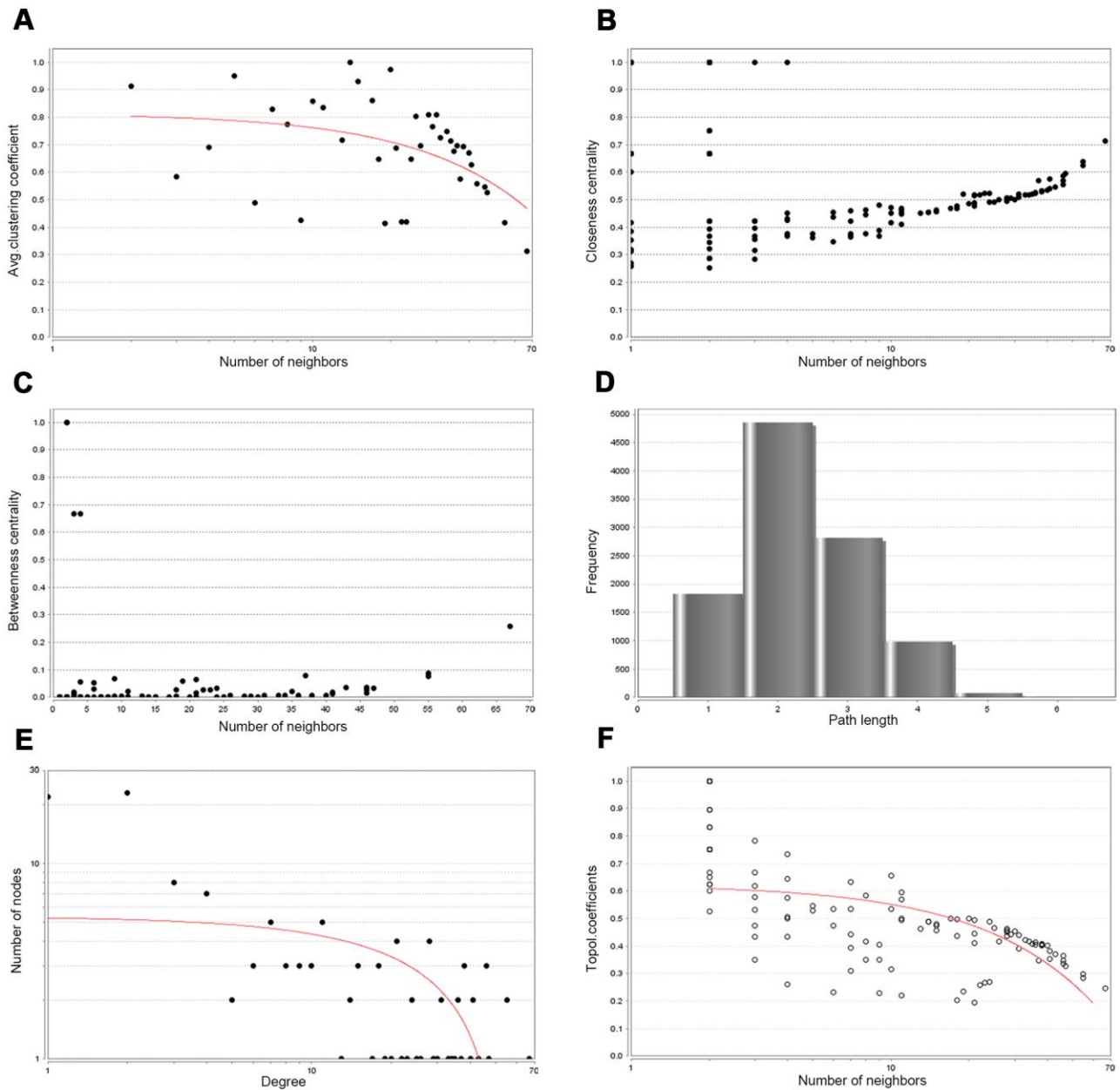
Next, Degree $>40$  was set as criterion to screen hub genes. Simultaneously, based on the module analysis, nine



**Figure 3. PPIs co-expression network from 279 co-DEGs.** The sphere color and size represent the degree of nodes with the line color indicating the combined score among them.

**Table 1. Topological parameters for co-DEGs PPIs network.**

Topological parameters	Comprehended values
Number of nodes	134
Clustering co-efficient	0.617
Network density	0.103
Network heterogeneity	1.098
Network centralization	0.407
Shortest paths	10572(59%)
Characteristic path length	2.303
Avg. number of neighbors	13.642



**Figure 4. Backbone network PPIs topology parameters.** (A) Avg. clustering coefficient. (B) Closeness centrality. (C) Betweenness centrality. (D) Shortest path length distribution. (E) Distribution of the node degree. (F) Topology coefficient.

significant modules were obtained (Supplementary Table 4). The seed nodes in these nine modules were also regarded as hub genes. According to the cut-off criteria, we collected 16 hub genes totally (Supplementary Table 5). Overall survival impacts of hub genes on patients with OSC at all stages were performed by Kaplan-Meier plotter (Supplementary Figure 3). A survival forest map for hub genes was shown as Supplementary Figure 3A and the hub genes-related survival curves were presented in Supplementary Figure 3B–3Q. As a result, we discovered 13 hub genes which were significantly associated with the overall survival of OSC patients among 16 hub genes, except for VIM [HR=0.92 (0.79–1.08), logrank  $P=0.3$ ] (Supplementary Figure 3C), SPARCL1 [HR=1.15 (0.96–1.37), logrank  $P=0.13$ ] (Supplementary Figure 3E), and CDCA8 [HR=1.14 (0.97–1.34), logrank  $P=0.13$ ] (Supplementary Figure 3N). Additionally, the expression levels of the hub genes in different pathological stages of ovary cancer samples were displayed in Supplementary Figure 4. The results indicated that among different pathological stages there had been notable alterations in the expression levels for NDC80 [Pr(>F)=0.0496] (Supplementary Figure 4F), MCM2 [Pr(>F)=0.000134] (Supplementary Figure 4G), KAT2B [Pr(>F)=0.0365] (Supplementary Figure 4I), CHTF18 [Pr(>F)=0.000685] (Supplementary Figure 4K), and BUB1B [Pr(>F)=0.00954] (Supplementary Figure 4P). The dynamic overall trends revealed that the expression of above hub genes decreased gradually with the continuous progression of ovary cancer (Supplementary Figure 4). Then, the GEPIA2 database was used to verify the expression levels of BUB1B, CHTF18, KAT2B, MCM2 and NDC80 (Supplementary Figure 5). As shown in Supplementary Figure 5, except for CDCA8 (Supplementary Figure 5E), the expression levels of the other genes (Supplementary Figure 5A–5D) were statistically significant between ovary cancer tissues and normal ovary tissue from TCGA (The Cancer Genome Atlas) and GTEx (the genotype-tissue expression) data. Consequently, BUB1B, MCM2, KAT2B and NDC80 were identified as the key genes in diagnosis and prognosis of OSC patients.

Then, the prognostic impact information of key genes on patients with OSC at different stages was explored by Kaplan-Meier plotter database (Figure 5). The key genes-related survival forest map at early stages (stages I+II) was shown in Figure 5A and relevant survival curves were respectively presented in Figure 5B–5E. The higher expression levels of NDC80 [HR=2.83 (1.19–6.73), logrank  $P=0.014$ ] (Figure 5B), MCM2 [HR=2.49 (1.04–5.95), logrank  $P=0.034$ ] (Figure 5C) and BUB1B [HR=2.82 (1.2–6.64), logrank  $P=0.013$ ] (Figure 5E) were notably related to poor overall survival at early stages in OSC patients. Meanwhile, the key genes-related survival forest map at advanced stages (stages

III+IV) was shown in Figure 5F and relevant survival curves were respectively presented in Figure 5G–5J, indicating that only one of four key genes was notably associated with the overall survival of OSC patients at advanced stages (Figure 5G–5J), in that high expression of MCM2 [HR=0.84 (0.71–1), logrank  $P=0.048$ ] (Figure 5H) was associated with improved overall survival in OSC patients at advanced stages. In summary, BUB1B and NDC80 were deemed as the core genes for early diagnosis of patients with OSC.

### **Association of expression patterns of core genes with EMT regulators**

The heatmap of expression levels of core genes and EMT regulators in GEO datasets revealed that the expressions of core genes and EMT regulators have exhibited significantly differences between OSC tissues and normal ovary tissues (Supplementary Figure 6). Further analyses indicated that correlation between core genes and EMT regulators were statistically significant (Figure 6A–6C,  $P<0.05$ ). And then STRING analyses for co-expression of core genes and EMT regulators verified that the core genes were closely related to the expression modulation of EMT regulators (Figure 6D). Meanwhile, functional KEGG enrichment analysis revealed that core genes and EMT regulators were mainly associated with adherens junction (hsa04520), Hippo signaling pathway (hsa04390), and Thyroid cancer (hsa05216) (Figure 6E and Supplementary Table 6). Furthermore, another online instrument NetworkAnalyst for ovary-specific PPIs network could also make out the notable roles of transcriptional regulation of core genes and EMT regulators in the development of ovarian cancer (Figure 6F).

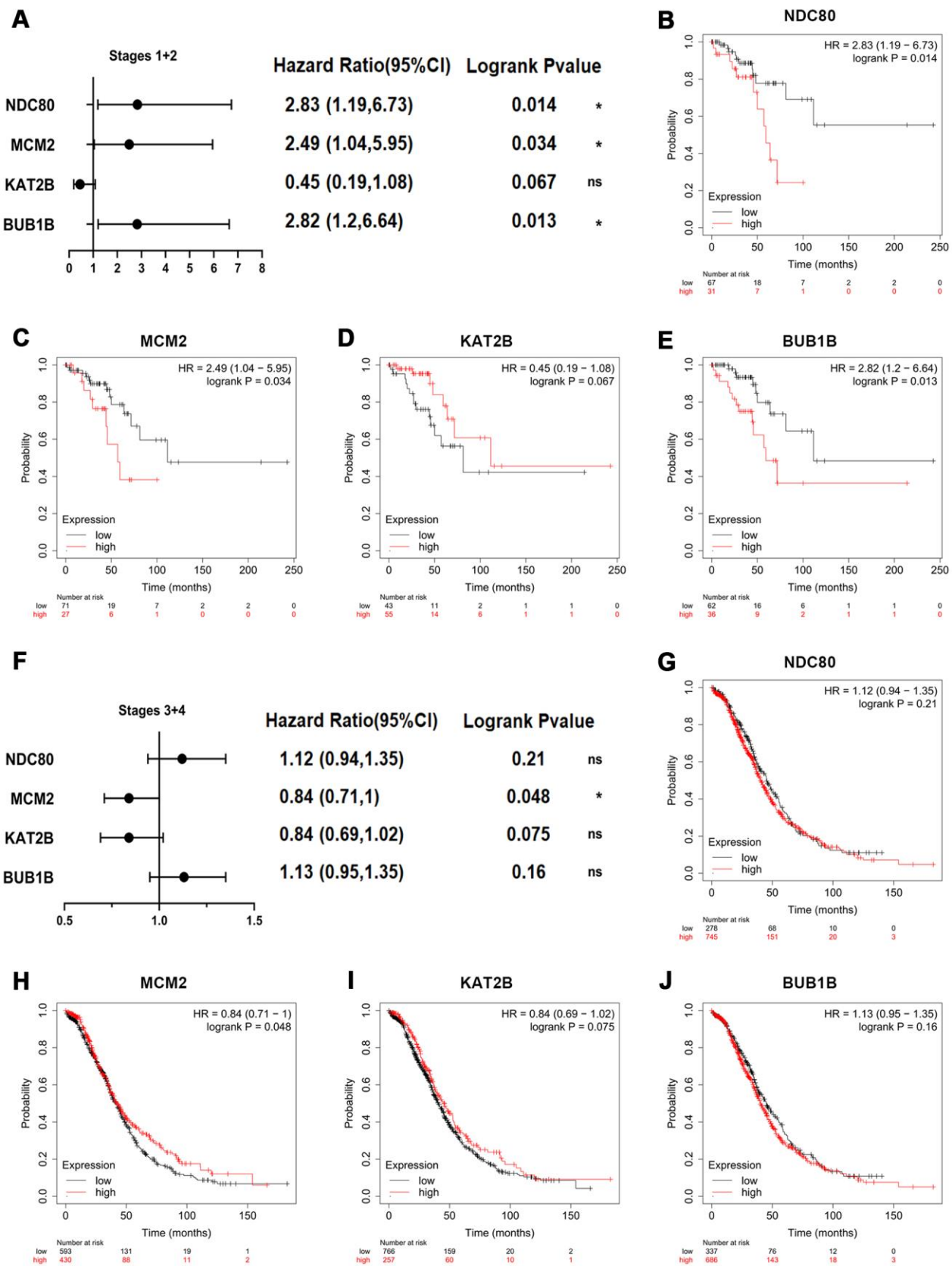
### **Characteristic alteration of core genes and EMT regulators**

The cBioPortal assay validated that variation, mutation count of the 2 core genes and 15 EMT regulators related to overall survival status were notably altered in 146 (47%) of queried patients or samples. Alterations in ELF3, including amplification and missense mutations (Supplementary Figure 7), were most often (8%) among them. Relevant gene amplification was accounted for the highest percentage among the different types of mutation, gene amplification, deep deletion and multiple alterations (Figure 6G).

