



Research article

Meta-analysis identifying epithelial-derived transcriptomes predicts poor clinical outcome and immune infiltrations in ovarian cancer

Dong-feng Li¹, Aisikeer Tulahong², Md. Nazim Uddin³, Huan Zhao¹ and Hua Zhang^{2,*}

¹ Department of Pharmacy, First Affiliated Hospital of Xinjiang Medical University, Urumqi 830011, China

² Department of Oncology, First Affiliated Hospital of Xinjiang Medical University, Urumqi 830011, China

³ School of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing 211198, China

* **Correspondence:** Email: zh4401@sina.com.

Abstract: *Background:* Previous studies revealed that the epithelial component is associated with the modulation of the ovarian tumor microenvironment (TME). However, the identification of key transcriptional signatures of laser capture microdissected human ovarian cancer epithelia remains lacking. *Methods:* We identified the differentially expressed transcriptional signatures of human ovarian cancer epithelia by meta-analysis of GSE14407, GSE2765, GSE38666, GSE40595, and GSE54388. Then we investigated the enrichment of KEGG pathways that are associated with epithelia-derived transcriptomes. Finally, we investigated the correlation of key epithelia-hub genes with the survival prognosis and immune infiltrations. Finally, we investigated the genetic alterations of key prognostic hub genes and their diagnostic efficacy in ovarian cancer epithelia. *Results:* We identified 1339 differentially expressed genes (DEGs) in ovarian cancer epithelia including 541 upregulated and 798 downregulated genes. We identified 21 (such as E2F4, FOXM1, TFDP1, E2F1, and SIN3A) and 11 (such as JUN, DDX4, FOSL1, NOC2L, and HMGA1) master transcriptional regulators (MTRs) that are interacted with upregulated and the downregulated genes in ovarian tumor epithelium, respectively. The STRING-based analysis identified hub genes (such as *CDK1*, *CCNB1*, *AURKA*, *CDC20*, and *CCNA2*) in ovarian cancer epithelia. The significant clusters of identified hub genes are associated with the enrichment of KEGG pathways including cell cycle, DNA replication, cytokine-cytokine receptor interaction, pathways in cancer, and focal adhesion. The upregulation of *SCNN1A* and *CDCA3* and the downregulation of *SOX6* are correlated with a shorter survival prognosis in ovarian cancer (OV). The expression level of *SOX6* is negatively correlated with immune score and positively correlated with tumor purity in OV. Moreover, *SOX6* is negatively correlated with the infiltration of TILs, CD8⁺ T cells, CD4⁺ Regulatory T cells, cytolytic activity, T cell activation, pDC, neutrophils, and macrophages in OV. Also, *SOX6* is negatively correlated with

various immune markers including *CD8A*, *PRF1*, *GZMA*, *GZMB*, *NKG7*, *CCL3*, and *CCL4*, indicating the immune regulatory efficiency of *SOX6* in the TME of OV. Furthermore, *SCNN1A*, *CDCA3*, and *SOX6* genes are genetically altered in OV and the expression levels of *SCNN1A* and *SOX6* genes showed diagnostic efficacy in ovarian cancer epithelia. *Conclusions:* The identified ovarian cancer epithelial-derived key transcriptional signatures are significantly correlated with survival prognosis and immune infiltrations, and may provide new insight into the diagnosis and treatment of epithelial ovarian cancer.

Keywords: ovarian cancer epithelia; meta-analysis; hub genes; survival prognosis; diagnostic efficacy

1. Introduction

Ovarian cancer is the most lethal gynecological malignancy worldwide [1]. The most common type of ovarian cancer is epithelial ovarian cancer and is the substantial cause of death in female cancer patients every year [2]. At the diagnosis stage, 58% of ovarian cancers are metastatic with a 5-year relative survival of only 30% of patients [2]. Ovarian surface epithelial carcinoma cells are capable of initiating ovarian cancer cells and are crucial for the malignant progression of cancer [3]. In the tumor microenvironment, epithelial component from high grade serous ovarian cancer is associated with cancer invasion, progression, and pathogenesis [4]. The differentially expressed genes of the epithelial component in ovarian cancer are associated with cell proliferation, invasion, motility, chromosomal instability, and gene silencing [5]. The epithelial component of ovarian cancer tissue also provided the prognostic factors, immunological factors, and molecular factors [6]. In ovarian cancer, key immune cells and soluble molecules in the TME regulated the prognosis [7]. These studies provide the clue that the human ovarian cancer epithelial-derived transcriptomes are associated with cancer initiation, invasion, migration, and metastasis.

Previous studies revealed the association of coding and non-coding biomarkers including differentially expressed genes (DEGs), mRNA, and miRNA on ovarian cancer pathogenesis. Ting Gui et al. identified biomarkers including key hub genes and signaling pathways that are involved with epithelial ovarian cancer progression [8]. Another study identified the biomarkers including DEGs, key hub genes, prognostic genes, significant clusters of genes, and functional enrichment analysis in epithelial ovarian cancer [9]. Daniela Matei et al. identified mRNA levels that are significantly deregulated in primary cultures of ovarian epithelial cells derived from epithelial ovarian carcinoma [10]. Immunogenic mRNA and protein expression by cell cultures of epithelial ovarian cancer are also crucial factors to identify the disease progression [11]. MicroRNA (miRNA), the substantial onco-regulators in cancer, is associated with clinicopathological characteristics of epithelial ovarian cancer [12]. MiRNA regulating the initiation, proliferation, cell cycle, survivability, and resistance of chemosensitivity in ovarian cancer [13]. Some studies demonstrated the significance of bioinformatics analysis, data mining, gene regulatory networks, and prediction of biomarkers in human diseases [14–17]. Altogether, these studies provide the clue that the bioinformatics strategies could identify the epithelial-derived transcriptomes that are associated with the pathogenesis of ovarian cancer.

Herein, we identified the deregulated transcriptional signatures in laser capture microdissected human ovarian cancer epithelia. Then we identified the TFs that are associated with regulating the

deregulated transcriptional signatures in ovarian cancer epithelia. Moreover, we identified the key hub genes and hub genes associated clusters from the PPI network. Furthermore, we investigated the correlation of key genes with survival prognosis and immune infiltrations. Finally, we investigated the genetic alterations of key prognostic hub genes and their diagnostic efficacy in ovarian cancer. The study may provide novel insights to investigate the functional regulatory mechanisms of immune-related biomarkers and help to develop immune-related targeted therapy in the treatment of ovarian cancer. Also, this study will help the experimental biologists to further carry out their research to explore the clinical association of these identified biomarkers to treat ovarian cancer patients.

2. Materials

2.1. Datasets

We searched the NCBI gene expression omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) using the keywords “epithelial”, “ovarian epithelial”, “ovarian surface epithelium tumor”, “cancer-associated epithelial”, “ovarian cancer”, and “tumor epithelial”, and identified five ovarian epithelial gene expression datasets from the same platform. These included five datasets have the following criteria: the datasets extracted from the same platform (Affymetrix human genome U133 plus 2.0 array), the study organisms had to be *Homo sapiens*, the datasets had to contain ovarian epithelial cancer and normal ovarian epithelial tissue samples, and the sample size is greater than 20 with minimum 5 control samples. The five selected datasets are GSE14407 [3], GSE27651 [18], GSE38666 [19,20], GSE40595 [4], and GSE54388 [21]. The total samples included 120 ovarian epithelial tumors and 43 control epithelial. Moreover, we downloaded the TCGA ovarian cancer cohort (n = 307) (<https://portal.gdc.cancer.gov/>) and normalized it by base-2 log transformation. Finally, we downloaded the clinical data of the TCGA ovarian cancer cohort for analyzing the survival differences between the two groups (<https://portal.gdc.cancer.gov/>).

The GEO datasets (GSE14407, GSE27651, GSE38666, GSE40595, and GSE54388) used in this study are available in The National Biotechnology Information Center Gene (NCBI-Gene) database (<https://www.ncbi.nlm.nih.gov/gene/>). TCGA-OV cohort is downloaded from the Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>). Genetical data of OV was utilized from the cBioPortal (<http://www.cbioportal.org/>).

2.2. Identification of differentially expressed genes (DEGs)

We used Network Analyst [22] to identify the DEGs between ovarian epithelial tumor and normal samples by a meta-analysis of five epithelial gene expression profiling datasets (GSE14407 [3], GSE27651 [18], GSE38666 [19,20], GSE40595 [4], and GSE54388 [21]). Datasets were normalized by quantile normalization or base-2 log transformation. We removed the batch effects of multiple datasets by using the ComBat method [23]. A total of 20,184 common genes were found by integrating the datasets from five datasets using the Network Analyst [22] tool. We used the R package “limma” for identifying the DEGs between tumor and normal samples, and Cochran’s combination test for performing the meta-analysis [24]. The false discovery rate (FDR) [25] was used to adjust for multiple tests. We identified the DEGs using a threshold of absolute combined effect size (ES) > 1.50 and FDR < 0.05.

2.3. Identification of master transcriptional regulators (MTRs) that are significantly associated with the DEGs

We utilized Cytoscape [26] plug-in iRegulon [27] to identify the MTRs for the upregulated and downregulated DEGs. In the Cytoscape plug-in iRegulon [27], a minimum normalized enrichment score (NES) > 3.0 was selected for each TFs. For this purpose, we used an extensive collection of TF motifs and a large collection of ChIP-seq tracks. The iRegulon method depends on a ranking-and-recovery system where all genes of the human genome are scored by a motif discovery step integrating the clustering of binding sites within cis-regulatory modules (CRMs) and the potential distal location of CRMs upstream or downstream of the transcription start site (TSS ± 10 kb). The recovery step calculates the normalized enrichment score (NES) of TFs for each set of genes, input for each of the individual analyses, leading to the prediction of the TFs based on NES and their putative direct target genes which exist in the input lists. This methodology optimizes the linking of TFs to motifs using both explicit annotations and predictions of TF orthologs and motif similarity. A transcription factor NES was computed for each group where an NES > 3.0 corresponds, and the maximum false discovery rate (FDR) on motif similarity was set as 0.001 [28].

2.4. Gene-set enrichment analysis

We performed gene-set enrichment analysis of the DEGs by using the GSEA [29]. We inputted all significant upregulated and downregulated DEGs into the GSEA tool [29] for identifying deregulated pathways. The KEGG [30] pathways significantly associated with the upregulated DEGs and the downregulated DEGs were identified, respectively. The P-value < 0.05 was considered significant when selecting the pathways.

2.5. Identifying hub genes and modular analysis from protein-protein interaction (PPI) network of DEGs

To better know the relationship among these identified DEGs, the PPI network was established using the STRING-based analysis [31]. To identify the rank of hub genes, we used Cytoscape plug-in tool cytoHubba [32]. Hub genes were identified based on the degree of interactions with neighbor genes. We selected the minimum required interaction score is 0.40 for identifying the PPI of DEGs. Hub genes were defined as a gene that was connected to a minimum of 10 other DEGs in the PPI. We visualize the PPI networks by utilizing the Cytoscape 3.6.1 software [26]. A Cytoscape plug-in molecular complex detection (MCODE) was employed to detect the modules from the PPI network [33]. We identified the significant modules based on the MCODE score and node number. The threshold of the MCODE was Node Score Cut-off: 0.2, Haircut: true, K-Core: 2, and maximum depth from Seed: 100.

2.6. Survival analysis of key hub genes

We used the R package “survival” to investigate the survival prognosis of ovarian cancer patients [34]. We used the clinical data of the TCGA OV cohort for analyzing the survival differences. We compared the overall survival (OS) of ovarian cancer patients that are classified based on gene expression levels (expression levels $>$ median versus expression levels $<$ median). Kaplan-Meier

survival curves were used to show the survival differences, and the log-rank test ($P < 0.05$) was utilized to evaluate the significance of survival differences between the groups.

2.7. ESTIMATE algorithmic for quantifying immune score and stromal score

ESTIMATE is an algorithmic tool based on the R package for predicting tumor purity, immune score (predicting the infiltrations of immune cells), and stromal score (predicting the infiltrations of stromal cells) which uses the gene expression profiles of 141 immune genes and 141 stromal genes [35]. The presence of infiltrated immune cells and stromal cells in tumor tissues were calculated using related gene expression matrix data, represented by immune score and stromal score, respectively [35]. Then we calculated the correlations of key genes with immune scores and stromal scores. The threshold value of correlation is $R > 0.20$, and P-value is not less than 0.001 (Spearman's correlation test).

2.8. Single-sample gene-set enrichment analysis (ssGSEA) and correlation of immune signatures with prognostic genes

One of the extension packages of GSEA, single-sample gene-set enrichment analysis (ssGSEA) was used to identify the enrichment scores of immune cells for each pairing of a sample and gene set in the tumor samples [36]. We collected the marker gene set for immune signatures and utilizing each gene set to quantify the ssGSEA scores of specific immune signatures [37–40]. We identified the enrichment levels (ssGSEA scores) of ten immune signatures included TILs, B cells, CD8+ T cells, CD4+ regulatory T cells, cytolytic activity, T cells activations, CAFs, pDC, macrophages, and neutrophils. All of the marker genes are displayed in Supplementary Table S1. Then we investigated Spearman's correlation between the ssGSEA scores and specific prognostic genes. The threshold of correlation of immune cells is the absolute value is not less than 0.20 with P-value < 0.05 .

2.9. Genetic alterations analysis of prognostic hub genes

We identified the genetic alterations associated with prognostic hub genes by using the cBioPortal (<http://www.cbioportal.org/>), an open-access tool for exploring and analyzing genetical alterations of multidimensional cancer studies [41]. In this study, we selected the TCGA OV epithelial data that contains 311 samples with mutation data and copy number alteration data (https://www.cbioportal.org/study/summary?id=ov_tcga).

2.10. Diagnostic efficacy evaluation for prognostic key genes

To assess diagnostic values of the prognostic genes, the receiver operating characteristic (ROC) curve was plotted and the area under the ROC curve (AUC) was calculated using the “pROC” R package [42] to evaluate the capability of distinguishing ovarian cancer epithelia and normal epithelia. The greater AUC value of individual genes indicated the differences between tumor and normal samples, and the key gene of AUC > 0.5 in the integrated five datasets was defined as a diagnostic efficiency of the gene [43]. If the P-value < 0.05 , the selection of the prognostic genes are considered statistically significant.

2.11. Statistical and computational analysis

We evaluated Pearson's or Spearman's correlation test to verify the significant levels between the two variables. For analyzing the correlations between the expression levels of hub genes and the enrichment levels (ssGSEA scores) of immune signatures, we used Spearman's correlation test because these data were not normally distributed [44]. For analyzing the correlations between the expression levels of hub genes with the expression levels of other marker genes, we utilized Pearson's correlation test because these data were normally distributed [44]. We utilized the Network Analyst [22] tool for calculating the average expression of a gene having multiple probes in the same expression dataset. We used the R package "ggplot2" for plotting the graphs in this study [45].

3. Results

3.1. Identification of differentially expressed genes in ovarian epithelium by a meta-analysis

Table 1. Gene expression pattern of top 25 upregulated genes in ovarian tumor epithelium relative to ovarian normal epithelium.

Entrez ID	Gene symbol	Combined ES	Adjusted P-value
64321	<i>SOX17</i>	3.6554	0
100133941	<i>CD24</i>	3.6041	7.24E-07
55765	<i>INAVA</i>	3.4924	0
2122	<i>MECOM</i>	3.4549	4.69E-12
4751	<i>NEK2</i>	3.404	2.02E-09
51514	<i>DTL</i>	3.3758	5.58E-07
10406	<i>WFDC2</i>	3.3672	0
51523	<i>CXXC5</i>	3.2352	0
1063	<i>CENPF</i>	3.1671	0.00011
4582	<i>MUC1</i>	3.1454	1.28E-13
701	<i>BUB1B</i>	3.1136	6.41E-10
4072	<i>EPCAM</i>	3.0962	3.57E-14
24137	<i>KIF4A</i>	3.0737	1.57E-13
7849	<i>PAX8</i>	3.0656	0
79581	<i>SLC52A2</i>	3.0212	0
1164	<i>CKS2</i>	2.9878	0
10112	<i>KIF20A</i>	2.9847	6.01E-07
57565	<i>KLHL14</i>	2.9794	1.89E-08
54845	<i>ESRP1</i>	2.9359	0
11130	<i>ZWINT</i>	2.9358	5.54E-05
81610	<i>FAM83D</i>	2.8954	8.96E-05
332	<i>BIRC5</i>	2.8935	0.000116
29968	<i>PSAT1</i>	2.8885	1.42E-13
1356	<i>CP</i>	2.8825	0

We identified 1339 differentially expressed genes (DEGs) between the ovarian epithelial tumor

and normal epithelial, which included 541 upregulated (Tables 1 and S2) and 798 downregulated (Tables 2 and S3) genes in the tumor ovarian epithelial when compared with normal ovarian epithelial based on combined Effect size (ES). ES is the difference between two group means divided by standard deviation, which is considered combinable and comparable across different studies [22]. Table 1 describes the regulatory status of the top 25 (The highest combined effect size) upregulated genes including *SOX17*, *CENPF*, *CD24*, *INAVA*, *MECOM*, *NEK2*, *DTL*, *WFDC2*, *PAX8*, *MUC1*, and *CXXC5*. Lin Zhao et al. reported that the expression of *SOX17*, *INAVA*, *WFDC2*, and *CD24* are upregulated in ovarian cancer [46]. It was found that *PAX8* and *SOX17* regulate tumor angiogenesis *in vitro* and *in vivo* in ovarian cancer [47]. *NEK2*, another upregulated gene in ovarian tumor epithelia, is associated with drug resistance in ovarian cancer [48].

Table 2. Gene expression pattern of top 25 downregulated genes in ovarian tumor epithelium relative to ovarian normal epithelium.

Entrez ID	Gene symbol	Combined ES	Adjusted P-value
55600	<i>ITLN1</i>	-8.0963	1.96E-06
150622	<i>SILC1</i>	-5.8072	2.46E-05
10351	<i>ABCA8</i>	-5.5625	2.58E-08
10216	<i>PRG4</i>	-5.3043	1.05E-06
316	<i>AOXI</i>	-5.264	1.87E-14
590	<i>BCHE</i>	-5.199	1.66E-05
8854	<i>ALDH1A2</i>	-5.039	1.76E-06
4886	<i>NPY1R</i>	-4.8974	8.46E-06
125	<i>ADH1B</i>	-4.872	3.77E-08
65055	<i>REEP1</i>	-4.8554	0
93663	<i>ARHGAP18</i>	-4.8506	0.00026109
4753	<i>NELL2</i>	-4.8473	6.74E-05
79804	<i>HAND2-ASI</i>	-4.7844	1.42E-06
4969	<i>OGN</i>	-4.7596	1.24E-07
84709	<i>MGARP</i>	-4.7055	2.86E-06
51555	<i>PEX5L</i>	-4.7042	0
126	<i>ADH1C</i>	-4.6513	1.57E-08
9737	<i>GPRASP1</i>	-4.6473	2.23E-05
8622	<i>PDE8B</i>	-4.6407	1.49E-08
56245	<i>C21orf62</i>	-4.6073	3.14E-06
339896	<i>GADL1</i>	-4.5668	9.75E-09
80310	<i>PDGFD</i>	-4.5255	5.88E-13
139221	<i>PWWP3B</i>	-4.3353	1.34E-12
3957	<i>LGALS2</i>	-4.2859	3.33E-05

In addition, Table 2 describes the regulatory status of the top 25 (The lowest combined effect size) downregulated genes including *ITLN1*, *SILC1*, *ABCA8*, *PRG4*, *AOXI*, *BCHE*, *ALDH1A2*, *NPY1R*, *ADH1B*, and *REEP1*. It was found that the downregulation of mesothelial cell-derived *ITLN1* in the omental tumor microenvironment facilitates ovarian cancer progression [49]. In the ovarian cancer cell line, *ABCA8* was significantly downregulated and is associated with drug-resistant [50]. A tumor

suppressor, ALDH1A2, was strongly downregulated 36-fold in 779 epithelial ovarian cancer cases compared with 18 normal controls [51]. Altogether, it indicates that the ovarian cancer epithelia-derived deregulated transcriptomes are associated with ovarian cancer pathogenesis.