### **GRN analysis for TF, miRNA, core genes and EMT regulators**

In order to further confirm the main functions of core genes and EMT regulators, the potential modulation relationship among core genes, EMT regulators and TFs

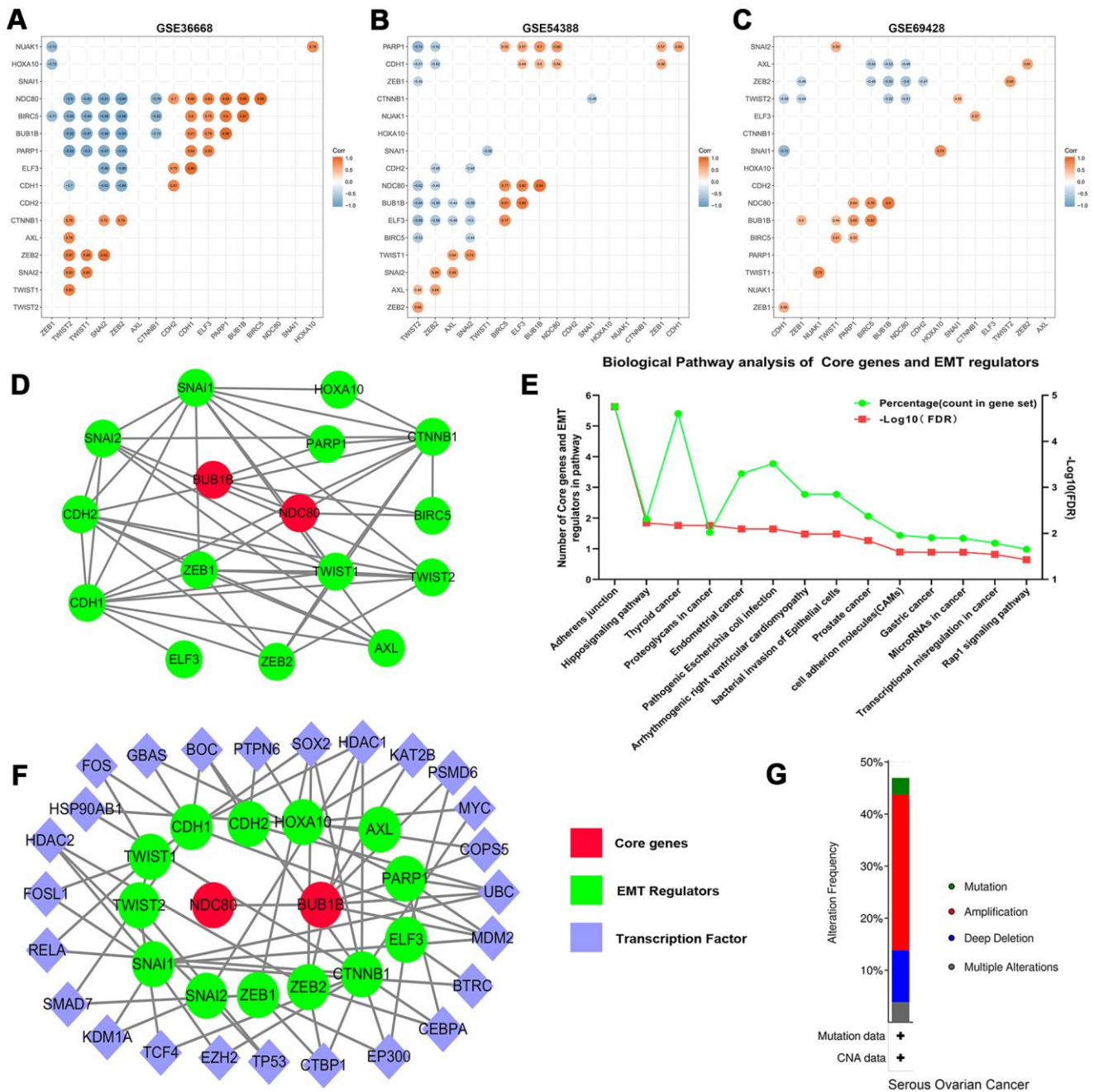




**Figure 5. Overall survival impact of key genes on patients with OSC at different stages. (A)** Survival prognosis forest map related to key genes at early stages (stages I+II) in OSC patients. Each point in the forest plot represents the Hazard ratio (HR) of the gene, and the



lines on both sides of the point represent the 95% confidence interval (95%CI). Survival curves were constructed by the Kaplan-Meier plotter based on the low and high expression of the key genes in OSC patients, including (B) NDC80, (C) MCM2, (D) KAT2B and (E) BUB1B. (F) Survival prognosis forest map related to key genes at advanced stages (stages III+IV) in OSC patients. Kaplan-Meier overall survival analysis for OSC patients with the expression of key genes, covering (G) NDC80, (H) MCM2, (I) KAT2B and (J) BUB1B. Logrank  $p$ -value<0.05 was considered statistically significant.



**Figure 6.** Correlation analysis of the core genes and EMT regulators in datasets (A) GES36668, (B) GES54388, (C) GSE69428. (D) The PPIs network of core genes and EMT regulators. The red color indicates core genes, and the green color predicts EMT regulators. (E) KEGG pathway enrichment analysis of core genes and EMT regulators. Only the enriched pathways with  $FDR < 0.05$  were presented. The green lines represent percentage of count in gene set and the red lines represent  $-\log_{10}(FDR)$ . (F) The ovary-specific PPIs integrated network of core genes and EMT regulators, the red color indicates core genes, the green color predicts EMT regulators and the purple color represents transcription factors. (G) An overview of the alteration of the core genes and EMT regulators in the genomics datasets of OSC in the TCGA database.

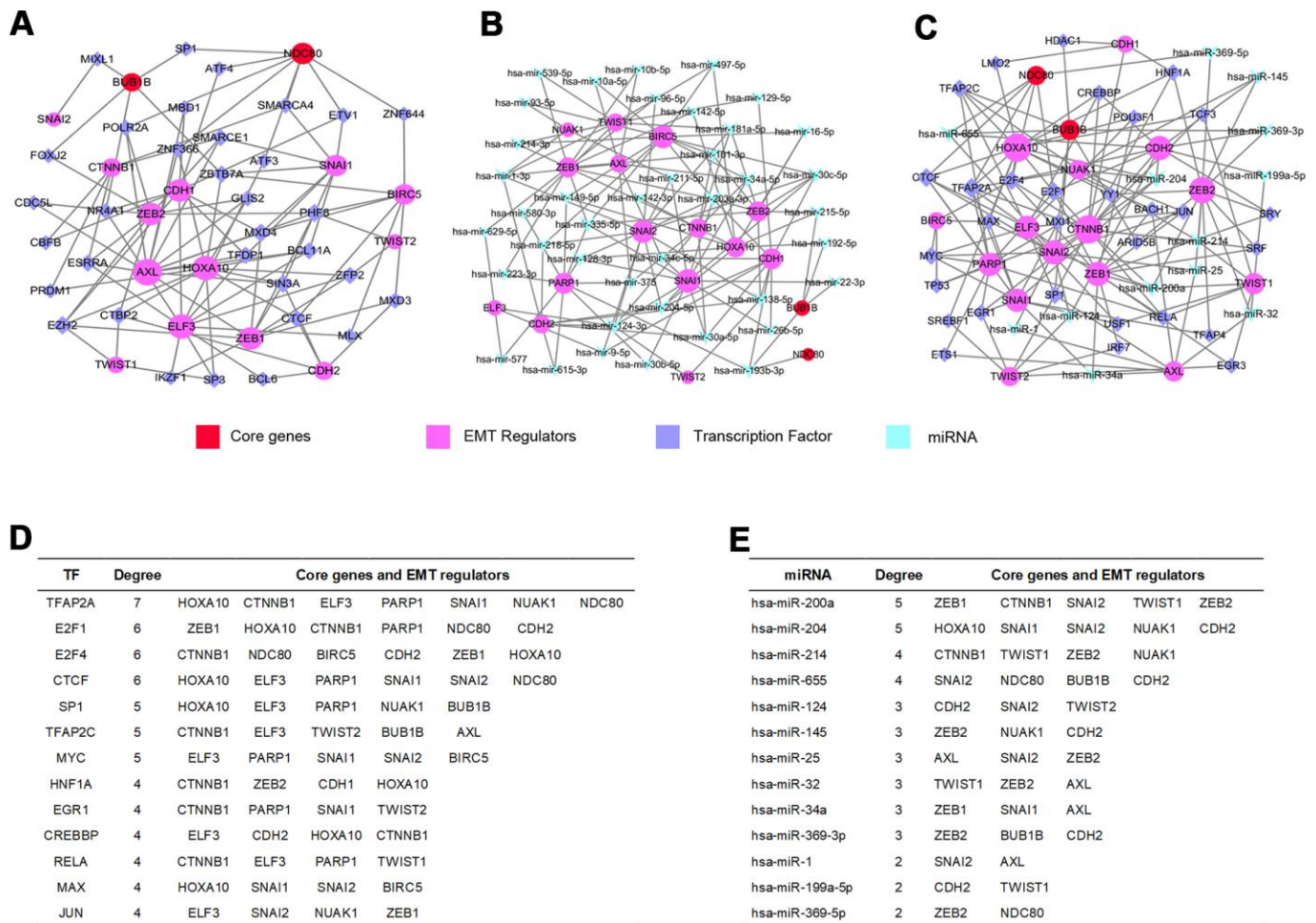
were discriminated based on TF and gene target data derived from the ENCODE ChIP-seq data (Figure 7A). Concurrently, a regulatory network among core genes, EMT regulators and miRNAs was compiled from the miRNA-gene interaction data collected from TarBase and miRTarBase (Figure 7B).

### TF-miRNA coregulatory interaction network

We constructed a TFs-miRNAs coregulatory network by collecting regulatory interaction information from the RegNetwork repository, containing 2 core genes, 15 EMT regulators, 30 TFs and 13 miRNAs with 152 edges (Figure 7C). The above analysis discovered that the transcription factor TFAP2A could monitor one core gene interacting with six EMT regulators (Figure 7D), and hsa-miR-655 could regulate two core genes (Figure 7E) among them.

### Core genes, EMT regulators, miRNA and TF validation

The expression of core genes including BUB1B (Supplementary Figure 8A) and NDC80 (Supplementary Figure 8B) were significantly increased in OSC patients among GSE36668, GSE54388 and GSE69428 datasets. BUB1B (Supplementary Figure 9A) and NDC80 (Supplementary Figure 9C) were validated at a transcription level in multiple cancer types based on the Oncomine database in that BUB1B (Median rank 132,  $p$ -value=1.54e-06) (Supplementary Figure 9B) and NDC80 (Median rank 192.5,  $p$ -value=4.04e-08) (Supplementary Figure 9D) were highly expressed in ovary cancer samples compared with normal ovary tissues. As for prognostic value of core TF, EMT regulator and miRNA, high expression of TFAP2A [HR=0.83[0.7-0.99], logrank  $P$ =0.033] was associated



**Figure 7. GRN analysis of core genes and EMT regulators.** (A) Network of TFs-core genes and EMT regulators was obtained from ENCODE database. (B) Network of miRNAs-core genes and EMT regulators was obtained from TarBase and miRTarBase database. (C) Integrative regulatory network of TFs-miRNAs-core genes and EMT regulators. (D) Core genes and EMT regulators regulated by TFs. (E) Core genes and EMT regulators modulated by miRNAs.

with the improved overall survival in OSC patients by Kaplan-Meier plotter (Figure 8A). However, the high expression of hsa-miR-655 [HR=1.68[1.33-2.13], logrank  $P=0.000013$ ] and ELF3 [HR=1.23[1.04-1.44], logrank  $P=0.014$ ] was linked with worse overall survival in OSC patients (Figure 8A). The expression levels of the EMT regulator and core TF were presented based on TCGA and GTEx data (Figure 8B, 8D), with consistent expression trend in three datasets (Supplementary Figure 8C, 8D). Moreover, the protein levels of EIF3 (Figure 8C) and TFAP2A (Figure 8E) were significantly higher in ovary cancer tissues than in normal ovary tissues based on HPA database. Furthermore, in the CCLE database, the expression levels of BUB1B (Supplementary Figure 10A), NDC80 (Supplementary Figure 10B), ELF3 (Supplementary Figure 10C), TFAP2A (Supplementary Figure 10D) and hsa-miR-655 (Supplementary Figure 11A) were confirmed among different ovarian cancer cell lines. Additional qRT-PCR also detected the consistent expression trend in SKOV3 cell line with CCLE (Supplementary Figure 10E–10H). Then, we explored the decreased expression of hsa-miR-655 and increased expression of hsa-miR-200a in ovarian cancer from the miRCancer database. At the same time, compared with normal blood samples, we also found the expression of hsa-miR-655 decreased and hsa-miR-200a increased in blood samples of patients with ovarian cancer (Supplementary Figure 11B, 11C and Supplementary Table 8).

All above-mentioned observations confirmed that BUB1B and NDC80 may be used as the key biomarkers at early stages for patients with OSC, and TFAP2A, ELF3 and hsa-miR-655 could play crucial roles in the EMT occurrence and pathological prognostic factors for OSC patients.

## DISCUSSION

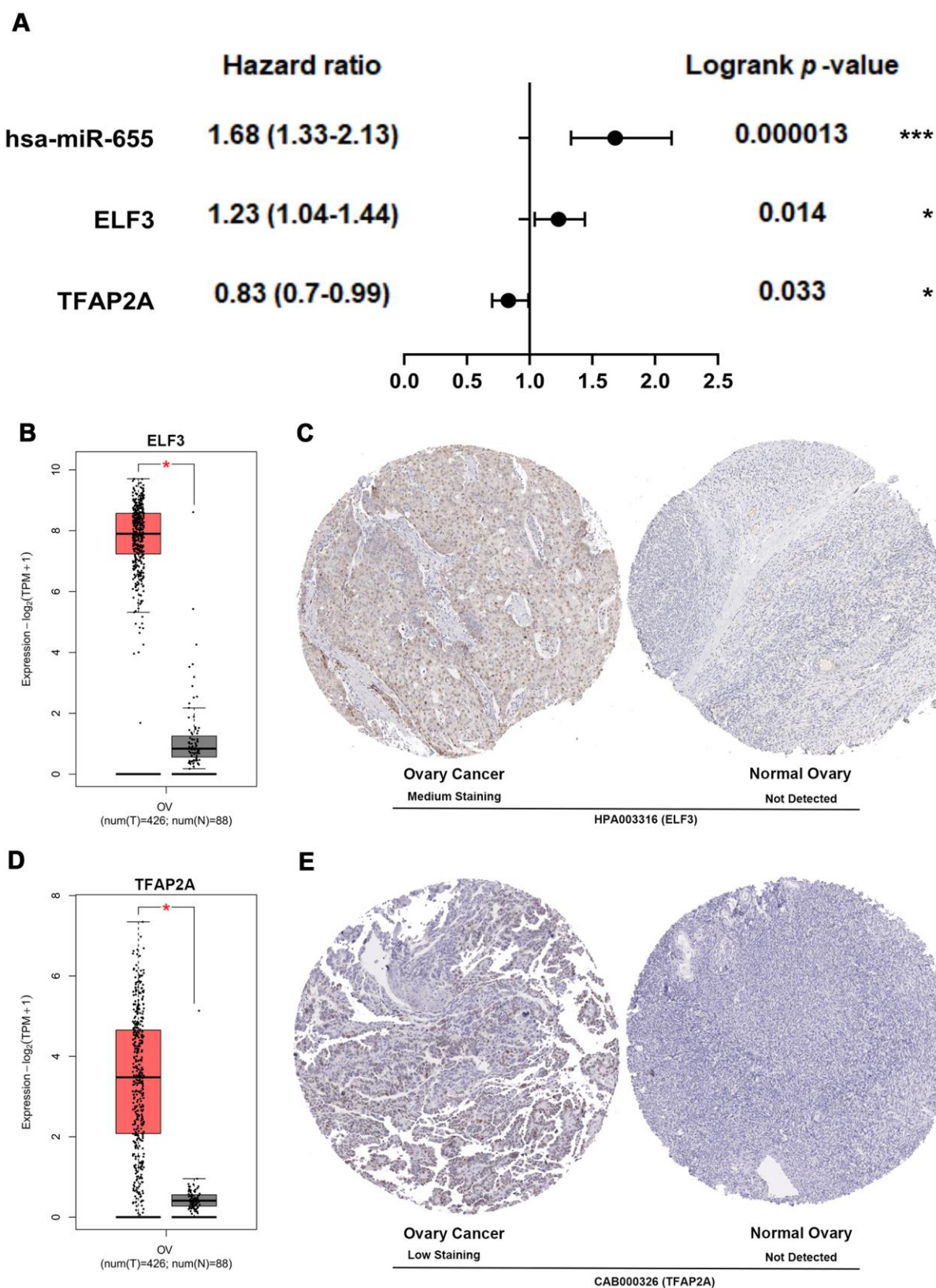
Since the ovary is located deep in the pelvis, about 70% of ovarian cancer cases have reached the advanced stage when they get diagnosed and distant metastases have already occurred, leading to its mortality rate ranking first among all kinds of gynecological malignant tumors [3]. Most patients with ovarian cancers may relapse after surgery or first line chemotherapy, and sometimes even after second line chemotherapy due to the ability of ovarian cancer stem cells to escape from these therapies or due to the reduced host immunosurveillance [24, 25]. Among them, OSC, as a common EOC and one of the most lethal gynecological tumors, accounts for 80–95% of ovarian malignancies [26]. Therefore, it is important to find reliable tumor biomarkers and explore the precise molecular mechanism of OSC for early clinical diagnosis, treatment and prognosis [27]. In recent years, with the

help of large-scale screening and the rapid development of bioinformatics, hundreds of genes alterations have been revealed to be closely related to the diagnosis, therapy and prognosis of tumors [22, 28, 29].