3.2. Master transcriptional regulators (MTRs) are significantly associated with the DEGs

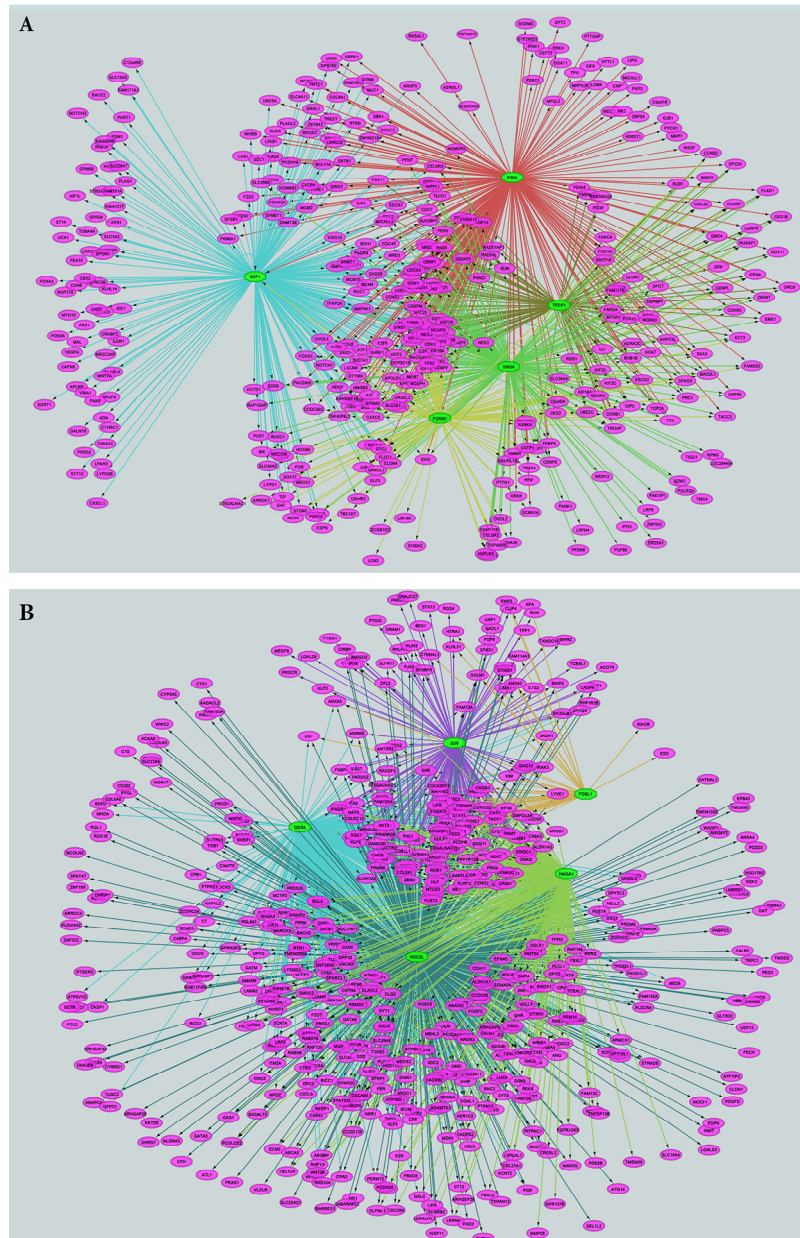


Figure 1. Regulatory networks of the master transcriptional regulators (MTRs) and their targeted differentially expressed genes (DEGs) between ovarian tumor epithelium and normal epithelium. A. Regulatory network of the top 5 MTRs and their targeted upregulated genes in ovarian tumor epithelium. B. Regulatory network of the top 5 MTRs and their targeted downregulated genes in ovarian tumor epithelium. The Green color octagon indicates MTRs, and the purple color oval indicates DEGs in ovarian tumor epithelial.

MTRs are crucial cancer-associated biomarkers and targets for metabolism-targeted cancer therapy [52]. To identify the significant MTRs that regulating the DEGs, we used the Cytoscape plug-in iRegulon tool. We identified 21 (E2F4, FOXM1, TFDP1, E2F1, SIN3A, PIR, SMAD1, E2F7, UBP1, NFYC, BDP1, E2F1, NFYA, MYBL2, TCF12, MTHFD1, TEAD4, MZF1, EP300, FHL2, and GRHL1) and 11 (JUN, DDX4, FOSL1, NOC2L, HMGA1, JUND, TCF12, EP300, NFIC, ZSCAN9, and FOS) MTRs for the upregulated and the downregulated genes in ovarian tumor epithelial, respectively (Table S4). We built the regulatory networks between the top 5 MTRs and upregulated DEGs (Figure 1A). In addition, we built the regulatory networks between the top 5 MTRs and downregulated DEGs (Figure 1B). In the networks, E2F4, top MTRs for upregulated genes, targeted 267 upregulated genes, and JUN, top-scored MTRs for downregulated genes, targeted 146 downregulated genes. Interestingly, three members of the E2 factor (E2F) family of transcription factors (E2F1, E2F7, and E2F1) were the MTRs that regulated the upregulated genes in ovarian tumors epithelium (Figure 1A). Deregulation of E2F transcription factors is associated with both proliferation-promoting and proliferation-inhibiting and their cross-talk is involved in the tumor biology of ovarian cancer and influences the clinical outcome of ovarian cancer [53,54]. Transcription factor JUN is associated with cellular proliferation, malignant transformation, and invasion in various tumors including ovarian cancer [55].

3.3. Epithelium-derived transcriptomes are associated with the enrichment of KEGG pathways

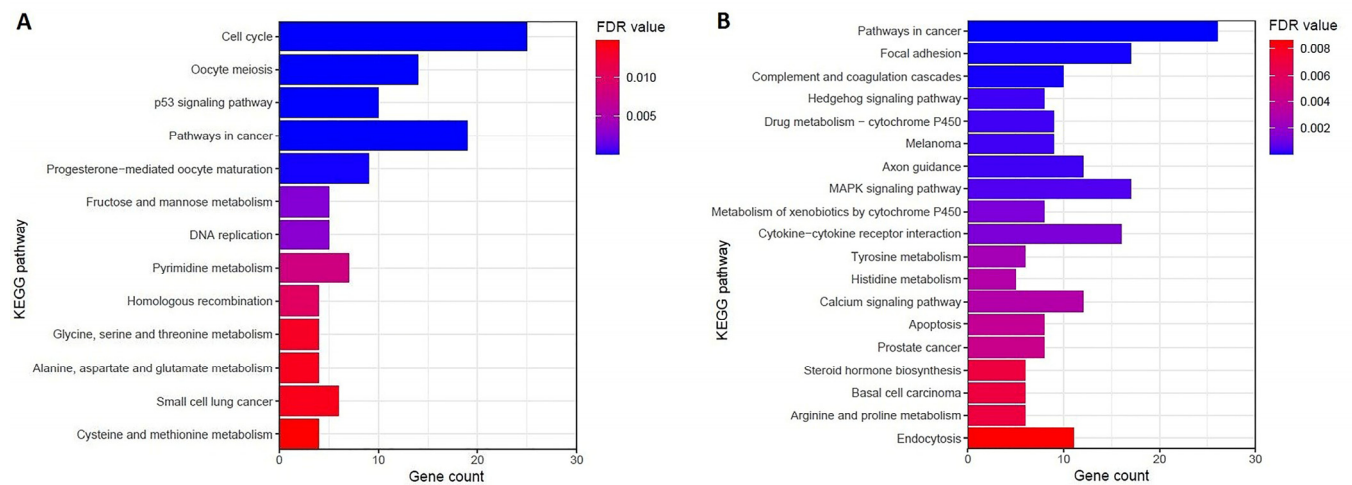


Figure 2. Significantly enriched KEGG pathways that are associated epithelium-derived DEGs. A. Significantly enriched 13 KEGG pathways that are associated with upregulated DEGs. B. Significantly enriched top 20 KEGG pathways that are associated with downregulated DEGs. FDR is false discovery rate.

The significantly enriched upregulated and downregulated biological pathways were identified by using the GSEA tool (Figure 2). GSEA identified 13 KEGG pathways that are significantly associated with upregulated DEGs (Figure 2A). The upregulated pathways included cell cycle, Oocyte meiosis, pathways in cancer, progesterone-mediated oocyte maturation, DNA replication, homologous recombination, and small cell lung cancer (Figure 2A). Besides, we identified the 42

KEGG pathways that are significantly linked with downregulated DEGs (Table S5). Some of these downregulated pathways are involved with immune regulation, including complement and coagulation cascades, cytokine-cytokine receptor interaction, Fc gamma R-mediated phagocytosis, apoptosis, and endocytosis (Table S5 and Figure 2B). Moreover, metabolic pathways including drug metabolism - cytochrome P450, metabolism of xenobiotics by cytochrome P450, tyrosine metabolism, histidine metabolism, arginine and proline metabolism, and steroid hormone biosynthesis are downregulated in the epithelium of ovarian cancer (Figure 2B).

3.4. STRING-based PPI analysis identified epithelial-derived hub genes and significant modules in ovarian cancer

We investigated the PPI of all significant epithelial-derived DEGs. The PPI information of STRING is inputted into the Cytoscape for identifying and visualizing the hub genes and significant clusters. We identified the 460 hub genes (minimum degree of interaction is 10 with other DEGs) (Table S6). The top 20 hub genes (with the maximum degree of interaction) including *CDK1*, *CCNB1*, *AURKA*, *CDC20*, *CCNA2*, *BUB1*, *TOP2A*, *BUB1B*, *CCNB2*, and *CDC45* are shown in Figure 3. The overexpression of *CDK1* is associated with cancer growth and survival rate in epithelial ovarian cancer [56]. *CCNB1*, another top hub gene, is abnormally expressed and significantly involved in carboplatin-resistant epithelial ovarian cancer [57]. The expression level of *CCNB2* is distinguished between ovarian cancer and normal tissues [58].

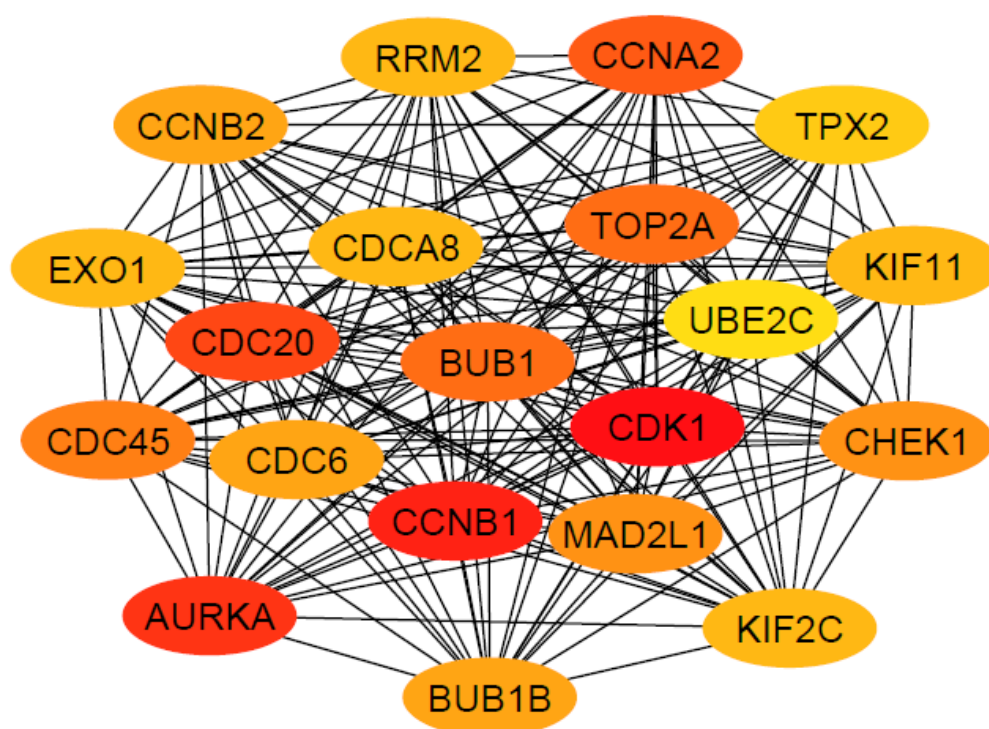


Figure 3. The top 20 hub genes and their degree of protein-protein interaction. The PPI network was established using the STRING-based analysis. Cytoscape plug-in cytoHubba tool was used to identify their degree of interactions.

We investigated the significant cluster-based analysis of 460 hub genes. The MCODE-based analysis identified 7 clusters (Score of MCODE > 5.0) from the original PPI networks. The description of MCODE derived clusters with their interacting gene lists is illustrated in Table 3. The top significant cluster 1 included 103 nodes and 4571 edges (Table 3). We identified the functional enrichment of KEGG pathways for all clusters by using the GSEA [29]. Interestingly, we found that all seven of the clusters are associated with the enrichment of KEGG pathways (FDR < 0.05). Gene set of cluster 1 is associated with the enrichment of the cell cycle and other cellular differentiation pathways (Table 3). Gene set of cluster 2 is mainly involved with immune regulation and cellular signaling (Table 3). Gene set of clusters 4 and 5 is mainly involved with cancer-associated pathways including pathways in cancer, Hedgehog signaling pathway, basal cell carcinoma, melanoma, Wnt signaling pathway, endocytosis, cytokine-cytokine receptor interaction, focal adhesion, regulation of actin cytoskeleton, MAPK signaling pathway, and TGF-beta signaling pathway (Table 3). Altogether, these results indicate that the epithelial tumor tissue-derived transcriptomes are contributed to ovarian cancer pathogenesis.

Table 3. MCODE identified significant 7 clusters from the PPI networks of DEGs and GSEA identified enrichment of KEGG pathways (FDR < 0.05) for a specific gene set of the individual cluster.

Cluster	Score of MCODE	Nodes	Edges	Node symbol	Enrichment of KEGG pathways (FDR < 0.05)
1	89.627	103	4571	<i>DEPDC1B, FOXM1, E2F8, KIF18B, AURKA, NUSAP1, PBK, TYMS, HJURP, DEPDC1, TOP2A, MCM2, RACGAP1, KIF20A, PRC1, CCNA2, CDC45, ECT2, MKI67, CDCA3, KIF4A, CDCA8, KIF2C, CEP55, TTK, TPX2, STIL, CENPF, CDCA5, CENPA, MELK, TACC3, KIF15, RAD51, SPAG5, FEN1, ESCO2, BIRC5, HMMR, CENPI, KIAA0101, MCM10, SPC25, RAD54L, CDC7, CASC5, CENPE, UHRF1, MND1, SGOL1, DTL, NEIL3, KIF11, TROAP, SKA3, NCAPG, ZWINT, DLGAP5, EXO1, TRIP13, FAM83D, FANCI, CENPU, NUF2, CDCA7, CDC25A, CCNE2, CDK1, PTTG1, DSCC1, CDC20, CKS2, E2F7, RRM2, SMC4, UBE2C, CKS1B, BUB1, PKMYT1, MAD2L1, MCM4, CCNB2, KIF23, BUB1B, KIF18A, GINS2, ESPL1, FAM64A, CCNB1, NCAPH, RAD51API, CDC6, OIP5, CENPM, ASF1B, CHEK1, CDKN3, KPNA2, CENPK, EZH2, NEK2, KIF14, TK1</i>	Cell cycle, Oocyte meiosis, Progesterone-mediated oocyte maturation, p53 signaling pathway, DNA replication

Continued on next page

Cluster	Score of MCODE	Nodes	Edges	Node symbol	Enrichment of KEGG pathways (FDR < 0.05)
2	16	16	120	<i>GNG12, GPSM2, NPY1R, CXCL6, GNG2, BDKRB1, HEBP1, CX3CL1, ADRA2C, APLNR, CXCR4, GNG11, LPAR3, ANXA1, PTGER3, GNAI1</i>	Chemokine signaling pathway, Neuroactive ligand-receptor interaction, Cytokine-cytokine receptor interaction, Leukocyte transendothelial migration, Axon guidance
3	7.6	26	95	<i>SOX9, CP, EPCAM, TIMP1, SPARCL1, FOXO1, ALDH1A1, KAT2B, CALCRL, STC2, CD24, LAMB1, GPR39, PTGDR, SDC2, PTGER4, KLF4, PTH2R, VWA1, CHGB, GOLM1, MEF2C, RAMP3, SCTR, ADCYAP1, CHRDL1</i>	Neuroactive ligand-receptor interaction
4	7.6	31	114	<i>LRP2, MMP7, PJA2, SYT1, TRIM4, WNT2B, NANOG, FBXL3, FBXL5, SPSB1, RCHY1, FGF13, GATA6, PACSIN3, STON2, MCAM, MUC1, PDGFRA, NOTCH1, BMP2, SH3GL2, CAV1, SFRP1, FBXL7, DAB2, FGF9, WNT7A, BMP4, SCARB2, MET, KITLG</i>	Pathways in cancer, Hedgehog signaling pathway, Basal cell carcinoma, Melanoma, Wnt signaling pathway, Endocytosis, Cytokine-cytokine receptor interaction, Melanogenesis, Focal adhesion, Regulation of actin cytoskeleton, MAPK signaling pathway, TGF-beta signaling pathway
5	6	9	24	<i>ANXA5, FGF2, VEGFA, KDR, SNAI2, ZEB2, VIM, FGF1, ZEB1</i>	Pathways in cancer, Melanoma, VEGF signaling pathway, Focal adhesion, Regulation of actin cytoskeleton, Cytokine-cytokine receptor interaction, MAPK signaling pathway
6	5.273	12	29	<i>RAB33A, PTX3, RAB39B, IL18, LCN2, NFKB1, HPSE, RAB31, RAB27B, GSDMD, JUP, RAB8B</i>	Cytosolic DNA-sensing pathway, Acute myeloid leukemia, NOD-like receptor signaling pathway
7	5	5	10	<i>MAP1LC3B, NBRI, GABARAPL2, PIK3C3, ATG101</i>	Regulation of autophagy

3.5. Epithelial derived hub transcriptomes are correlated with survival prognosis in ovarian cancer

We investigated the survival significance of the epithelium-derived all 460 significant hub genes in TCGA OV data. Our analysis revealed that the epithelium-derived upregulated genes included *SCNNIA* and *CDC43* (Figure 4A,B) and downregulated *SOX6* gene (Figure 4C) is significantly correlated with shorter survival time of ovarian cancer patients (Figure 4). The overexpression of *SCNNIA* exerts substantial roles in cell growth, invasion, and migration in ovarian cancer through regulating the epithelial to mesenchymal transition, and is a potential indicator for a patient's prognosis [59]. Chongxiang Chen et al. reported that the expression of the *CDC43* gene is associated

with the survival and tumorigenesis through the PLK1 pathway in ovarian cancer [60]. SOX6, a tumor suppressor, is downregulated in various cancers and associated with the inhibition of cellular proliferation, invasion, and tumor cell-induced angiogenesis of ovarian cancer cells [61].