In our study, three datasets for OSC samples, namely GSE36668, GSE54388 and GSE69428, were selected and downloaded from GEO database. Then we set up comprehensively bioinformatic scheme to perform co-DEGs, GO and KEGG pathway functional enrichment, PPIs network and overall survival analysis at different pathological stages. Through the PPIs network regulation of co-DEGs and the prognostic investigation of the hub genes-related survival rates at different pathological stages, we identified two core genes (BUB1B and NDC80) associated with the early diagnosis and prognosis of OSC. One of them, BUB1B, a key component of the mitotic checkpoint complex, is localized to the kinetochore and plays a pivotal role in inhibiting anaphase-promoting complex/cyclosome (APC/C), delaying the onset of anaphase and ensuring proper chromosome segregation. In addition, BUB1B was also remarkably enriched in cell cycle and in Human T-cell leukemia virus (HTLV)-1 infection on KEGG pathway database [30, 31]. Another core gene, NDC80, encodes a component of the NDC80 kinetochore complex, which consists of a N-terminal microtubule binding domain and a C-terminal coiled-coiled domain able to interact with other components of the complex. Its molecular functions are to organize and stabilize microtubule-kinetochore interactions for proper chromosome segregation. And NDC80 was notably enriched in cell cycle, covering overall mitotic phase at the metaphase, anaphase, and prometaphase transition [32, 33], with impairment function of mitotic spindle checkpoint found in many types of cancer. Moreover, emerging evidence also suggests that dysregulation of cell cycle signaling cross-cascade is commonly been observed in a broad range of human cancers [34, 35]. Especially, relevant co-DEGs were also most striking in cell cycle pathway associated highly with the tumorigenesis and progression of ovary cancer [36, 37].

Since the EMT plays a complex role in tumor metastasis and recurrence by enhancing cell invasion and migration or other possible accesses [38], it is crucial to elucidate molecular mechanisms regulating EMT biological events for improving treatment of patients with OSC [11]. Consequently, we screened 15 EMT regulatory factors involved closely in the diverse pathological features of ovarian cancer. Then, we explored gene alteration characteristics and GRN of two core genes and 15 EMT regulators. Subsequently, it was demonstrated that the EMT regulator ELF3 was the most frequent in genetic variation, thus eliciting the





**Figure 8. Multidimensional validation and efficacy evaluation with the core TF, EMT regulator and miRNA. (A)** Survival prognosis forest map related to TFAP2A, hsa-miR-655 and ELF3 in patients with OSC. **(B)** The expression level of ELF3 from ovary cancer samples (red) and normal ovary samples (gray). **(C)** Validation of ELF3 from the HPA database. **(D)** The expression level of TFAP2A from ovary cancer samples and normal ovary samples. **(E)** Validation of TFAP2A from the HPA database.



development and relapse of OSC. Meanwhile, in the GRN analysis for core genes and EMT regulators, it was also found that TFAP2A and hsa-miR-655 could exert a crucial function in the EMT modulation of ovarian cancer.

What was also intriguing was that TFAP2A could modulate the expression of target EMT regulator ELF3, positively, via interacting with their transcription start site (TSS) through the TF-miRNA comprehensive regulation analysis. Simultaneously, it has been demonstrated that TFAP2A could influence the transcription of target genes involved in many different types of cancer [39]. For example, relevant studies indicated that TFAP2A promotes the proliferation, migration and invasion of breast cancer cells [40], and Shi et.al revealed that TFAP2A promotes nasopharyngeal carcinoma cell proliferation and inhibits apoptosis [41]. Furthermore, microRNAs, as endogenous transcripts with almost 22 nucleotides in length, are considered as possible causative agents in cancer [42]. However, previous studies have identified hsa-miR-655 as a novel EMT-suppressive microRNA [43, 44]. Additionally, there is emerging evidence that women with endometriosis have a higher risk of developing ovarian cancer due to original disease progression and malignant transformation. Namely, endometriosis may be one of the main causes of ovarian cancer [45].

Overall, by integrating multiple microarrays of gene expression profiles, BUB1B and NDC80 have been identified to be vital in early stages for OSC development. TF-miRNA comprehensive investigations for core genes and EMT regulators elucidated that TFAP2A, hsa-miR-655/200a and ELF3 could exert crucial function and prognostic potential in the development and progression of OSC. Eventually, our investigation into bioinformatics for the core genes related to EMT biological process in OSC may bring an unusual perspective for the early diagnosis and prognosis evaluation of patients with OSC.

## CONCLUSIONS

In summary, BUB1B and NDC80 activation could play a pivotal role in the occurrence and development of OSC at stages I+II, and thus might serve as early clinical diagnosis biomarkers for patients with OSC. Furthermore, the gene variation and GRN analysis revealed that ELF3, TFAP2A and hsa-miR-655/200a could collectively coordinate BUB1B and NDC80 to modulate EMT biological process on the development and progression of OSC, which may serve as the potential therapeutic target-points.

## MATERIALS AND METHODS

### Microarray data collection

The raw expression profiles of GSE36668 [46], GSE54388 [20] and GSE69428 [47] were downloaded from GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) based on microarray platform GPL570 (Affymetrix Human Genome U133 Plus 2.0Array). Details of each microarray data were provided in Supplementary Table 1. A microRNA expression profile (GSE31568) of blood was collected from ovarian cancer patients [48]. The dataset GSE31568 based on the platform of GPL9040 (febit Homo Sapiens miRBase 13.0) containing 15 samples with ovarian cancer and 70 samples without cancer was also downloaded from GEO database.

### Data preprocessing and Differential Expression Genes (DEGs) analysis

All data were normalized using NormalizeBetweenArray function from R package 'LIMMA' of the bioconductor project [49]. Data before and after normalization were shown in Supplementary Figure 1A–1C respectively. Next, we performed differential genes analyses ( $\text{LogFC} \geq 1$  or  $\text{LogFC} \leq -1$ , adjusted  $p$  value  $< 0.05$ ) by comparing OSC with normal ovary using 'LIMMA' R package.

### Screening co-DEGs and construction the PPIs network

The selected DEGs were separately uploaded to an online tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>), which could identify the co-DEGs among GSE36668, GSE54388 and GSE69428 datasets. Then, we utilized the online database STRING (Version 11.0, <https://string-db.org/>) to visualize the PPIs among co-DEGs [50]. To avoid an inaccurate PPIs network, we used a cutoff  $\geq 0.9$  (high-confidence interaction score) to obtain the striking PPIs and visualized in Cytoscape Version 3.6.1 [51]. Next, these most significant modules in the PPIs network were screened using MCODE, a package of Cytoscape, which could identify clusters in large protein networks according to the topology to build significant function modules. The criteria for selection included 5 aspects: MCODE score  $> 5$ , node score cut-off = 0.2, degree cut-off = 2, Max depth = 100, and k-score = 2.

### Analyzing the backbone network

The NetworkAnalyzer package in Cytoscape was utilized to explore the topological parameters and centrality measures such as the Avg. clustering coefficient, distribution of the node degree, topological coefficients, shortest path length distribution, betweenness centrality,

and closeness centrality for directed and undirected networks of co-DEGs PPIs backbone network.

### Functional enrichment analyses of co-DEGs

The co-DEGs were further analyzed via the Database for Annotation, Visualization and Integrated Discovery (DAVID, version 6.8, <https://david.ncifcrf.gov/>) to perform the GO and KEGG [52, 53]. The R ggplot2 package was adopted to visualize these data.

### Hub genes identification and analysis

The hub genes were selected for degree > 40 nodes or seed genes of PPIs network significant function modules. In order to prove the hub genes related to OSC prognosis, overall survival analyses were performed using Kaplan-Meier Plotter (<http://kmplot.com/analysis/>). Patients with OSC were categorized into high-expression group and low-expression group, according to the expression of specific genes. The overall survival related to hub genes was analyzed for OSC patients at all stages in above 2 groups. The analysis results were visualized in the forms of survival prognosis forest map and survival curves. Logrank  $p$ -value < 0.05 was regarded as statistically significant. Then, Gene Expression Profiling Interactive Analysis (GEPIA2, <http://gepia2.cancer-pku.cn/#index>) was performed to explore the alteration among ovary cancer samples at different pathological stages. ANOVA was accomplished to evaluate the statistical significance of variations.  $Pr(>F)$  < 0.05 was regarded as statistically significant.

### Key and core genes validated

We first identified the key genes from hub genes by Logrank  $p$ -value < 0.05 in overall survival and by  $Pr(>F)$  < 0.05 in different pathological stages of patients with ovary cancer. Furthermore, we also analyzed patients with OSC for key genes in high-expression and low-expression groups at early stages (stages I+II) and advanced stages (stages III+IV). Simultaneously, we utilized GEPIA2 to confirm the expression of the key genes between OSC tissues and normal ovary tissues. According to the Hazard ratio and Logrank  $p$ -value of key genes in overall survival analysis for early stages (stages I+II) and advanced stages (stages III+IV), we defined the core genes from the key genes for early diagnosis in patients with OSC.

### EMT regulators selection and analysis

In order to identify the EMT-related regulatory genes in OSC development, we have compiled 15 EMT-related regulatory factors from published literatures [11–15, 18–20]. Then we systematically evaluated the expressed

EMT-associated regulators and core genes in datasets GSE36668, GSE54388 and GSE69428, and then R ComplexHeatmap and dendextend packages were adopted to visualize them. Meanwhile, we used the R ggcorrplot package to estimate the correlation of core genes and EMT regulators. Then, Online database STRING was used to analysis the PPIs network and functional enrichment of core genes and EMT regulators. Next, we used another online tool NetworkAnalyst (<http://www.networkanalyst.ca/>) to visualize ovary specific PPIs network of core genes and EMT regulators. Furthermore, the online database cBioPortal for cancer genomics (<https://www.cbioportal.org/>) was used to analysis the genetic variation, mutation count and overall survival status related to the core genes and EMT regulators in OSC.

### GRN analysis of core genes and EMT regulators

We compiled TF, core gene and EMT regulator co-network and analyzed the GRN by uploading the core gene and EMT regulator to NetworkAnalyst. The TF and gene target data were derived from the ENCODE (Encyclopedia of DNA Elements) ChIP-seq data. Only those objects with peak intensity signal < 500 and the predicted regulatory potential score < 1 could be selected using BETA Minus algorithm. Next, we compiled miRNAs-core genes and EMT regulators co-network. The miRNA-gene interaction data validated by comprehensive experiments were collected from TarBase and miRTarBase. Soon after, we established the TF-miRNA integrated modulation network. Then, the integrated network was respectively visualized in Cytoscape to identify the core TF, EMT regulator and miRNA.

### Core genes, TF, EMT regulators and miRNA validation

We used Oncomine (<http://www.oncomine.com>) to evaluate the core genes on transcriptional level in multiple cancer types and relevant studies. The overall survival analyses related to core TF, EMT regulator and miRNA were performed using Kaplan-Meier Plotter. Then, we evaluated the significant core TF, EMT regulator expression among the GRN analysis and further validated using immunohistochemistry (IHC) from the Human Protein Atlas database (HPA, <https://www.proteinatlas.org/>). Simultaneously, we utilized the GEPIA2 to affirm the expression of core TF or EMT regulator between OSC tissues and normal ovary tissues.

### Experimental validation using quantitative real-time PCR in ovarian cancer lines

We explored the expression of core TF, EMT regulator and miRNA among different ovarian cancer cell lines in

Cancer Cell Line Encyclopedia (CCLE). Then, ovarian cancer cell line (SKOV3) was obtained from American Type Culture Collection (Manassas, VA, USA) and maintained in Roswell Park Memorial Institute (RPMI)-1640 medium (Sigma-Aldrich, St Louis, MO, USA) with 10% fetal bovine serum (FBS) in a humidified atmosphere containing 5% CO<sub>2</sub> at 37° C. Simultaneously, the human ovarian surface epithelium cells (HOSEC) were used as control. Total RNA was isolated by using a RNeasy Mini Kit (Qiagen) and cDNA was extracted by reverse transcription kit (Takara, Dalian, China). Gene expression was measured by qRT-PCR (Lightcycler96, Roche, Basel, Switzerland) using a SYBR Green™ Premix Ex Taq™ II (Takara, Dalian, China) and following the manufacturer's instructions. The primers used were shown in Supplementary Table 7.

### Statistical analyses

The significances of differences between two groups were analyzed using non-parametric test or *t*-test based on data distribution characteristics in Graphad Prism 8. The log-rank test was used to identify the differences in overall survival rate at different stages between low-expression and high-expression groups of hub genes using Kaplan-Meier Plots. Logrank *p*-value<0.05 was considered statistically significant in survival rate. Correlation analysis was calculated using R ggcorrplot package. All analyses were conducted using software R Studio 3.5.3. *P*-value < 0.05 was considered statistically significant.

### Abbreviations

OSC: ovarian serous cancer/carcinoma; EMT: epithelial-mesenchymal transformation; EOC: epithelial ovarian cancer; GEO: Gene Expression Omnibus; co-DEGs: common differential expression genes; PPIs: Protein-protein interactions; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Gene and Genome; GRN: Gene Regulatory Networks; TF: transcription factor; ELF3: E74 like ETS transcription factor 3; TFAP2A: transcription factor AP-2 alpha; BUB1B: BUB1 mitotic checkpoint serine/threonine kinase B; NDC80: NDC80 kinetochore complex component; CDH1/E-cadherin: cadherin 1; CDH2/N-cadherin: cadherin 2; ZEB1/2: zinc finger E-box binding homeobox 1/2; SNAI1/2: snail family transcriptional repressor 1/2; TWIST1/2: twist family bHLH transcription factor 1/2; AXL: AXL receptor tyrosine kinase; BIRC5: baculoviral IAP repeat containing 5; CTNBB1: catenin beta 1; PARP-1: poly(ADP-ribose) polymerase 1; ARK5/NUAK1: NUAK family kinase 1; HOXA10: homeobox A10.