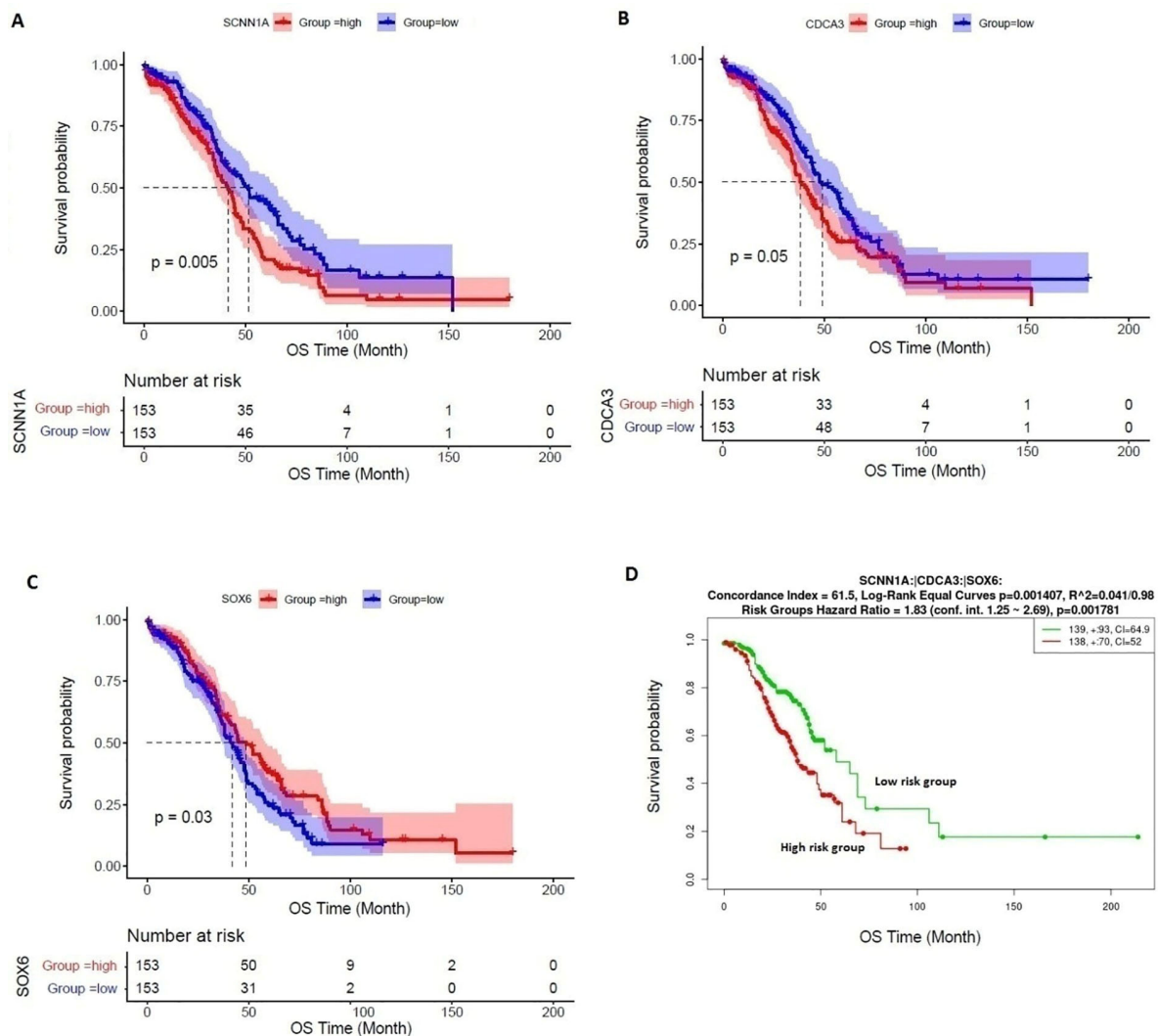


Figure 4. Identification of prognostic hub genes in ovarian cancer. A-B. Upregulated genes that included *SCNN1A* and *CDCA3* are significantly correlated with shorter survival time in ovarian cancer. C. Downregulated *SOX6* gene is significantly correlated with shorter survival time in ovarian cancer. D. The GEO dataset GSE9891 was used for the analysis of survival differences in the SurvExpress tool. three prognostic gene signatures are significantly associated with shorter overall survival time in the high-risk groups.

To verify the survival significance of the key three genes (*SCNN1A*, *CDCA3*, and *SOX6*), we inputted these prognostic three gene signatures into the SurvExpress tool (<http://bioinformatica.mty.itesm.mx/SurvExpress>) [62]. We used a GEO dataset GSE9891 ($n = 285$) [63] that is a built-in dataset in the SurvExpress tool (<http://bioinformatica.mty.itesm.mx/SurvExpress>). SurvExpress split the samples into two groups (high-risk group and low-risk group) based on the prognostic index and identify the significant survival differences between the two groups. Interestingly, we found that the three gene signatures are prognostic in overall survival (Figure 4D). The high-risk

group patients had significantly lower survival times than low-risk group patients (Figure 4D).

3.6 Epithelial tumor-derived prognostic hub genes are associated with immune infiltrations in ovarian cancer

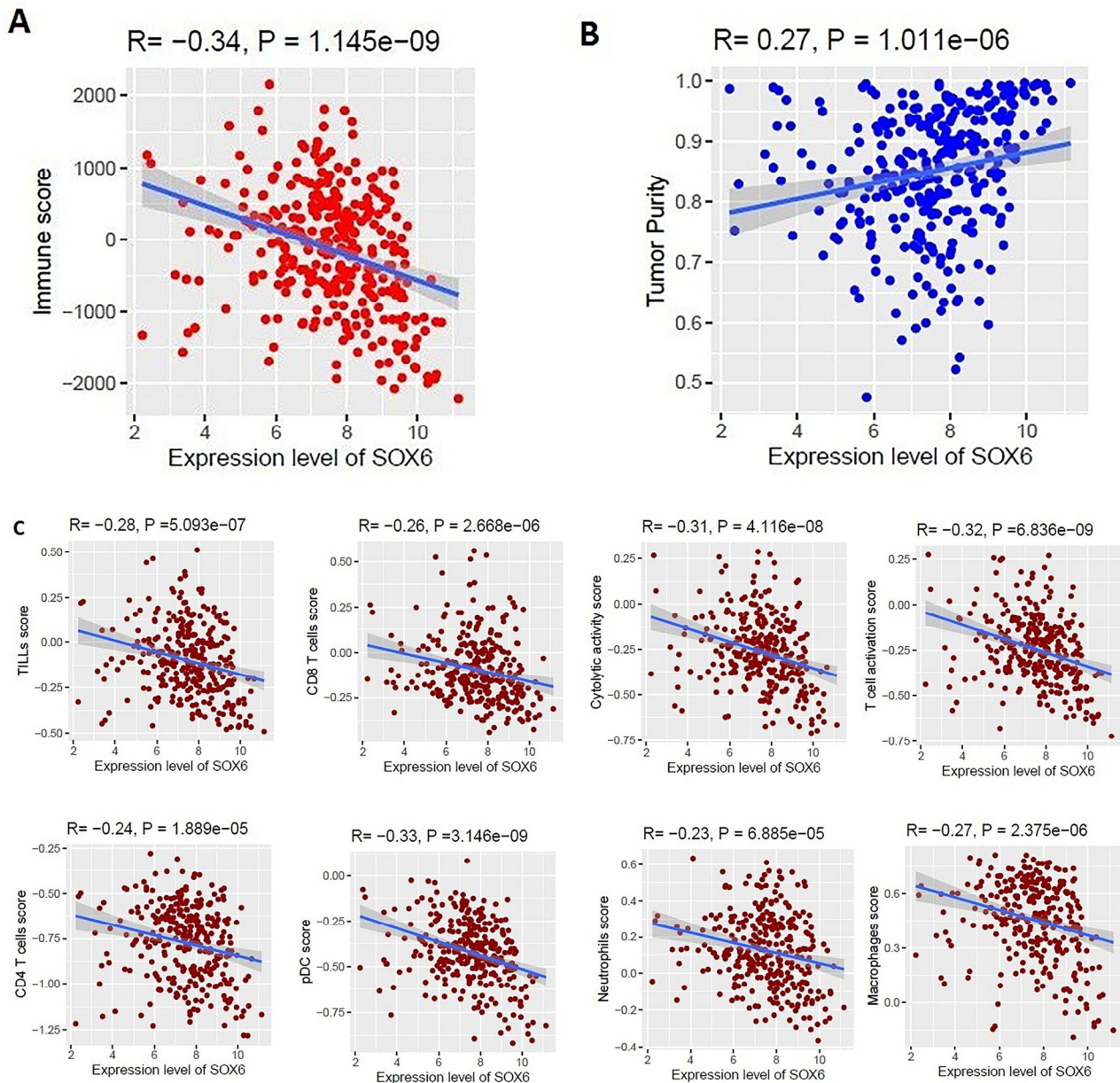


Figure 5. The expression level of *SOX6* is associated with the tumor microenvironment and immune infiltration in ovarian cancer. A. The expression level of *SOX6* is negatively correlated with immune scores ($R = -0.34, P < 0.001$). B. The expression level of *SOX6* is positively correlated with tumor purity ($R = 0.27, P < 0.001$). C. The expression level of *SOX6* is negatively correlated with ssGSEA scores of TILs, CD8+ T cells, CD4+ Regulatory T cells, cytolytic activity, T cell activation, pDC, neutrophils, and macrophages. (Spearman's correlation test, $P < 0.001$).

Table 4. Correlated immune markers with *SOX6* in ovarian cancer (TCGA OV cohort). R is Pearson's correlation; P is p-value; FDR is false discovery rate.

Immune signatures	ID	Immune markers	Name	R	P value	FDR value
CD8 T cell	925	<i>CD8A</i>	CD8a molecule	-0.26	5.87E-06	8.74E-05
Cytolytic activity	3001	<i>GZMA</i>	granzyme A	-0.27	1.73E-06	3.45E-05
	5551	<i>PRF1</i>	perforin 1	-0.30	1.25E-07	4.38E-06
T cell activation	3001	<i>GZMA</i>	granzyme A	-0.27	1.73E-06	3.45E-05
	3002	<i>GZMB</i>	granzyme B	-0.27	1.31E-06	2.74E-05
	4818	<i>NKG7</i>	natural killer cell granule protein 7	-0.26	3.72E-06	6.16E-05
	5551	<i>PRF1</i>	perforin 1	-0.30	1.25E-07	4.38E-06
	6348	<i>CCL3</i>	C-C motif chemokine ligand 3	-0.26	3.53E-06	5.90E-05
	6351	<i>CCL4</i>	C-C motif chemokine ligand 4	-0.31	3.79E-08	1.75E-06
	8530	<i>CST7</i>	cystatin F	-0.23	3.78E-05	3.76E-04
CD4 regulatory T cell	1493	<i>CTLA4</i>	cytotoxic T-lymphocyte associated protein 4	-0.26	4.69E-06	7.33E-05
	50943	<i>FOXP3</i>	forkhead box P3	-0.24	2.65E-05	2.80E-04
Macrophages	968	<i>CD68</i>	CD68 molecule	-0.29	3.66E-07	1.03E-05
	1536	<i>CYBB</i>	cytochrome b-245 beta chain	-0.24	2.61E-05	2.77E-04
	10457	<i>GPNMB</i>	glycoprotein nmb	-0.23	6.45E-05	5.71E-04
	23601	<i>CLEC5A</i>	C-type lectin domain containing 5A	-0.22	1.22E-04	9.47E-04
Neutrophils	2124	<i>EVI2B</i>	ecotropic viral integration site 2B	-0.23	6.91E-05	6.03E-04
	4332	<i>MNDA</i>	myeloid cell nuclear differentiation antigen	-0.25	6.50E-06	9.44E-05
	55350	<i>VNN3</i>	vanin 3	-0.24	2.35E-05	2.55E-04
pDC	2833	<i>CXCR3</i>	C-X-C motif chemokine receptor 3	-0.27	1.02E-06	2.24E-05
	3002	<i>GZMB</i>	granzyme B	-0.27	1.31E-06	2.74E-05
	122618	<i>PLD4</i>	phospholipase D family member 4	-0.20	3.03E-04	1.98E-03
	171558	<i>PTCRA</i>	pre T cell antigen receptor alpha	-0.25	1.14E-05	1.45E-04
TILs	914	<i>CD2</i>	CD2 molecule	-0.23	4.20E-05	4.09E-04
	915	<i>CD3D</i>	CD3d molecule	-0.25	1.41E-05	1.71E-04
	916	<i>CD3E</i>	CD3e molecule	-0.24	2.49E-05	2.68E-04
	919	<i>CD247</i>	CD247 molecule	-0.25	1.11E-05	1.42E-04
	925	<i>CD8A</i>	CD8a molecule	-0.26	5.87E-06	8.74E-05
		<i>CD86</i>	CD86 molecule	-0.29	1.73E-07	5.66E-06
	952	<i>CD38</i>	CD38 molecule	-0.25	1.20E-05	1.50E-04
	962	<i>CD48</i>	CD48 molecule	-0.31	2.05E-08	1.11E-06
	963	<i>CD53</i>	CD53 molecule	-0.30	7.56E-08	2.97E-06
	1043	<i>CD52</i>	CD52 molecule	-0.32	9.68E-09	6.45E-07
	1536	<i>CYBB</i>	cytochrome b-245 beta chain	-0.24	2.61E-05	2.77E-04
	1794	<i>DOCK2</i>	dedicator of cytokinesis 2	-0.21	2.88E-04	1.91E-03
	2124	<i>EVI2B</i>	ecotropic viral integration site 2B	-0.23	6.91E-05	6.03E-04
	2533	<i>FYB1</i>	FYN binding protein 1	-0.22	7.16E-05	6.19E-04
	2841	<i>GPR18</i>	G protein-coupled receptor 18	-0.25	8.27E-06	1.13E-04
3071	<i>NCKAP1L</i>	NCK associated protein 1 like	-0.26	4.66E-06	7.29E-05	

Continued on next page

Immune signatures	ID	Immune markers	Name	R	P value	FDR value
	3560	<i>IL2RB</i>	interleukin 2 receptor subunit beta	-0.27	1.99E-06	3.82E-05
	3561	<i>IL2RG</i>	interleukin 2 receptor subunit gamma	-0.27	2.48E-06	4.48E-05
	3587	<i>IL10RA</i>	interleukin 10 receptor subunit alpha	-0.24	1.75E-05	2.04E-04
	3676	<i>ITGA4</i>	integrin subunit alpha 4	-0.22	7.79E-05	6.61E-04
	3932	<i>LCK</i>	LCK proto-oncogene, Src family tyrosine kinase	-0.24	1.95E-05	2.21E-04
	3937	<i>LCP2</i>	lymphocyte cytosolic protein 2	-0.31	2.37E-08	1.21E-06
	4689	<i>NCF4</i>	neutrophil cytosolic factor 4	-0.27	2.38E-06	4.35E-05
	5341	<i>PLEK</i>	pleckstrin	-0.28	6.03E-07	1.50E-05
	5788	<i>PTPRC</i>	protein tyrosine phosphatase receptor type C	-0.24	2.37E-05	2.57E-04
	5790	<i>PTPRCAP</i>	protein tyrosine phosphatase receptor type C associated protein	-0.23	4.11E-05	4.02E-04
	6352	<i>CCL5</i>	C-C motif chemokine ligand 5	-0.29	2.63E-07	7.84E-06
	6793	<i>STK10</i>	serine/threonine kinase 10	-0.20	3.71E-04	2.33E-03
	6846	<i>XCL2</i>	X-C motif chemokine ligand 2	-0.32	1.36E-08	8.17E-07
	7293	<i>TNFRSF4</i>	TNF receptor superfamily member 4	-0.22	7.34E-05	6.31E-04
	8530	<i>CST7</i>	cystatin F	-0.23	3.78E-05	3.76E-04
	9046	<i>DOK2</i>	docking protein 2	-0.28	3.81E-07	1.06E-05
	9404	<i>LPXN</i>	leupaxin	-0.21	2.27E-04	1.57E-03
	10791	<i>VAMP5</i>	vesicle associated membrane protein 5	-0.24	2.26E-05	2.48E-04
	10859	<i>LILRB1</i>	leukocyte immunoglobulin like receptor B1	-0.24	2.04E-05	2.29E-04
	10870	<i>HCST</i>	hematopoietic cell signal transducer	-0.25	7.56E-06	1.05E-04
	11151	<i>CORO1A</i>	coronin 1A	-0.39	2.12E-12	1.36E-09
	22914	<i>KLRK1</i>	killer cell lectin like receptor K1	-0.22	1.19E-04	9.31E-04
	23157	<i>SEPTIN6</i>	septin 6	-0.20	3.52E-04	2.24E-03
	29851	<i>ICOS</i>	inducible T cell costimulator	-0.23	3.46E-05	3.50E-04
	51316	<i>PLAC8</i>	placenta associated 8	-0.22	1.21E-04	9.45E-04
	54440	<i>SASH3</i>	SAM and SH3 domain containing 3	-0.30	8.70E-08	3.32E-06
		<i>GIMAP4</i>	GTPase, IMAP family member 4	-0.27	1.06E-06	2.31E-05
	55340	<i>GIMAP5</i>	GTPase, IMAP family member 5	-0.24	1.99E-05	2.24E-04
	55423	<i>SIRPG</i>	signal regulatory protein gamma	-0.27	2.30E-06	4.24E-05
	55843	<i>ARHGAP15</i>	Rho GTPase activating protein 15	-0.26	3.03E-06	5.27E-05
	63940	<i>GPSM3</i>	G protein signaling modulator 3	-0.24	3.09E-05	3.19E-04
	64098	<i>PARVG</i>	parvin gamma	-0.25	1.24E-05	1.53E-04
	64231	<i>MS4A6A</i>	membrane spanning 4-domains A6A	-0.25	6.90E-06	9.87E-05
	64333	<i>ARHGAP9</i>	Rho GTPase activating protein 9	-0.27	1.91E-06	3.72E-05
	80342	<i>TRAF3IP3</i>	TRAF3 interacting protein 3	-0.25	1.30E-05	1.60E-04
	139818	<i>DOCK11</i>	dedicator of cytokinesis 11	-0.28	5.53E-07	1.41E-05

Since survival time of patients is correlated with immunological responses in human cancers [64], it is essential to identify the correlation of prognostic hub genes with the immune infiltrations. We investigated the correlations between the expression levels of three hub genes (*SCNN1A*, *CDCA3*,

and *SOX6*) and the levels of immune signatures and tumor purity in the TME of OV. Interestingly, we found that the expression level of *SOX6* is negatively correlated with immune scores ($R=-0.34$, $P < 0.001$) (Figure 5A), and positively correlated with tumor purity ($R=0.27$, $P<0.001$) (Figure 5B), but the expression level of *SCNN1A* and *CDC43* are not correlated with immune scores and tumor purity. Then, we investigated the correlation of three prognostic hub genes (*SCNN1A*, *CDC43*, and *SOX6*) with the several immune stimulatory and inhibitory signatures including B cells, TILs, CD8+ T cells, CD4+ Regulatory T cells, cytolytic activity, T cell activation, pDC, neutrophils, CAFs, and macrophages. We found that the expression of *SOX6* is negatively correlated with the infiltration of TILs, CD8+ T cells, CD4+ Regulatory T cells, cytolytic activity, T cell activation, pDC, neutrophils, and macrophages (Figure 5C). The expression level of *SCNN1A* and *CDC43* is not correlated with immune infiltrations. Therefore, we identified the correlations of *SOX6* with the immune markers (all markers selected from the significant immune signatures). Interestingly, we found that the expression of *SOX6* negatively correlated with 68 immune markers including *CD8A*, *PRF1*, *GZMA*, *GZMB*, *NKG7*, *PRF1*, *CCL3*, *CCL4*, *CST7*, *CXCR3*, and *IL10RA* in ovarian cancer (Table 5). Tumor immune infiltration is a key indicator in the progression of ovarian cancer [65]. In epithelial ovarian cancer, tumor-infiltrating T cells are predictors of prognosis and biological basis of treatment outcomes [66]. In intratumoral TME, the accumulation of CD8+ T cells is associated with the survival of high-grade serous ovarian carcinoma patients [67]. Our immunological analysis indicated that the epithelial-derived-transcriptomes are associated with the modulation of the immune microenvironment in ovarian cancer.

3.7 Three prognostic hub genes are mutated in ovarian cancer

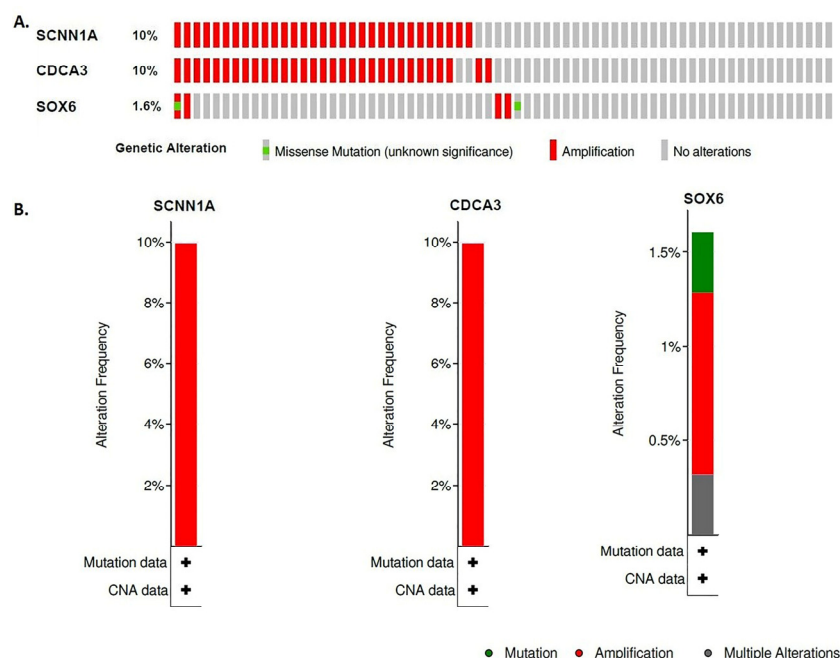


Figure 6. Genetical alterations of *SCNN1A*, *CDC43*, and *SOX6* in the ovarian epithelial tumor. A. *SCNN1A* (10%), *CDC43* (10%), and *SOX6* (1.6 %) genes are mutated in the ovarian epithelial tumor. B. The genetic alterations of *SCNN1A* and *CDC43* are amplification and the genetic alterations of *SOX6* are mutation, amplification, and multiple alterations in the ovarian epithelial tumor.

We used the OV epithelial tumor data ($n = 311$) in cBioPortal (<http://www.cbioportal.org/>) to identify the genetic alterations of three prognostic hubs genes (*SCNN1A*, *CDCA3*, and *SOX6*). Queried *SCNN1A*, *CDCA3*, and *SOX6* genes are altered in 36 (12%) of queried patients/samples.