## AUTHOR CONTRIBUTIONS

YN.Z and XC.C conceived the study. XC.C supervised the whole project. YN.Z and X.W, performed the data curation and analysis. YN.Z wrote the manuscript, XC.C participated in the manuscript editing and discussion.

## ACKNOWLEDGMENTS

We thank GEO for data sharing.

## CONFLICTS OF INTEREST

The authors disclose no potential conflicts of interest.

## FUNDING

This investigation was supported by High-tech Research and Development Program (2014ZX09101041-002) and the Chinese Postdoctoral Science Foundation (2020M673232).

## REFERENCES

1. Della Pepa C, Tonini G, Pisano C, Di Napoli M, Cecere SC, Tambaro R, Facchini G, Pignata S. Ovarian cancer standard of care: are there real alternatives? *Chin J Cancer*. 2015; 34:17–27. <https://doi.org/10.5732/cjc.014.10274> PMID:25556615
2. Bast RC Jr, Hennessy B, Mills GB. The biology of ovarian cancer: new opportunities for translation. *Nat Rev Cancer*. 2009; 9:415–28. <https://doi.org/10.1038/nrc2644> PMID:19461667
3. Herzog TJ. Recurrent ovarian cancer: how important is it to treat to disease progression? *Clin Cancer Res*. 2004; 10:7439–49. <https://doi.org/10.1158/1078-0432.CCR-04-0683> PMID:15569973
4. Beauchamp MC, Yasmeen A, Knafo A, Gotlieb WH. Targeting insulin and insulin-like growth factor pathways in epithelial ovarian cancer. *J Oncol*. 2010; 2010:257058. <https://doi.org/10.1155/2010/257058> PMID:20069126
5. Langdon SP, Herrington CS, Hollis RL, Gourley C. Estrogen signaling and its potential as a target for therapy in ovarian cancer. *Cancers (Basel)*. 2020; 12:1647. <https://doi.org/10.3390/cancers12061647> PMID:32580290
6. Teeuwssen M, Fodde R. Wnt signaling in ovarian cancer stemness, EMT, and therapy resistance. *J Clin Med*. 2019; 8:1658.

- <https://doi.org/10.3390/jcm8101658>  
PMID:[31614568](https://pubmed.ncbi.nlm.nih.gov/31614568/)
7. Bates RC, Mercurio AM. The epithelial-mesenchymal transition (EMT) and colorectal cancer progression. *Cancer Biol Ther.* 2005; 4:365–70.  
<https://doi.org/10.4161/cbt.4.4.1655> PMID:[15846061](https://pubmed.ncbi.nlm.nih.gov/15846061/)
  8. Thiery JP, Sleeman JP. Complex networks orchestrate epithelial-mesenchymal transitions. *Nat Rev Mol Cell Biol.* 2006; 7:131–42.  
<https://doi.org/10.1038/nrm1835> PMID:[16493418](https://pubmed.ncbi.nlm.nih.gov/16493418/)
  9. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol.* 2019; 20:69–84.  
<https://doi.org/10.1038/s41580-018-0080-4>  
PMID:[30459476](https://pubmed.ncbi.nlm.nih.gov/30459476/)
  10. Nawshad A, Lagamba D, Polad A, Hay ED. Transforming growth factor-beta signaling during epithelial-mesenchymal transformation: implications for embryogenesis and tumor metastasis. *Cells Tissues Organs.* 2005; 179:11–23.  
<https://doi.org/10.1159/000084505> PMID:[15942189](https://pubmed.ncbi.nlm.nih.gov/15942189/)
  11. Lutgendorf SK, Penedo F, Goodheart MJ, Dahmouch L, Arevalo JM, Thaker PH, Slavich GM, Sood AK, Cole SW. Epithelial-mesenchymal transition polarization in ovarian carcinomas from patients with high social isolation. *Cancer.* 2020; 126:4407–13.  
<https://doi.org/10.1002/cncr.33060> PMID:[32691853](https://pubmed.ncbi.nlm.nih.gov/32691853/)
  12. Weingarten C, Jenudi Y, Tshuva RY, Moskovich D, Alfandari A, Hercbergs A, Davis PJ, Ellis M, Ashur-Fabian O. The interplay between epithelial-mesenchymal transition (EMT) and the thyroid hormones- $\alpha\beta$  axis in ovarian cancer. *Horm Cancer.* 2018; 9:22–32.  
<https://doi.org/10.1007/s12672-017-0316-3>  
PMID:[29260382](https://pubmed.ncbi.nlm.nih.gov/29260382/)
  13. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, Van Keymeulen A, Brown D, Moers V, Lemaire S, De Clercq S, Minguijón E, Balsat C, et al. Identification of the tumour transition states occurring during EMT. *Nature.* 2018; 556:463–68.  
<https://doi.org/10.1038/s41586-018-0040-3>  
PMID:[29670281](https://pubmed.ncbi.nlm.nih.gov/29670281/)
  14. Antony J, Tan TZ, Kelly Z, Low J, Choolani M, Recchi C, Gabra H, Thiery JP, Huang RY. The GAS6-AXL signaling network is a mesenchymal (Mes) molecular subtype-specific therapeutic target for ovarian cancer. *Sci Signal.* 2016; 9:ra97.  
<https://doi.org/10.1126/scisignal.aaf8175>  
PMID:[27703030](https://pubmed.ncbi.nlm.nih.gov/27703030/)
  15. Zhang HY, Li JH, Li G, Wang SR. Activation of ARK5/miR-1181/HOXA10 axis promotes epithelial-mesenchymal transition in ovarian cancer. *Oncol Rep.* 2015; 34:1193–202.  
<https://doi.org/10.3892/or.2015.4113> PMID:[26151663](https://pubmed.ncbi.nlm.nih.gov/26151663/)
  16. Wu D, He X, Wang W, Hu X, Wang K, Wang M. Long noncoding RNA SNHG12 induces proliferation, migration, epithelial-mesenchymal transition, and stemness of esophageal squamous cell carcinoma cells via post-transcriptional regulation of BMI1 and CTNNB1. *Mol Oncol.* 2020; 14:2332–51.  
<https://doi.org/10.1002/1878-0261.12683>  
PMID:[32239639](https://pubmed.ncbi.nlm.nih.gov/32239639/)
  17. Lin Y, Yang Z, Xu A, Dong P, Huang Y, Liu H, Li F, Wang H, Xu Q, Wang Y, Sun D, Zou Y, Zou X, et al. PIK3R1 negatively regulates the epithelial-mesenchymal transition and stem-like phenotype of renal cancer cells through the AKT/GSK3 $\beta$ /CTNNB1 signaling pathway. *Sci Rep.* 2015; 5:8997.  
<https://doi.org/10.1038/srep08997> PMID:[25757764](https://pubmed.ncbi.nlm.nih.gov/25757764/)
  18. Makvandi M, Pantel A, Schwartz L, Schubert E, Xu K, Hsieh CJ, Hou C, Kim H, Weng CC, Winters H, Doot R, Farwell MD, Pryma DA, et al. A PET imaging agent for evaluating PARP-1 expression in ovarian cancer. *J Clin Invest.* 2018; 128:2116–26.  
<https://doi.org/10.1172/JCI97992> PMID:[29509546](https://pubmed.ncbi.nlm.nih.gov/29509546/)
  19. Su S, Lin X, Ding N, Zhang H, Zhang Q, Ding Y, Hou X, Tian Y. Effects of PARP-1 inhibitor and ERK inhibitor on epithelial mesenchymal transitions of the ovarian cancer SKOV3 cells. *Pharmacol Rep.* 2016; 68:1225–29.  
<https://doi.org/10.1016/j.pharep.2016.08.001>  
PMID:[27668317](https://pubmed.ncbi.nlm.nih.gov/27668317/)
  20. Yeung TL, Leung CS, Wong KK, Gutierrez-Hartmann A, Kwong J, Gershenson DM, Mok SC. ELF3 is a negative regulator of epithelial-mesenchymal transition in ovarian cancer cells. *Oncotarget.* 2017; 8:16951–63.  
<https://doi.org/10.18632/oncotarget.15208>  
PMID:[28199976](https://pubmed.ncbi.nlm.nih.gov/28199976/)
  21. Li Y, Wang Q, Ning N, Tang F, Wang Y. Bioinformatic analysis reveals MIR502 as a potential tumour suppressor in ovarian cancer. *J Ovarian Res.* 2020; 13:77.  
<https://doi.org/10.1186/s13048-020-00683-y>  
PMID:[32660514](https://pubmed.ncbi.nlm.nih.gov/32660514/)
  22. Wu C, He L, Wei Q, Li Q, Jiang L, Zhao L, Wang C, Li J, Wei M. Bioinformatic profiling identifies a platinum-resistant-related risk signature for ovarian cancer. *Cancer Med.* 2020; 9:1242–53.  
<https://doi.org/10.1002/cam4.2692> PMID:[31856408](https://pubmed.ncbi.nlm.nih.gov/31856408/)
  23. Gortzak-Uzan L, Ignatchenko A, Evangelou AI, Agochiya M, Brown KA, St Onge P, Kireeva I, Schmitt-Ulms G, Brown TJ, Murphy J, Rosen B, Shaw P, Jurisica I, Kislinger T. A proteome resource of ovarian cancer ascites: integrated proteomic and bioinformatic