We found that the *SCNN1A* and *CDCA3* are altered in 10 % of patients (Figure 6A). In addition, *SOX6* is altered in 1.6 % of patients (Figure 6A). The genetic alterations of *SCNN1A* and *CDCA3* included amplification and the genetic alterations of *SOX6* included mutation, amplification, and multiple alterations (Figure 6B).

3.8 Diagnostic efficacy evaluation of prognostic in combined 5 datasets of ovarian epithelium

We speculate that these six genes (*SCNN1A*, *CDCA3*, and *SOX6*) have diagnostic value in combined 5 datasets of ovarian epithelium. We used the combined 5 datasets (GSE14407 [3], GSE27651 [18], GSE38666 [19,20], GSE40595 [4], and GSE54388 [21]) of the ovarian epithelium to validate our hypothesis, and the results showed that the ROC curve of the expression levels of *SCNN1A* (AUC = 0.60) and *SOX6* (AUC = 0.744) showed excellent diagnostic value for ovarian epithelial tumor and normal ovarian epithelial cells (Figure 7).

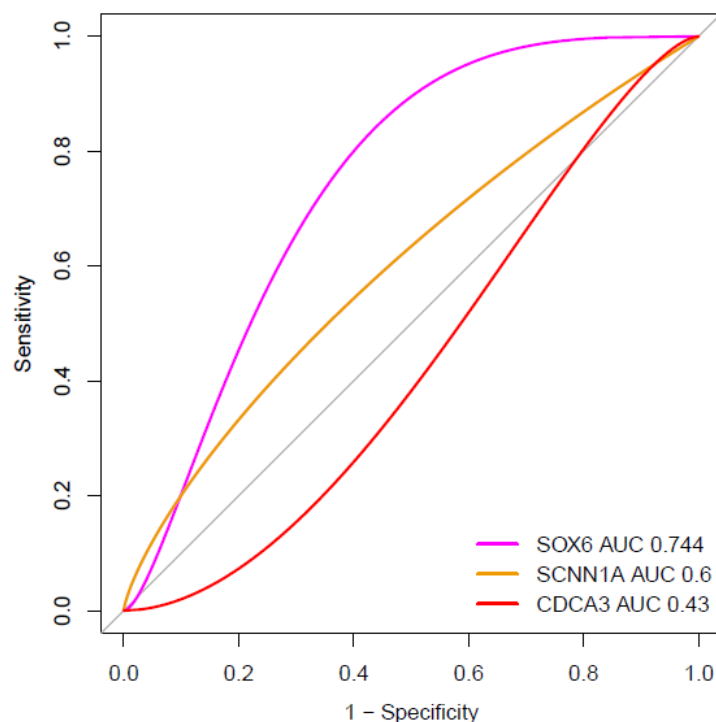


Figure 7. Evaluation of diagnostic efficacy of key prognostic genes in combined 5 datasets of ovarian epithelium.

The receiver operating characteristic (ROC) curve of prognostic genes in combined 5 datasets (GSE14407 [3], GSE27651 [18], GSE38666 [19,20], GSE40595 [4], and GSE54388 [21]) of ovarian tumor epithelium and normal ovarian epithelium. The expression levels of *SCNN1A* (AUC = 0.60) and *SOX6* (AUC = 0.744) showed diagnostic value in the ovarian epithelial tumor.

4. Discussion

Since bioinformatic studies are crucial for identifying the key signaling genes and pathways in human diseases [68–71], we aimed to explore the ovarian cancer epithelium-derived transcriptomes for identifying the substantial biomarkers. We identified DEGs in ovarian tumor epithelium by a meta-analysis of five gene expression datasets (Tables S2 and S3). Based on the significant DEGs, we identified pathways that are associated with the significant DEGs. The upregulated pathways were mainly involved in cellular development, cancer, and metabolism, and the downregulated pathways involved in immune-regulation and metabolism (Figure 2). The cell cycle, the top enriched upregulated pathway, is involved in ovarian cancer pathogenesis [68]. The downregulation of immune cells is directly associated with enhancing the pathogenesis of ovarian cancer through the release of various cytokines and chemokines [72]. Therefore, our investigated results are consistent with the role of DEGs in the enrichment of pathways that cause ovarian cancer pathogenesis. These results revealed the abnormal alterations of cellular development, immune regulation, and metabolism-related pathways in the compartment of the ovarian cancer epithelium. In addition, we identified hub genes that are interacted with other DEGs. We identified 21 and 11 transcription factors that are associated with regulating the upregulated and downregulated DEGs, respectively (Table S4). In ovarian epithelial cancer, TFs are substantial contributors to cancer risk and somatic development [73]. These TFs could be a unique target for the development of novel precision medicine strategies for ovarian cancer.

Moreover, we found that several clusters are associated with the hub gene signatures. These clusters are associated with the enrichment level of KEGG pathways (Table 3). Pathway analysis revealed that the significant clusters are mainly involved with cancer, immune regulation, and cellular signaling (Table 3). Then, we found that the expression of three hub genes (*SCNN1A*, *CDCA3*, and *SOX6*) are significantly correlated with the shorter overall survival time of ovarian cancer patients (Figure 4). Since the level of immune infiltration is an independent predictor of a patient's survival in cancer [74], we analyzed the correlation of *SCNN1A*, *CDCA3*, and *SOX6* with the immune infiltration levels in the ovarian tumor. First, we revealed the association of these three hub genes with the ovarian tumor microenvironment. The expression level of *SOX6* is negatively correlated with immune scores, indicating that *SOX6* is associated with lower the immunity (Figure 5A,B). Second, the expression level of *SOX6* is negatively correlated with infiltrations of TILs, CD8⁺ T cells, CD4⁺ Regulatory T cells, cytolytic activity, T cell activation, pDC, neutrophils, and macrophages in epithelial ovarian cancer, suggesting that the *SOX6* expression is linked with lowering the immune infiltrations in ovarian cancer (Figure 5C). Third, the 68 immune markers (we selected all markers of the significant immune signatures (Figure 4C)) including *CD8A*, *GZMA*, *PRF1*, *GZMB*, *NKG7*, *CCL3*, *CCL4*, and *CCL5* are negatively correlated with the expression levels of *SOX6*, demonstrating that the *SOX6* expression is related to the reduced expression levels of immune markers (Table 4). Altogether. It indicates that the expression of *SOX6* is associated with the suppression of tumor immunity in ovarian epithelial cancer. The expression of *SCNN1A* and *CDCA3* is not associated with the regulation of immune infiltrations in epithelial cancer. In addition, *SCNN1A*, *CDCA3*, and *SOX6* are genetically altered in ovarian cancer. The genetic alteration of *SCNN1A* and *CDCA3* is amplification and the genetic alterations of *SOX6* are mutation, amplification, and multiple alterations in ovarian cancer (Figure 6).

We downloaded and integrated publicly available five datasets for this study. Some of the

previous studies scatteredly used these datasets to implement various purposes [75–77]. For example, Dan Yang et al. identified hub genes and therapeutic drugs by using the four datasets including GSE54388 and GSE40595 [75]. However, none of the authors specifically using these five datasets together for identifying the key biomarkers in the epithelial compartment of ovarian cancer. As per our knowledge, this is the first time, we have integrated the five transcriptomic datasets from the same platform. One of the major advantages of this study is that we selected only laser capture microdissected human ovarian cancer epithelial samples with control. Another advantage of this study is that we integrated the microarray datasets from the same platform (Affymetrix Human Genome U133 Plus 2.0) to reduce the data platform heterogeneity. Moreover, we did a meta-analysis to identify the key genes in ovarian tumor epithelial samples. This meta-analysis method usually gives more conservative results (less DEGs but more confident) [22]. The major drawback of this study is that the ovarian epithelial-associated key genes and networks identified by bioinformatics analysis have not been validated by experimental analysis. Thus, although our findings could provide potential biomarkers for ovarian cancer diagnosis and prognosis, as well as therapeutic targets, further experimental and clinical validation is necessary to transform these results into practical application in ovarian cancer treatments.

5. Conclusions

The identification of ovarian cancer epithelial-derived key biomarkers may provide insight into the association of these genes with survival prognosis and tumor immunity. Epithelial-derived key transcriptomes may be crucial indicators of the effectiveness of ovarian cancer diagnosis and treatment.

Acknowledgments

All authors have contributed to the research concept, design, and interpretation of the data. All authors review and approved the final version of the manuscript.

The external fund was not collected for this project.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- 1 L. M. Hurwitz, P. F. Pinsky, B. Trabert, General population screening for ovarian cancer, *Lancet*, **397** (2021), 2128–2130.
- 2 R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer statistics, 2021, *CA Cancer J. Clin.*, **71** (2021), 7–33.
- 3 N. J. Bowen, L. D. Walker, L. V. Matyunina, S. Logani, K. A. Totten, B. B. Benigno, 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.*, **2** (2009), 71.