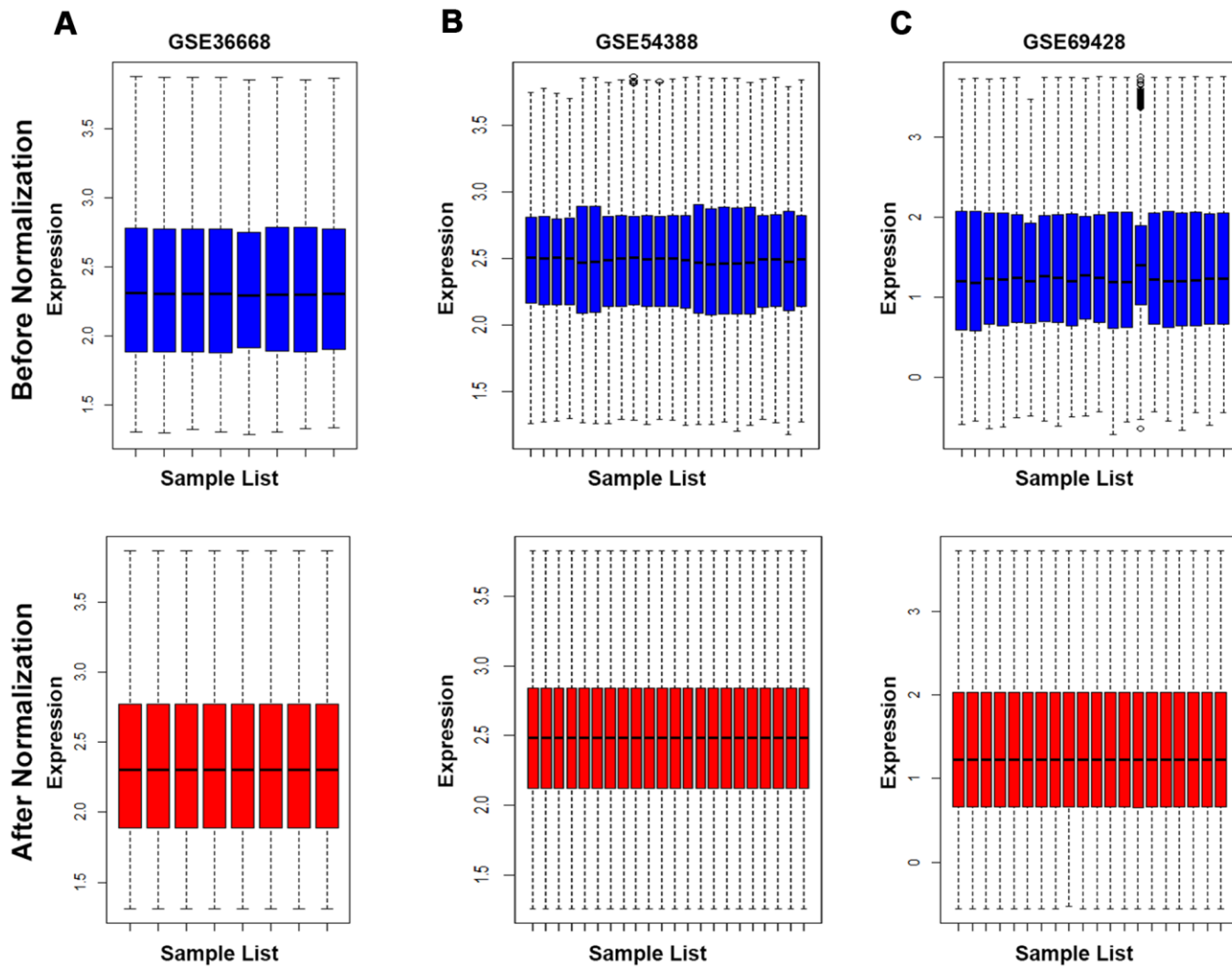


- analyses to identify putative biomarkers. *J Proteome Res.* 2008; 7:339–51.  
<https://doi.org/10.1021/pr0703223> PMID:[18076136](https://pubmed.ncbi.nlm.nih.gov/18076136/)
24. Laganà AS, Sofo V, Vitale SG, Triolo O. Epithelial ovarian cancer inherent resistance: may the pleiotropic interaction between reduced immunosurveillance and drug-resistant cells play a key role? *Gynecol Oncol Rep.* 2016; 18:57–58.  
<https://doi.org/10.1016/j.gore.2016.09.004> PMID:[27995176](https://pubmed.ncbi.nlm.nih.gov/27995176/)
  25. Barra F, Laganà AS, Ghezzi F, Casarin J, Ferrero S. Nintedanib for advanced epithelial ovarian cancer: a change of perspective? summary of evidence from a systematic review. *Gynecol Obstet Invest.* 2019; 84:107–17.  
<https://doi.org/10.1159/000493361> PMID:[30304728](https://pubmed.ncbi.nlm.nih.gov/30304728/)
  26. Penny SM. Ovarian Cancer: An Overview. *Radiol Technol.* 2020; 91:561–75. PMID:[32606233](https://pubmed.ncbi.nlm.nih.gov/32606233/)
  27. Laganà AS, Colonese F, Colonese E, Sofo V, Salmeri FM, Granese R, Chiofalo B, Ciancimino L, Triolo O. Cytogenetic analysis of epithelial ovarian cancer's stem cells: an overview on new diagnostic and therapeutic perspectives. *Eur J Gynaecol Oncol.* 2015; 36:495–505. PMID:[26513872](https://pubmed.ncbi.nlm.nih.gov/26513872/)
  28. Wang Q, Wang Z, Bao Z, Zhang C, Wang Z, Jiang T. PABPC1 relevant bioinformatic profiling and prognostic value in gliomas. *Future Oncol.* 2020; 16:4279–88.  
<https://doi.org/10.2217/fon-2019-0268> PMID:[31797689](https://pubmed.ncbi.nlm.nih.gov/31797689/)
  29. Zhou X, Fan Y, Ye W, Jia B, Yang Y, Liu Y. Identification of the novel target genes for osteosarcoma therapy based on comprehensive bioinformatic analysis. *DNA Cell Biol.* 2020; 39:1172–80.  
<https://doi.org/10.1089/dna.2020.5377> PMID:[32584170](https://pubmed.ncbi.nlm.nih.gov/32584170/)
  30. Ma Q, Liu Y, Shang L, Yu J, Qu Q. The FOXM1/BUB1B signaling pathway is essential for the tumorigenicity and radioresistance of glioblastoma. *Oncol Rep.* 2017; 38:3367–75.  
<https://doi.org/10.3892/or.2017.6032> PMID:[29039578](https://pubmed.ncbi.nlm.nih.gov/29039578/)
  31. Lee E, Pain M, Wang H, Herman JA, Toledo CM, DeLuca JG, Yong RL, Paddison P, Zhu J. Sensitivity to BUB1B inhibition defines an alternative classification of glioblastoma. *Cancer Res.* 2017; 77:5518–29.  
<https://doi.org/10.1158/0008-5472.CAN-17-0736> PMID:[28855212](https://pubmed.ncbi.nlm.nih.gov/28855212/)
  32. Kim JO, Zelter A, Umbreit NT, Bollozos A, Riffle M, Johnson R, MacCoss MJ, Asbury CL, Davis TN. The Ndc80 complex bridges two Dam1 complex rings. *Elife.* 2017; 6:e21069.  
<https://doi.org/10.7554/eLife.21069> PMID:[28191870](https://pubmed.ncbi.nlm.nih.gov/28191870/)
  33. Bharadwaj R, Qi W, Yu H. Identification of two novel components of the human NDC80 kinetochore complex. *J Biol Chem.* 2004; 279:13076–85.  
<https://doi.org/10.1074/jbc.M310224200> PMID:[14699129](https://pubmed.ncbi.nlm.nih.gov/14699129/)
  34. da Silva JN, Ranzi AD, Carvalho CT, Scheide TV, Strey YT, Graziottin TM, Bica CG. Cell cycle markers in the evaluation of bladder cancer. *Pathol Oncol Res.* 2020; 26:175–81.  
<https://doi.org/10.1007/s12253-018-0389-5> PMID:[29524168](https://pubmed.ncbi.nlm.nih.gov/29524168/)
  35. Yang B, Bai H, Sa Y, Zhu P, Liu P. Inhibiting EMT, stemness and cell cycle involved in baicalin-induced growth inhibition and apoptosis in colorectal cancer cells. *J Cancer.* 2020; 11:2303–17.  
<https://doi.org/10.7150/jca.37242> PMID:[32127957](https://pubmed.ncbi.nlm.nih.gov/32127957/)
  36. Jiang L, Hou R. Tetrandrine reverses paclitaxel resistance in human ovarian cancer via inducing apoptosis, cell cycle arrest through  $\beta$ -catenin pathway. *Onco Targets Ther.* 2020; 13:3631–39.  
<https://doi.org/10.2147/OTT.S235533> PMID:[32431514](https://pubmed.ncbi.nlm.nih.gov/32431514/)
  37. Karagoz K, Mehta GA, Khella CA, Khanna P, Gatza ML. Integrative proteogenomic analyses of human tumours identifies ADNP as a novel oncogenic mediator of cell cycle progression in high-grade serous ovarian cancer with poor prognosis. *EBioMedicine.* 2019; 50:191–202.  
<https://doi.org/10.1016/j.ebiom.2019.11.009> PMID:[31767542](https://pubmed.ncbi.nlm.nih.gov/31767542/)
  38. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: acquisition of Malignant and stem cell traits. *Nat Rev Cancer.* 2009; 9:265–73.  
<https://doi.org/10.1038/nrc2620> PMID:[19262571](https://pubmed.ncbi.nlm.nih.gov/19262571/)
  39. Kołat D, Kałuzińska Ż, Bednarek AK, Płuciennik E. The biological characteristics of transcription factors AP-2 $\alpha$  and AP-2 $\gamma$  and their importance in various types of cancers. *Biosci Rep.* 2019; 39:BSR20181928.  
<https://doi.org/10.1042/BSR20181928> PMID:[30824562](https://pubmed.ncbi.nlm.nih.gov/30824562/)
  40. Allouche A, Nolens G, Tancredi A, Delacroix L, Mardaga J, Fridman V, Winkler R, Boniver J, Delvenne P, Begon DY. The combined immunodetection of AP-2 $\alpha$  and YY1 transcription factors is associated with ERBB2 gene overexpression in primary breast tumors. *Breast Cancer Res.* 2008; 10:R9.  
<https://doi.org/10.1186/bcr1851> PMID:[18218085](https://pubmed.ncbi.nlm.nih.gov/18218085/)
  41. Shi D, Xie F, Zhang Y, Tian Y, Chen W, Fu L, Wang J, Guo W, Kang T, Huang W, Deng W. TFAP2A regulates nasopharyngeal carcinoma growth and survival by targeting HIF-1 $\alpha$  signaling pathway. *Cancer Prev Res (Phila).* 2014; 7:266–77.  
<https://doi.org/10.1158/1940-6207.CAPR-13-0271> PMID:[24335623](https://pubmed.ncbi.nlm.nih.gov/24335623/)

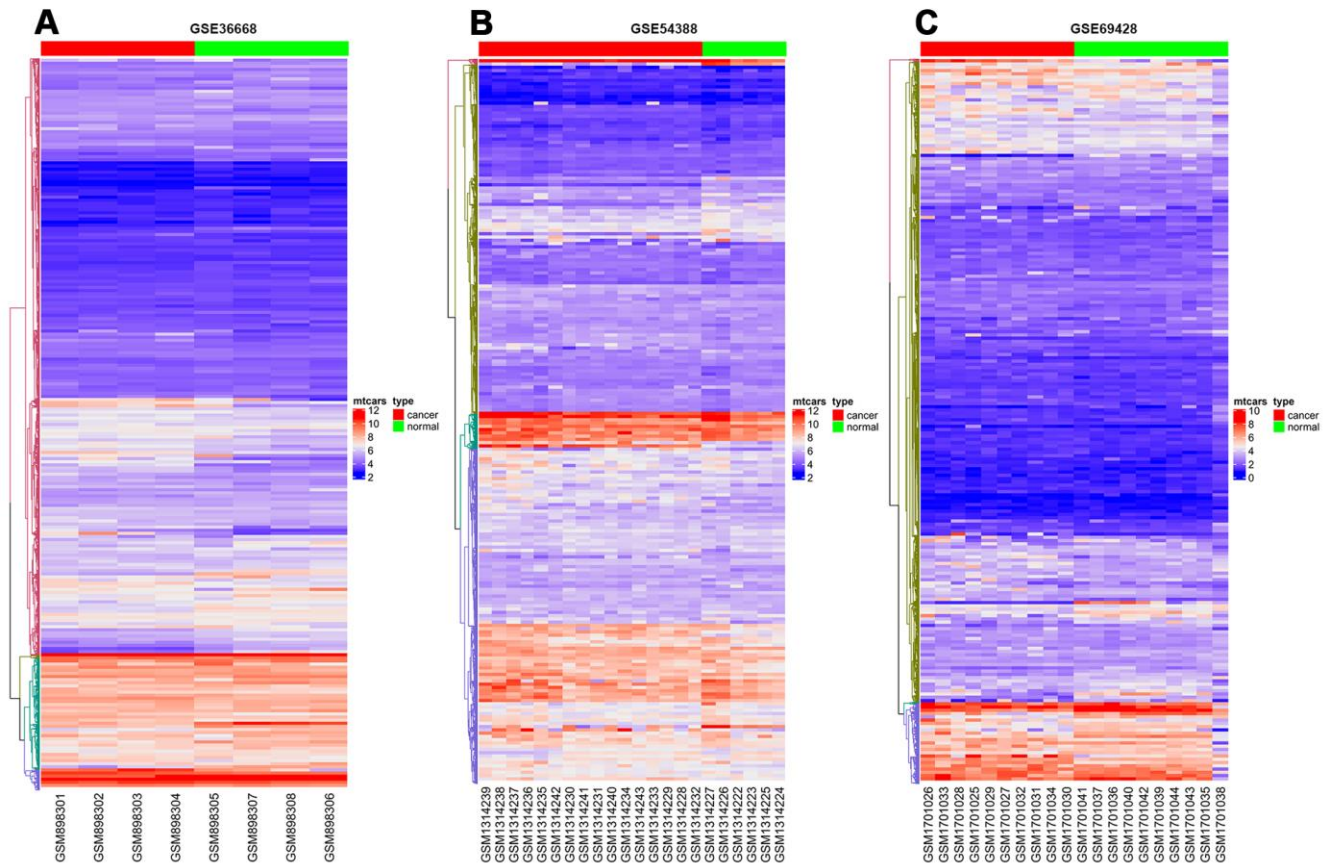
42. Liu Y, Cai Q, Bao PP, Su Y, Cai H, Wu J, Ye F, Guo X, Zheng W, Zheng Y, Shu XO. Tumor tissue microRNA expression in association with triple-negative breast cancer outcomes. *Breast Cancer Res Treat.* 2015; 152:183–91.  
<https://doi.org/10.1007/s10549-015-3460-x>  
PMID:26062749
43. Harazono Y, Muramatsu T, Endo H, Uzawa N, Kawano T, Harada K, Inazawa J, Kozaki K. miR-655 is an EMT-suppressive microRNA targeting ZEB1 and TGFBR2. *PLoS One.* 2013; 8:e62757.  
<https://doi.org/10.1371/journal.pone.0062757>  
PMID:23690952
44. Lv ZD, Kong B, Liu XP, Jin LY, Dong Q, Li FN, Wang HB. miR-655 suppresses epithelial-to-mesenchymal transition by targeting Prrx1 in triple-negative breast cancer. *J Cell Mol Med.* 2016; 20:864–73.  
<https://doi.org/10.1111/jcmm.12770>  
PMID:26820102
45. Králíčková M, Laganà AS, Ghezzi F, Vetvicka V. Endometriosis and risk of ovarian cancer: what do we know? *Arch Gynecol Obstet.* 2020; 301:1–10.  
<https://doi.org/10.1007/s00404-019-05358-8>  
PMID:31745637
46. Elgaaen BV, Olstad OK, Sandvik L, Odegaard E, Sauer T, Staff AC, Gautvik KM. ZNF385B and VEGFA are strongly differentially expressed in serous ovarian carcinomas and correlate with survival. *PLoS One.* 2012; 7:e46317.  
<https://doi.org/10.1371/journal.pone.0046317>  
PMID:23029477
47. Yamamoto Y, Ning G, Howitt BE, Mehra K, Wu L, Wang X, Hong Y, Kern F, Wei TS, Zhang T, Nagarajan N, Basuli D, Torti S, et al. *In vitro* and *in vivo* correlates of physiological and neoplastic human Fallopian tube stem cells. *J Pathol.* 2016; 238:519–30.  
<https://doi.org/10.1002/path.4649> PMID:26415052
48. Keller A, Leidinger P, Bauer A, Elsharawy A, Haas J, Backes C, Wendschlag A, Giese N, Tjaden C, Ott K, Werner J, Hackert T, Ruprecht K, et al. Toward the blood-borne miRNome of human diseases. *Nat Methods.* 2011; 8:841–43.  
<https://doi.org/10.1038/nmeth.1682> PMID:21892151
49. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics.* 2005; 21:2067–75.  
<https://doi.org/10.1093/bioinformatics/bti270>  
PMID:15657102
50. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019; 47:D607–13.  
<https://doi.org/10.1093/nar/gky1131> PMID:30476243
51. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13:2498–504.  
<https://doi.org/10.1101/gr.1239303> PMID:14597658
52. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009; 37:1–13.  
<https://doi.org/10.1093/nar/gkn923>  
PMID:19033363
53. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009; 4:44–57.  
<https://doi.org/10.1038/nprot.2008.211>  
PMID:19131956

# SUPPLEMENTARY MATERIALS

## Supplementary Figures

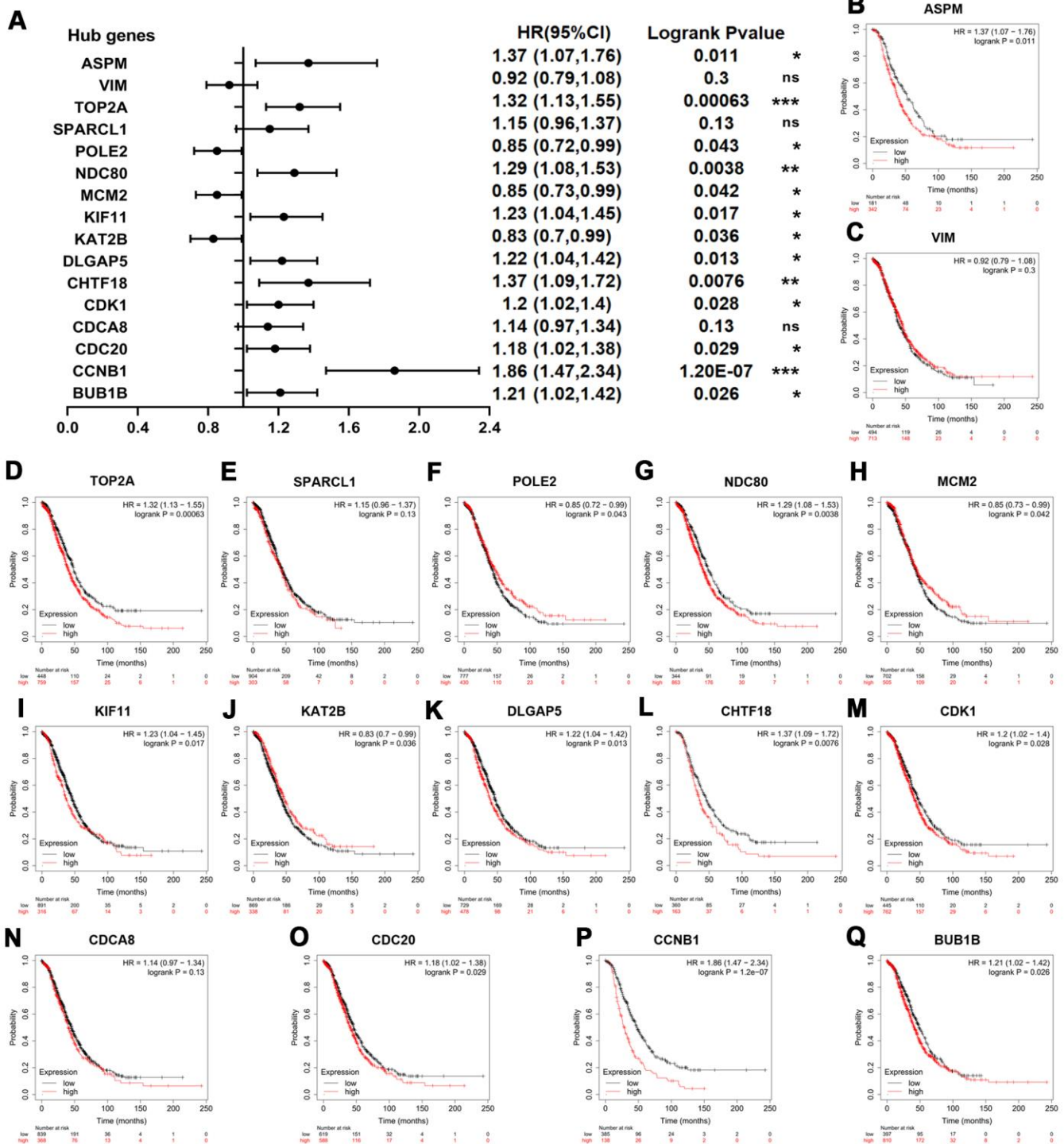


**Supplementary Figure 1. Normalization of the raw data in GEO database.** (A) Normalization of GSE36668 dataset. (B) Normalization of GSE54388 dataset. (C) Normalization of GSE69428 dataset. Blue represents raw data and red represents data after normalization.

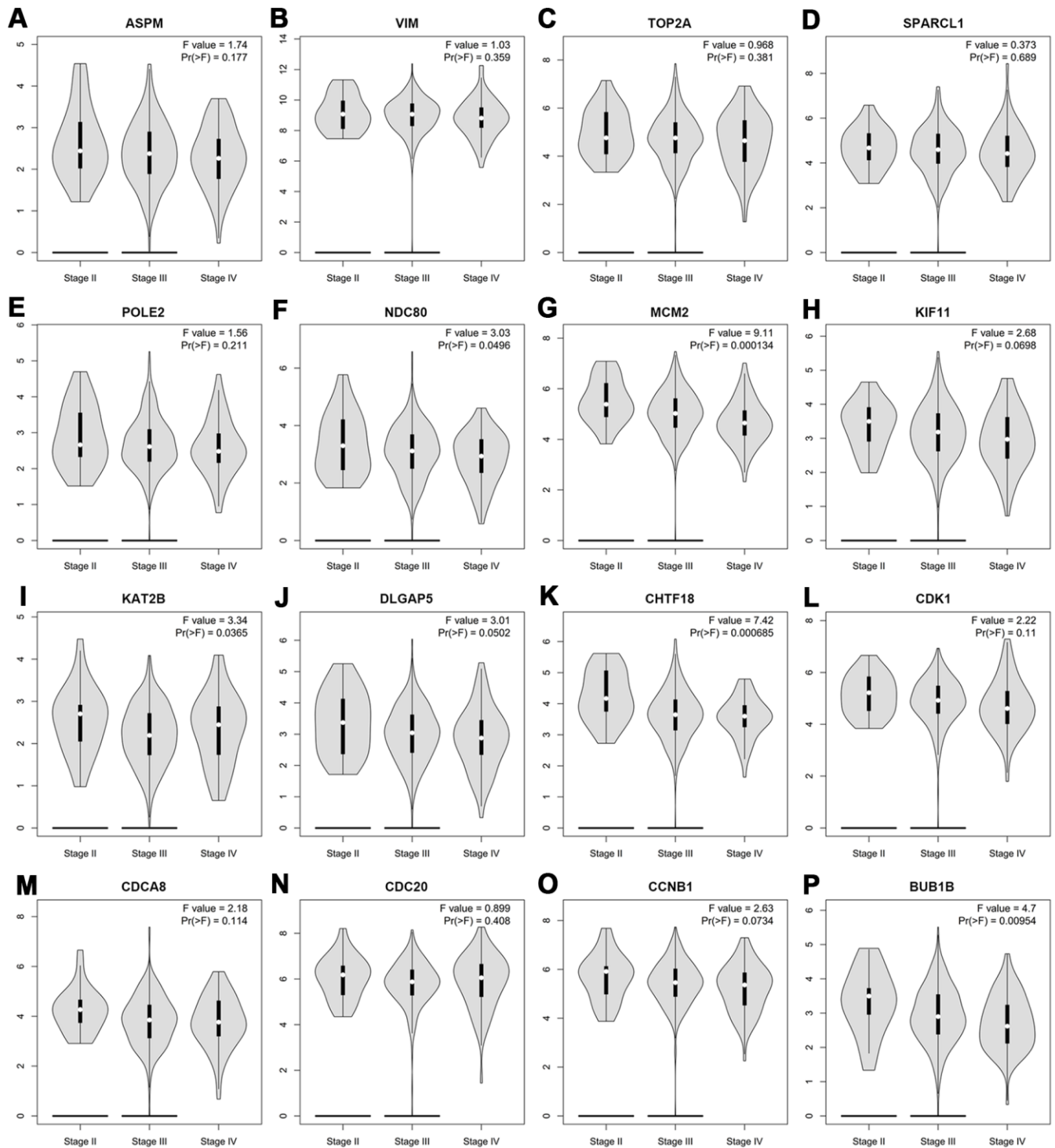


**Supplementary Figure 2.** The heatmap of gene expression in (A) GSE36668, (B) GSE54388 and (C) GSE69428. Red indicates a relatively high expression and blue indicates a relatively low expression.

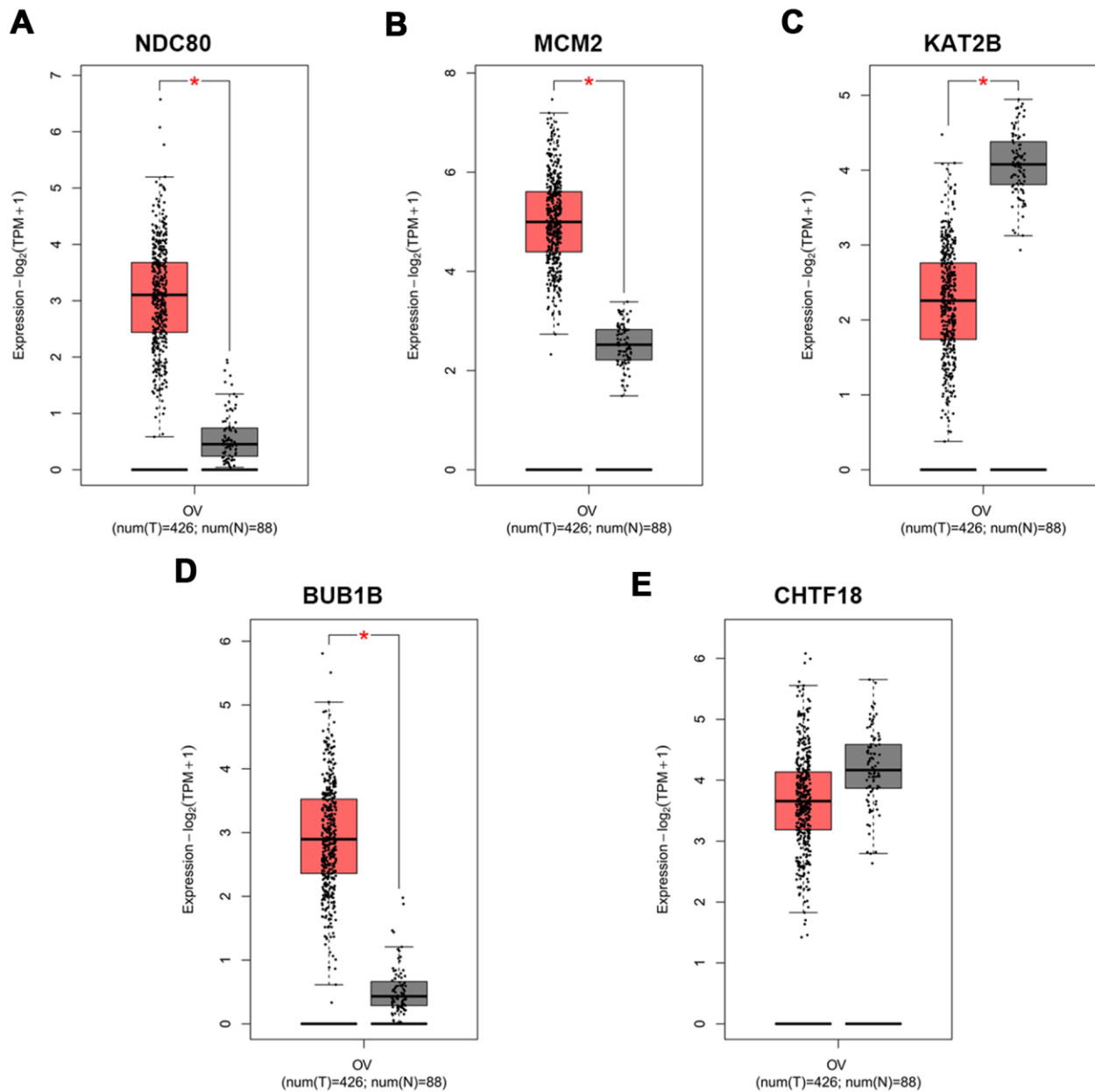




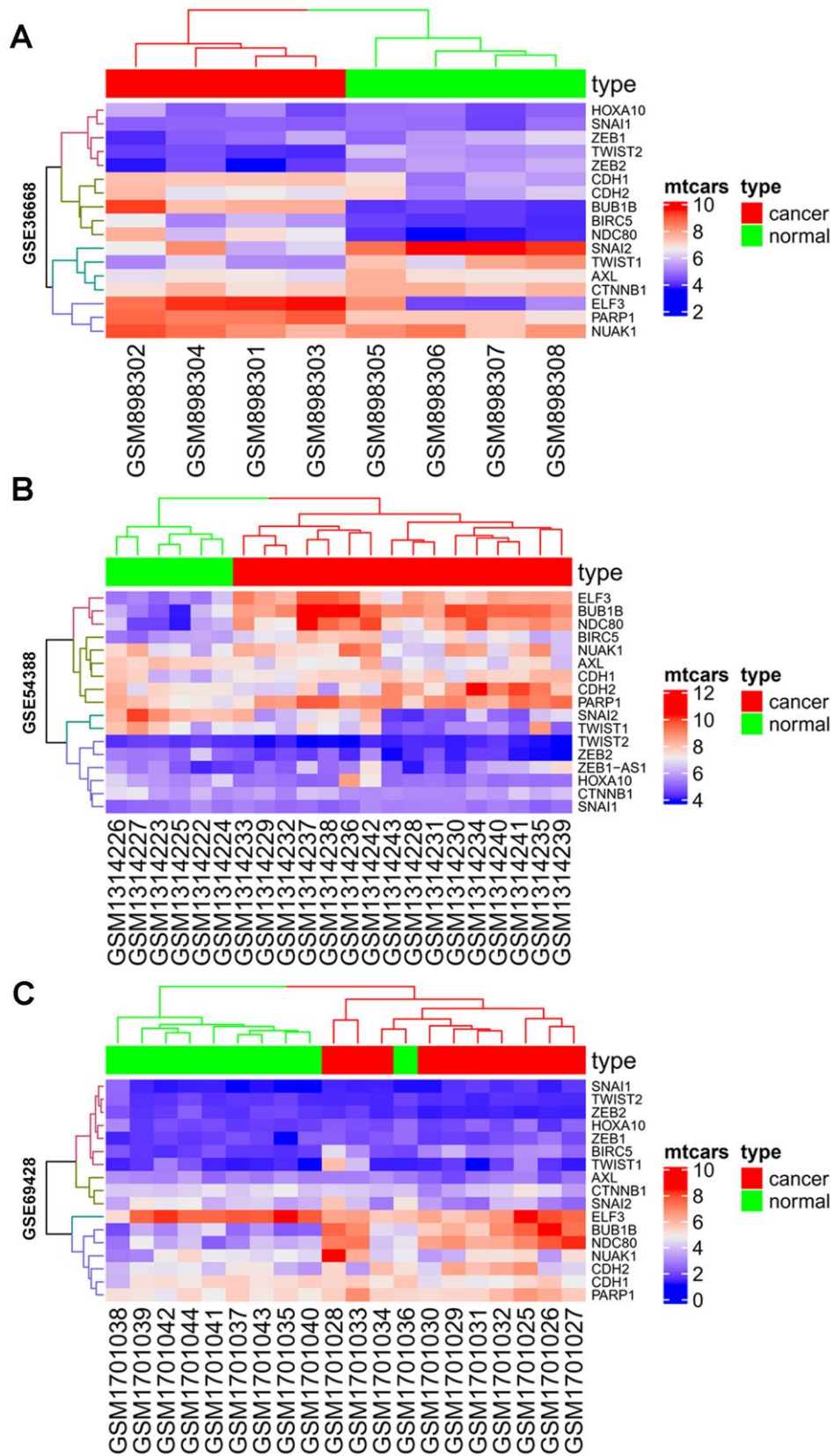
**Supplementary Figure 3. Overall survival impact of hub genes on patients with OSC at all stages.** (A) Survival prognosis forest map related to hub genes at all stages in OSC patients. (B–Q) Kaplan-Meier analysis for overall survival involved in the key genes expressed in OSC patients, covering (B) ASPM, (C) VIM, (D) TOP2A, (E) SPARCL1, (F) POLE2, (G) NDC80, (H) MCM2, (I) KIF11, (J) KAT2B, (K) DLGAP5, (L) CHTF18, (M) CDK1, (N) CDCA8, (O) CDC20, (P) CCNB1 and (Q) BUB1B. Logrank  $p$ -value < 0.05 was considered statistically significant.



**Supplementary Figure 4. The expression levels of hub genes in ovary cancer samples at different pathological stages.** The hub genes (A) ASPM, (B) VIM, (C) TOP2A, (D) SPARCL1, (E) POLE2, (F) NDC80, (G) MCM2, (H) KIF11, (I) KAT2B, (J) DLGAP5, (K) CHTF18, (L) CDK1, (M) CDCA8, (N) CDC20, (O) CCNB1 and (P) BUB1B were analyzed respectively from the GEPIA2. ANOVA was performed to assess the statistical significance of variations.  $Pr(>F) < 0.05$  was considered statistically significant.

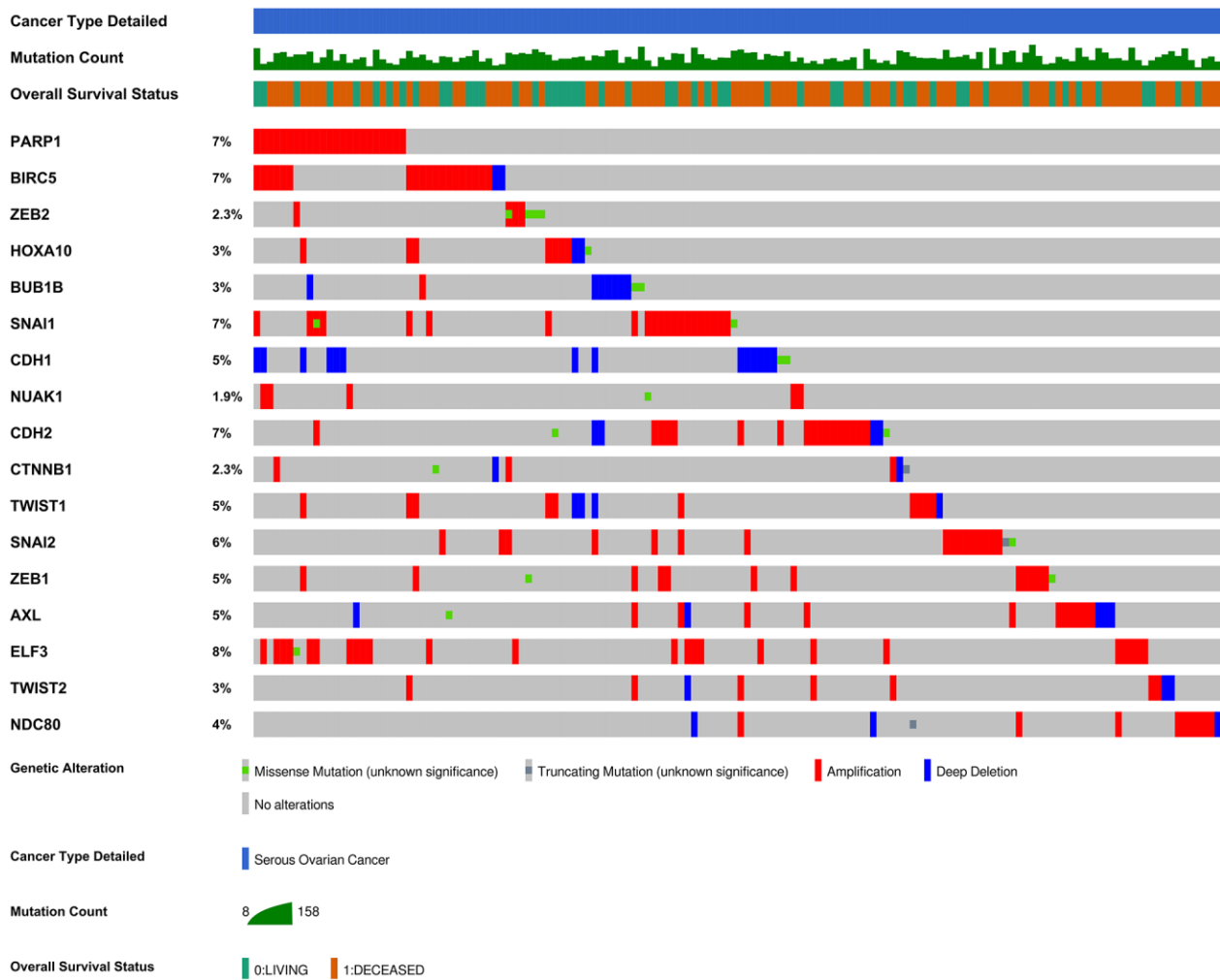


**Supplementary Figure 5. The expression levels of key genes in ovary cancer samples and normal ovary samples.** Based on TCGA and GTEx data from GEPIA2, we validated the expression levels of the key genes in ovarian cancer (n = 426, red) and normal tissues (n = 88, gray). (A) NDC80, (B) MCM2, (C) KAT2B, (D) BUB1B and (E) CHTF18.

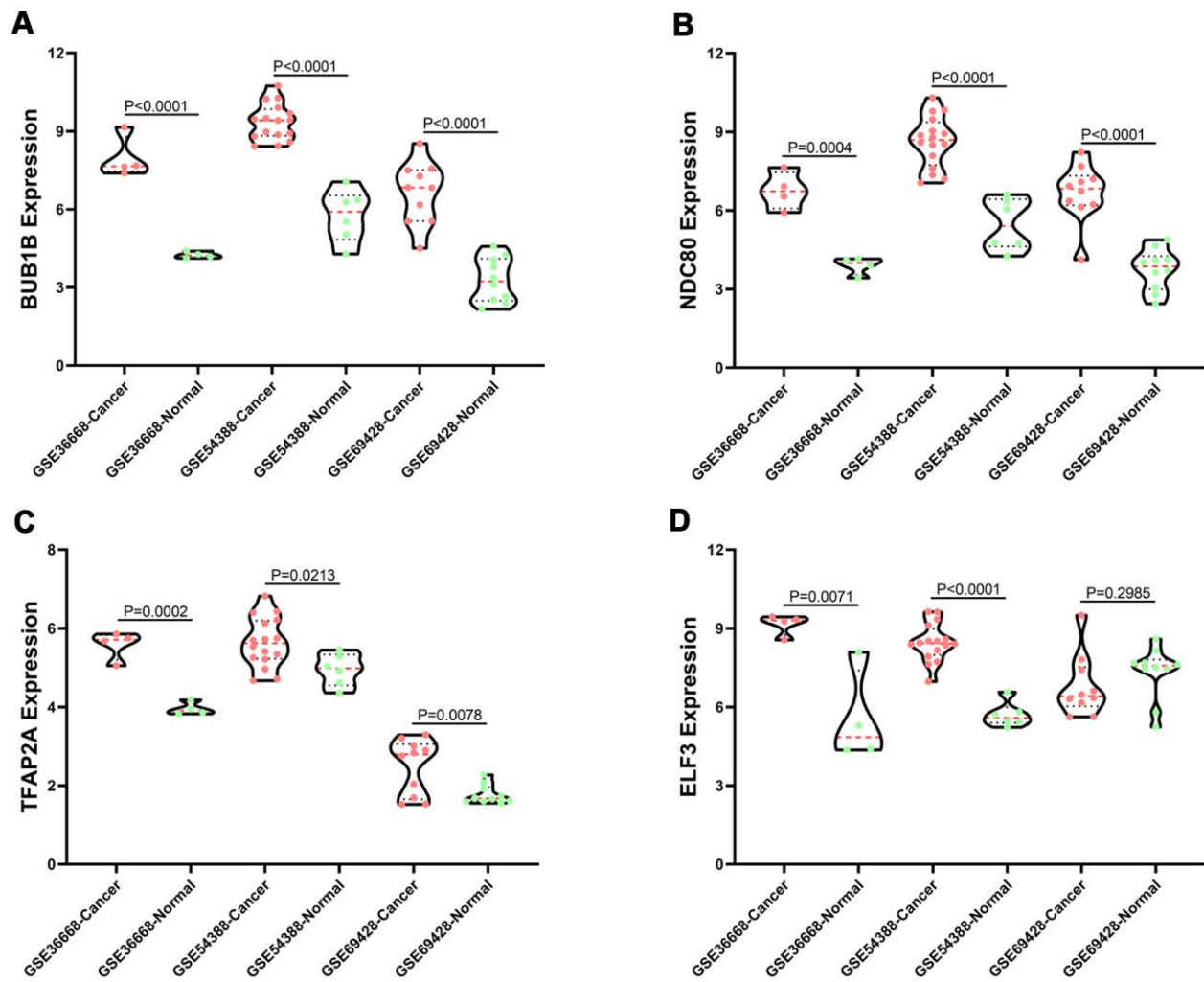


**Supplementary Figure 6.** The heatmap of differentially expressed core genes and EMT regulators between OSC tissues (red) and normal ovary tissues (green) in the datasets of (A) GSE36668, (B) GSE54388, (C) GSE69428. Blue indicates a relatively low expression and red indicates a relatively high expression.



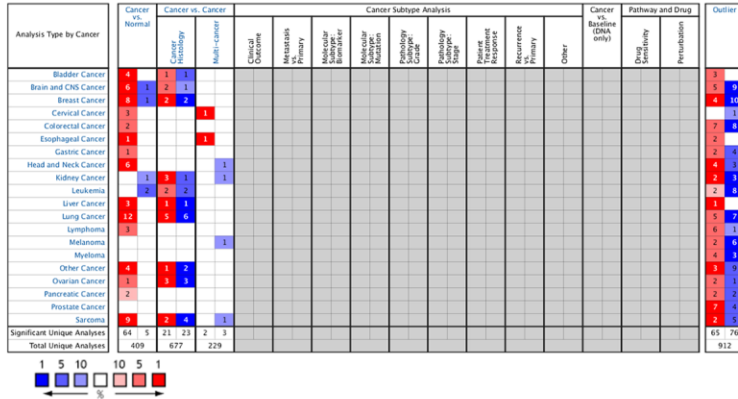


Supplementary Figure 7. The genetic variation, mutation count and overall survival status related to the core genes and EMT regulators are shown as a visual summary for a set of serous ovarian cancer samples from the TCGA database.



**Supplementary Figure 8. Validation of core genes with relevant TF and EMT regulator in GSE36668, GSE54388 and GSE69428 datasets. The expression level analyses of (A) BUB1B, (B) NDC80, (C) TFAP2A, and (D) ELF3.**

### A Disease Summary for BUB1B



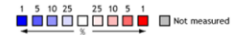
Cell color is determined by the best gene rank percentile for the analyses within the cell.

NOTE: An analysis may be counted in more than one cancer type.

Median Rank	p-Value	Gene
132.0	1.54E-6	BUB1B

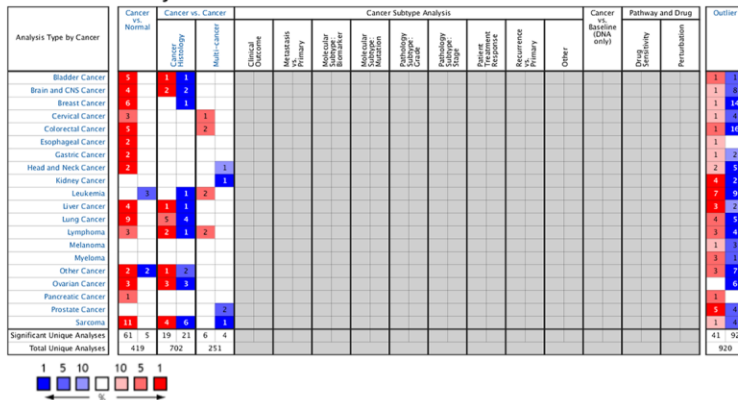
1. p=6.56E-16 Flod Change=4.893 Rank=101
2. p=3.07E-6 Flod Change=2.312 Rank=163
3. p=2.47E-9 Flod Change=2.782 Rank=22
4. p=2.56E-7 Flod Change=8.640 Rank=330

- Legend**
1. Ovarian Cancer Type: Ovarian Carcinoma *Anglesio Ovarian, Mol Cancer Res, 2008*
  2. Ovarian Serous Adenocarcinoma vs. Normal *Lu Ovarian, Clin Cancer Res, 2004*
  3. Ovarian Cancer Type: Ovarian Carcinoma *Sieben Ovarian, J Clin Oncol, 2005*
  4. Ovarian Serous Cystadenocarcinoma vs. Normal *TCGA Ovarian, No Associated Paper, 2013*



The rank for a gene is the median rank for that gene across each of the analyses. The p-Value for a gene is its p-Value for the median-ranked analysis.

### C Disease Summary for NDC80



Cell color is determined by the best gene rank percentile for the analyses within the cell.

NOTE: An analysis may be counted in more than one cancer type.

Median Rank	p-Value	Gene
192.5	4.04E-8	NDC80

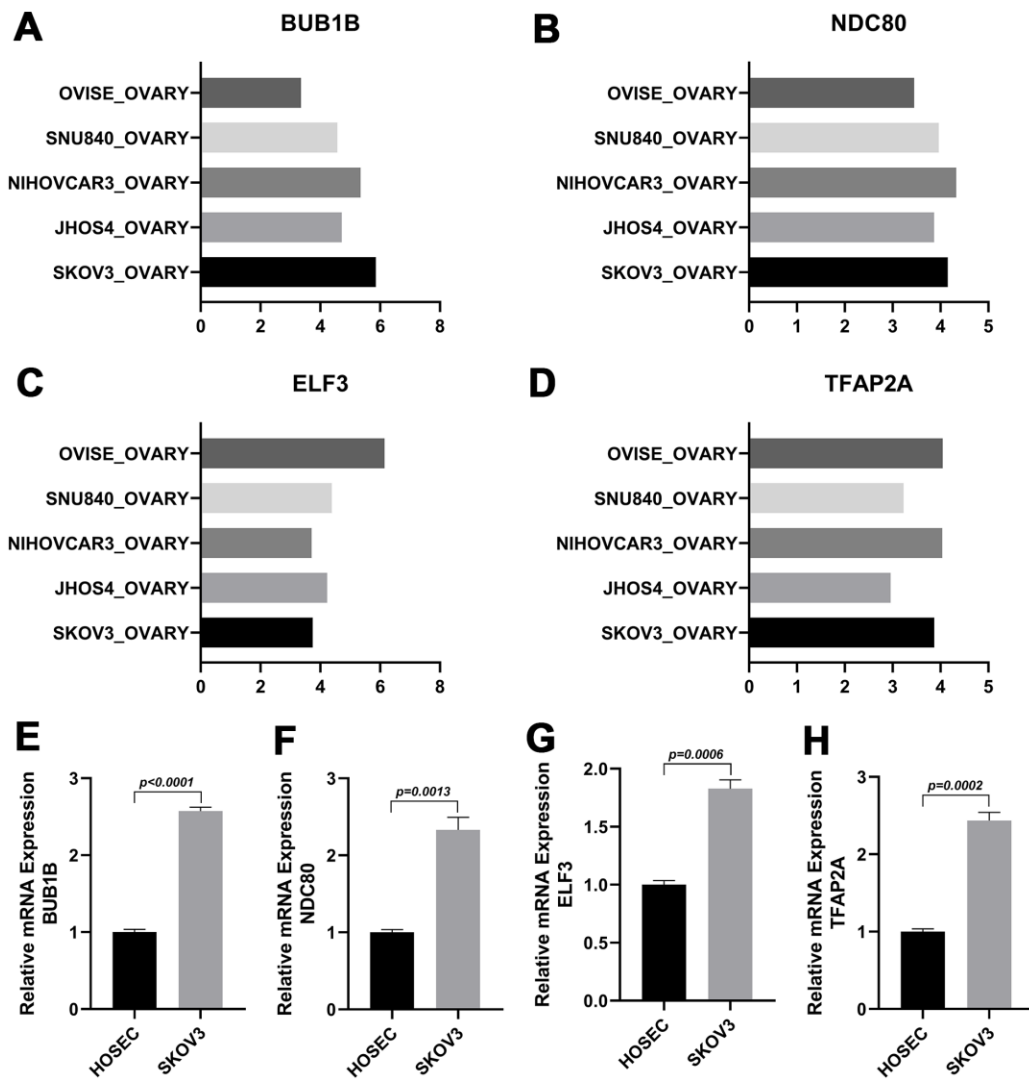
1. p=1.03E-14 Flod Change=4.501 Rank=149
2. p=3.13E-5 Flod Change=1.859 Rank=360
3. p=1.86E-7 Flod Change=2.291 Rank=61
4. p=8.08E-8 Flod Change=6.175 Rank=236

- Legend**
1. Ovarian Cancer Type: Ovarian Carcinoma *Anglesio Ovarian, Mol Cancer Res, 2008*
  2. Ovarian Serous Adenocarcinoma vs. Normal *Lu Ovarian, Clin Cancer Res, 2004*
  3. Ovarian Cancer Type: Ovarian Carcinoma *Sieben Ovarian, J Clin Oncol, 2005*
  4. Ovarian Serous Cystadenocarcinoma vs. Normal *TCGA Ovarian, No Associated Paper, 2013*



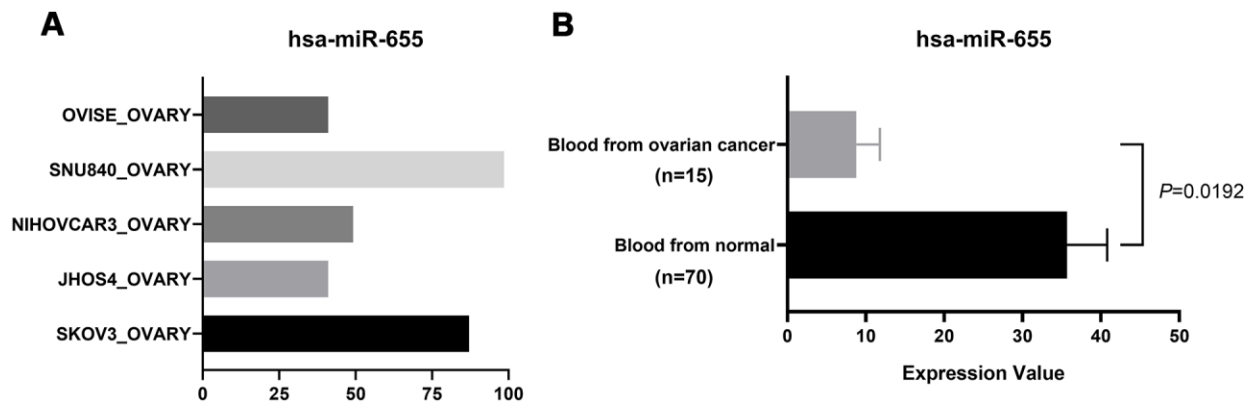
The rank for a gene is the median rank for that gene across each of the analyses. The p-Value for a gene is its p-Value for the median-ranked analysis.

**Supplementary Figure 9. OncoPrint analysis of NDC80 and BUB1B expression in cancer tissues vs normal tissues.** Validation of BUB1B (A) and NDC80 (C) on a transcriptional level in multiple cancer types and relevant studies. Heatmap represents BUB1B (B) and NDC80 (D) expression in ovarian cancer samples vs normal ovary samples. 1. ovarian carcinoma, Anglesio Ovarian, Mol Cancer Res, 2008; 2. ovarian carcinoma, Sieben Ovarian, J Clin Oncol, 2005; 3. Ovarian serous Adenocarcinoma vs Normal Lu Ovarian, Clin Cancer Res, 2004; 4. Ovarian serous Cystadenocarcinoma vs Normal TCGA Ovarian, No Associated Paper, 2013.



**Supplementary Figure 10. The expression levels of core genes relevant with TF and EMT regulator in ovarian cancer cell lines.** Analysis for BUB1B (A), NDC80 (B), ELF3 (C) and TFAP2A (D) from the CLE in five different ovarian cancer cell lines. qRT-PCR assay was adopted to evaluate the expression of BUB1B (E), NDC80 (F), ELF3 (G) and TFAP2A (H) in SKOV3 compared with HOSEC as normal control.





**Supplementary Figure 11. The expression level of hsa-miR-655.** Analysis of hsa-miR-655 (A) from the CCLE in five different ovarian cancer cell lines. (B) The hsa-miR-655 expression and (C) the differential expression in blood miRNAs from ovarian cancer patients modulating core genes and EMT regulators compared with normal explored in GSE31568.

## Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 2, 3, 8.

**Supplementary Table 1. Details of ovarian serous cancer associated microarray datasets from GEO database.**

Reference	sample	GEO	Platform	Normal ovary	Ovarian serous cancer
Elgaaen et al (2012)	Ovarian	GSE36668	GPL570	4	4
Yeung et al (2017)	Ovarian	GSE54388	GPL570	6	16
Yamamoto et al (2015)	Ovarian	GSE69428	GPL570	10	10

**Supplementary Table 2. The details of significant DEGs from GSE36668, GSE54388 and GSE69428 datasets.**

**Supplementary Table 3. GO Terms and KEGG functional enrichment analyses of co-DEGs.**

**Supplementary Table 4. Module analysis of co-DEGs using MCODE of Cytoscope.**

MCODE_Cluster	MCODE_Node_Status	MCODE_Score	name	TopologicalCoefficient
Cluster 1	Clustered	21.32307692	CDK1	0.24610988
Cluster 1	Clustered	21.32307692	CCNB1	0.28399209
Cluster 1	Clustered	21.32307692	CDC20	0.29765066
Cluster 1	Clustered	21.32307692	KIF11	0.32624113
Cluster 1	Clustered	21.32307692	TOP2A	0.33646245
Cluster 1	Clustered	21.32307692	CDCA8	0.34704034
Cluster 1	Clustered	21.32307692	BUB1B	0.36419753
<b>Cluster 1</b>	<b>Seed</b>	<b>21.32307692</b>	<b>ASPM</b>	<b>0.36976744</b>
Cluster 1	Clustered	21.32307692	DLGAP5	0.38181271
Cluster 1	Clustered	21.32307692	NDC80	0.35257453
Cluster 1	Clustered	21.32307692	KIF2C	0.40125
Cluster 1	Clustered	21.32307692	CENPF	0.40493421
Cluster 1	Clustered	21.32307692	TTK	0.40350877
Cluster 1	Clustered	21.32307692	KIF20A	0.40773229
Cluster 1	Clustered	18.8034188	RRM2	0.34723854
Cluster 1	Clustered	21.32307692	NUSAP1	0.41392318
Cluster 1	Clustered	20.51	NCAPG	0.40590278
Cluster 1	Clustered	20.37	TPX2	0.40785714
Cluster 1	Clustered	20.37	CEP55	0.41801471
Cluster 1	Clustered	20.69565217	UBE2C	0.42312009
Cluster 1	Clustered	19.98290598	PBK	0.44112903
Cluster 1	Clustered	18	ESPL1	0.39016897
Cluster 1	Clustered	20.44	KIF4A	0.45512821
Cluster 1	Clustered	19.60474308	KIF15	0.44137931
Cluster 1	Clustered	19.49275362	SPAG5	0.45767196
Cluster 1	Clustered	20.69565217	MELK	0.46383929
Cluster 1	Clustered	18.90952381	NUF2	0.45238095
Cluster 1	Clustered	19.76284585	KIF23	0.43717728
Cluster 1	Clustered	17.81052632	PRC1	0.41557223
Cluster 1	Clustered	18.81904762	PTTG1	0.46615385
Cluster 1	Clustered	18.90952381	MKI67	0.48839662
Cluster 1	Clustered	17.90526316	CDCA3	0.50185185

Cluster 2	Clustered	14	KNTC1	0.48070175
Cluster 2	Clustered	14	CENPU	0.47387387
Cluster 2	Clustered	14	ZWINT	0.45855856
Cluster 2	Clustered	14	RCC2	0.48841699
Cluster 2	Clustered	14	CDCA5	0.48841699
Cluster 3	Clustered	5.727272727	PCNA	0.23539618
Cluster 3	Clustered	5.785714286	MCM10	0.22094361
Cluster 3	Clustered	5.5	FEN1	0.31566265
Cluster 3	Clustered	5	DTL	0.34993614
Cluster 3	Clustered	5	DNA2	0.40613027
Cluster 3	Clustered	5	TIMELESS	0.41666667
Cluster 3	Clustered	5.785714286	ORC6	0.35030864
Cluster 3	Clustered	5	RMI2	0.44387755
<b>Cluster 3</b>	<b>Seed</b>	<b>6</b>	<b>CHTF18</b>	<b>0.53571429</b>
Cluster 3	Clustered	5.785714286	CDC7	0.30892857
Cluster 3	Clustered	5	RAD51AP1	0.52857143
Cluster 4	Clustered	7.418181818	MCM3	0.26919476
Cluster 4	Clustered	8	MCM7	0.26763285
<b>Cluster 4</b>	<b>Seed</b>	<b>8</b>	<b>MCM2</b>	<b>0.25791624</b>
Cluster 5	Clustered	3.771428571	RFC4	0.19430761
<b>Cluster 5</b>	<b>Seed</b>	<b>4.581818182</b>	<b>POLE2</b>	<b>0.2029321</b>
Cluster 5	Clustered	4	ISG15	0.50641026
Cluster 6	Clustered	2	LAMC1	0.66666667
<b>Cluster 6</b>	<b>Seed</b>	<b>2</b>	<b>SPARCL1</b>	<b>0.83333333</b>
Cluster 6	Clustered	2	STC2	0.83333333
<b>Cluster 7</b>	<b>Seed</b>	<b>2</b>	<b>VIM</b>	<b>1</b>
Cluster 7	Clustered	2	TNNI3	1
Cluster 7	Clustered	2	TNNT1	1
<b>Cluster 8</b>	<b>Seed</b>	<b>2</b>	<b>KAT2B</b>	<b>0.43333333</b>
Cluster 8	Clustered	2	E2F3	0.53333333
Cluster 8	Clustered	2	MEF2C	0.75
Cluster 9	Clustered	2	GAPDH	1
Cluster 9	Clustered	2	TPI1	1
<b>Cluster 9</b>	<b>seed</b>	<b>2</b>	<b>GPI</b>	<b>1</b>

**Supplementary Table 5. Hub genes evaluated by PPI network analysis Degree>40 or the seed genes in MOCODE analysis of Cytoscape.**

Entrez ID	Gene symbol	Description	Alternate names	Chromosome	Map location
63922	CHTF18	chromosome transmission fidelity factor 18	C16orf41 C321D2.2 C321D2.3 C321D2.4 CHL12 Ctf18 RUVBL	16	16p13.3
4171	MCM2	minichromosome maintenance complex component 2	BM28 CCNL1 CDCL1 D3S3194 MITOTIN cdc19	3	3q21
5427	POLE2	polymerase (DNA directed), epsilon 2, accessory subunit	DPE2	14	14q21-q22
8404	SPARCL1	SPARC-like 1 (hevin)	MAST 9 MAST9 PIG33 SC1	4	4q22.1
7431	VIM	vimentin	CTRCT30 HEL113	10	10p13
8850	KAT2B	K(lysine) acetyltransferase 2B	CAF P CAF PCAF	3	3p24
983	CDK1	cyclin-dependent kinase 1	CDC2 CDC28A P34CDC2	10	10q21.1
891	CCNB1	cyclin B1	CCNB	5	5q12
991	CDC20	cell division cycle 20	CDC20A bA276H19.3 p55CDC	1	1p34.1
3832	KIF11	kinesin family member 11	EG5 HKSP KNSL1 MCLMR TRIP5	10	10q24.1
7153	TOP2A	topoisomerase (DNA) II alpha 170kDa	TOP2 TP2A	17	17q21-q22
55143	CDCA8	cell division cycle associated 8	BOR BOREALIN DasraB MESRGP	1	1p34.3
701	BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B	BUB1beta BUBR1 Bub1A MAD3L MVA1 SSK1 hBUBR1	15	15q15
259266	ASPM	asp (abnormal spindle) homolog, microcephaly associated (Drosophila)	ASP Calmbp1 MCPH5	1	1q31
9787	DLGAP5	discs, large (Drosophila) homolog-associated protein 5	DLG7 HURP	14	14q22.3
10403	NDC80	NDC80 kinetochore complex component	HEC HEC1 HsHec1 KNTC2 TID3 hsNDC80	18	18p11.32

**Supplementary Table 6. KEGG enrichment of core genes and EMT regulators.**

Pathway ID	Description	Count in gene set	False Discovery Rate
hsa04520	Adherens junction	4 of 71	1.77E-05
hsa04390	Hippo signaling pathway	3 of 152	0.0059
hsa05216	Thyroid cancer	2 of 37	0.0067
hsa05205	Proteoglycans in cancer	3 of 195	0.0067
hsa05213	Endometrial cancer	2 of 58	0.008
hsa05130	Pathogenic Escherichia coli infection	2 of 53	0.008
hsa05412	Arrhythmogenic right ventricular cardiomyopathy	2 of 72	0.0103
hsa05100	bacterial invasion of Epithelial cells	2 of 72	0.0103
hsa05215	Prostate cancer	2 of 97	0.0142
hsa04514	cell adhesion molecules(CAMs)	2 of 139	0.0254
hsa05226	Gastric cancer	2 of 147	0.0257
hsa05206	MicroRNAs in cancer	2 of 149	0.0257
hsa05202	Transcriptional misregulation in cancer	2 of 169	0.0284
hsa04015	Rap1 signaling pathway	2 of 203	0.0373



**Supplementary Table 7. Primers used for the real-time PCR.**

<b>Gene</b>	<b>Primers</b>	<b>Sequences (5'-3')</b>
$\beta$ -actin	Forward	CCACGAAACTACCTTCAACTCC
	Reverse	GTGATCTCCTTCTGCATCCTGT
ELF3	Forward	GCAAACACCCTGGACTTGAC
	Reverse	CTGTCTCTCATCACGGTGC
TFAP2A	Forward	ATATCCGTTACGCCGATCC
	Reverse	CCTCGCAGTCCTCGTACTTG
BUB1B	Forward	GGGATGGGTCTTCTGGAAA
	Reverse	GGCATTTCAGAATCCGCACAA
NDC80	Forward	CCTGGTGTTTTGATGACCGC
	Reverse	TCCAGTTTCCTGACACGACC

**Supplementary Table 8. The details of differential expression for miRNAs from GSE31568.**

## **Supplementary Videos**

Please browse Full Text version to see the data of Supplementary Videos 1, 2, 3.

**Supplementary Video 1. Multidimensional dynamic scaling for GSE36668 dataset. Related to Figure 1A.**

**Supplementary Video 2. Multidimensional dynamic scaling for GSE54388 dataset. Related to Figure 1B.**

**Supplementary Video 3. Multidimensional dynamic scaling for GSE69428 dataset. Related to Figure 1C.**