- 4 T. L. Yeung, C. S. Leung, K. K. Wong, G. Samimi, M. S. Thompson, J. Liu, et al., TGF- β modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment, *Cancer Res.*, **73** (2013), 5016–5028.
- 5 J. Farley, L. L. Ozbun, M. J. Birrer, Genomic analysis of epithelial ovarian cancer, *Cell Res.*, **18** (2008), 538–548.
- 6 M. Ezzati, A. Abdullah, A. Sharifabrizi, J. Hou, M. Kopf, J. K. Stedman, et al., Recent advancements in prognostic factors of epithelial ovarian carcinoma, *Int. Sch. Res. Not.*, **2014** (2014), e953509.
- 7 M. L. Drakes, P. J. Stiff, Regulation of ovarian cancer prognosis by immune cells in the tumor microenvironment, *Cancers*, **10** (2018), 302.
- 8 T. Gui, C. Yao, B. Jia, K. Shen, Identification and analysis of genes associated with epithelial ovarian cancer by integrated bioinformatics methods, *PloS One*, **16** (2021), e0253136.
- 9 J. Liu, H. Meng, S. Li, Y. Shen, H. Wang, W. Shan, et al., Identification of potential biomarkers in association with progression and prognosis in epithelial ovarian cancer by integrated bioinformatics analysis, *Front. Genet.*, **10** (2019), 1031.
- 10 D. Matei, T. G. Graeber, R. L. Baldwin, B. Y. Karlan, J. Rao, D. D. Chang, Gene expression in epithelial ovarian carcinoma, *Oncogene*, **21** (2002), 6289–6298.
- 11 P. Israelsson, E. Dehlin, I. Nagaev, E. Lundin, U. Ottander, L. Mincheva-Nilsson, Cytokine mRNA and protein expression by cell cultures of epithelial ovarian cancer-methodological considerations on the choice of analytical method for cytokine analyses, *Am. J. Reprod. Immunol.*, **84** (2020), e13249.
- 12 K. P. Prahm, C. K. Høgdall, M. A. Karlsen, I. J. Christensen, G. W. Novotny, E. Høgdall, MicroRNA characteristics in epithelial ovarian cancer, *PloS One*, **16** (2021), e0252401.
- 13 A. A. Alshamrani, Roles of microRNAs in ovarian cancer tumorigenesis: two decades later, what have we learned?, *Front. Oncol.*, **10** (2020), 1084.
- 14 S. Udhaya Kumar, D. Thirumal Kumar, R. Bithia, S. Sankar, R. Magesh, M. Sidenna, et al., Analysis of differentially expressed genes and molecular pathways in familial hypercholesterolemia involved in atherosclerosis: a systematic and bioinformatics approach, *Front. Genet.*, **11** (2020), 734.
- 15 J. Wan, S. Jiang, Y. Jiang, W. Ma, X. Wang, Z. He, et al., Corrigendum: data mining and expression analysis of differential lncRNA ADAMTS9-AS1 in prostate cancer, *Front. Genet.*, **11** (2020), 361.
- 16 U. K. S., B. Rajan, T. K. D., A. P. V., T. Abunada, S. Younes, et al., Involvement of essential signaling cascades and analysis of gene networks in diabetes, *Genes*, **11** (2020), 1256.
- 17 D. Fu, B. Zhang, L. Yang, S. Huang, W. Xin, Development of an immune-related risk signature for predicting prognosis in lung squamous cell carcinoma, *Front. Genet.*, **11** (2020), 978.
- 18 E. R. King, C. S. Tung, Y. T. M. Tsang, Z. Zu, G. T. M. Lok, M. T. Deavers, et al., The anterior gradient homolog 3 (AGR3) gene is associated with differentiation and survival in ovarian cancer, *Am. J. Surg. Pathol.*, **35** (2011), 904–912.
- 19 C. Huang, E. A. Clayton, L. V. Matyunina, L. D. McDonald, B. B. Benigno, F. Vannberg, et al., Machine learning predicts individual cancer patient responses to therapeutic drugs with high accuracy, *Sci. Rep.*, **8** (2018), 16444.
- 20 L. N. Lili, L. V. Matyunina, L. D. Walker, B. B. Benigno, J. F. McDonald, Molecular profiling predicts the existence of two functionally distinct classes of ovarian cancer stroma, *BioMed Res. Int.*, **2013** (2013), 846387.

- 21 T. L. Yeung, C. S. Leung, K. K. Wong, A. Gutierrez-Hartmann, J. Kwong, D. M. Gershenson, et al., ELF3 is a negative regulator of epithelial-mesenchymal transition in ovarian cancer cells, *Oncotarget*, **8** (2017), 16951–16963.
- 22 J. Xia, E. E. Gill, R. E. W. Hancock, NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data, *Nat. Protoc.*, **10** (2015), 823–844.
- 23 W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods, *Biostat. Oxf. Engl.*, **8** (2007), 118–127.
- 24 W. G. Cochran, The combination of estimates from different experiments, *Biometrics*, **10** (1954), 101–129.
- 25 Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B Methodol.*, **57** (1995), 289–300.
- 26 P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.*, **13** (2003), 2498–2504.
- 27 R. Janky, A. Verfaillie, H. Imrichová, B. Van de Sande, L. Standaert, V. Christiaens, et al., iRegulon: from a gene list to a gene regulatory network using large motif and track collections, *PLoS Comput. Biol.*, **10** (2014), e1003731.
- 28 J. Wang, Y. Wu, M. N. Uddin, R. Chen, J. P. Hao, Identification of potential key genes and regulatory markers in essential thrombocythemia through integrated bioinformatics analysis and clinical validation, *Pharmacogenomics Pers. Med.*, **14** (2021), 767–784.
- 29 A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci.*, **102** (2005), 15545–15550.
- 30 M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: new perspectives on genomes, pathways, diseases and drugs, *Nucleic Acids Res.*, **45** (2017), D353–D361.
- 31 D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, et al., STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Res.*, **47** (2019), D607–D613.
- 32 C. H. Chin, S. H. Chen, H. H. Wu, C. W. Ho, M. T. Ko, C. Y. Lin, cytoHubba: identifying hub objects and sub-networks from complex interactome, *BMC Syst. Biol.*, **8** (2014), S11.
- 33 G. D. Bader, C. W. Hogue, An automated method for finding molecular complexes in large protein interaction networks, *BMC Bioinformatics*, **4** (2003), 2.
- 34 T. Therneau, A package for survival analysis in R, 2021. Available from: <https://cran.r-project.org/web/packages/survival/vignettes/survival.pdf>.
- 35 K. Yoshihara, M. Shahmoradgoli, E. Martínez, R. Vegesna, H. Kim, W. Torres-Garcia, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, *Nat. Commun.*, **4** (2013), 2612.
- 36 S. Hänzelmann, R. Castelo, J. Guinney, GSEA: gene set variation analysis for microarray and RNA-seq data, *BMC Bioinformatics*, **14** (2013), 7.
- 37 I. Tirosh, B. Izar, S. M. Prakadan, M. H. Wadsworth, D. Treacy, J. J. Trombetta, et al., Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq, *Science*, **352** (2016), 189–196.

- 38 D. Aran, Z. Hu, A. J. Butte, xCell: digitally portraying the tissue cellular heterogeneity landscape, *Genome Biol.*, **18** (2017), 220.
- 39 T. Davoli, H. Uno, E. C. Wooten, S. J. Elledge, Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy, *Science*, **355** (2017), eaaf8399.
- 40 Z. Liu, M. Li, Z. Jiang, X. Wang, A comprehensive immunologic portrait of triple-negative breast cancer, *Transl. Oncol.*, **11** (2018), 311–329.
- 41 E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, et al., The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, *Cancer Discov.*, **2** (2012), 401–404.
- 42 X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. C. Sanchez, et al., pROC: an open-source package for R and S+ to analyze and compare ROC curves, *BMC Bioinformatics*, **12** (2011), 77.
- 43 Q. Yang, R. Wang, B. Wei, C. Peng, L. Wang, G. Hu, et al., Candidate biomarkers and molecular mechanism investigation for glioblastoma multiforme utilizing WGCNA, *BioMed Res. Int.*, **2018** (2018), e4246703.
- 44 M. N. Uddin, R. Akter, M. Li, Z. Abdelrahman, Expression of SARS-COV-2 cell receptor gene ACE2 is associated with immunosuppression and metabolic reprogramming in lung adenocarcinoma based on bioinformatics analyses of gene expression profiles, *Chem. Biol. Interact.*, **335** (2021), 109370.
- 45 V. Gómez-Rubio, ggplot2 - elegant graphics for data analysis, *J. Stat. Softw.*, **77** (2017), 1–3.
- 46 L. Zhao, Y. Li, Z. Zhang, J. Zou, J. Li, R. Wei, et al., Meta-analysis based gene expression profiling reveals functional genes in ovarian cancer, *Biosci. Rep.*, **40** (2020), BSR20202911.
- 47 D. Chaves-Moreira, M. A. Mitchell, C. Arruza, P. Rawat, S. Sidoli, R. Nameki, et al., PAX8 orchestrates an angiogenic program through interaction with SOX17, preprint, *BioRxiv*, 2020.09.09.290387.
- 48 X. Liu, Y. Gao, Y. Lu, J. Zhang, L. Li, F. Yin, Upregulation of NEK2 is associated with drug resistance in ovarian cancer, *Oncol. Rep.*, **31** (2014), 745–754.
- 49 C. L. Au-Yeung, T. L. Yeung, A. Achreja, H. Zhao, K. P. Yip, S. Y. Kwan, et al., ITLN1 modulates invasive potential and metabolic reprogramming of ovarian cancer cells in omental microenvironment, *Nat. Commun.*, **11** (2020), 3546.
- 50 R. Januchowski, P. Zawierucha, M. Andrzejewska, M. Ruciński, M. Zabel, Microarray-based detection and expression analysis of ABC and SLC transporters in drug-resistant ovarian cancer cell lines, *Biomed. Pharmacother.*, **67** (2013), 240–245.
- 51 Y. Wang, F. Shao, L. Chen, ALDH1A2 suppresses epithelial ovarian cancer cell proliferation and migration by downregulating STAT3, *Oncotargets Ther.*, **11** (2018), 599–608.
- 52 S. R. Rosario, M. D. Long, H. C. Affronti, A. M. Rowsam, K. H. Eng, D. J. Smiraglia, Pan-cancer analysis of transcriptional metabolic dysregulation using The Cancer Genome Atlas, *Nat. Commun.*, **9** (2018), 5330.
- 53 D. Reimer, S. Sadr, A. Wiedemair, S. Stadlmann, N. Concin, G. Hofstetter, et al. Clinical relevance of E2F family members in ovarian cancer-an evaluation in a training set of 77 patients, *Clin. Cancer Res.*, **13** (2007), 144–151.
- 54 L. Zhan, Y. Zhang, W. Wang, E. Song, Y. Fan, B. Wei, E2F1: a promising regulator in ovarian carcinoma, *Tumor Biol.*, **37** (2016), 2823–2831.

- 55 S. Hein, S. Mahner, C. Kanowski, T. Löning, F. Jänicke, K. Milde-Langosch, Expression of Jun and Fos proteins in ovarian tumors of different malignant potential and in ovarian cancer cell lines, *Oncol. Rep.*, **22** (2009), 177–183.
- 56 W. Yang, H. Cho, H. Y. Shin, J. Y. Chung, E. S. Kang, E. J. Lee, et al., Accumulation of cytoplasmic Cdk1 is associated with cancer growth and survival rate in epithelial ovarian cancer, *Oncotarget*, **7** (2016), 49481–49497.
- 57 S. Wei, J. Liu, Y. Shi, X. Zhang, Y. Yang, Q. Song, Exploration of the sequential gene changes in epithelial ovarian cancer induced by carboplatin via microarray analysis, *Mol. Med. Rep.*, **16** (2017), 3155–3160.
- 58 W. Li, Z. Liu, B. Liang, S. Chen, X. Zhang, X. Tong, et al., Identification of core genes in ovarian cancer by an integrative meta-analysis, *J. Ovarian Res.*, **11** (2018), 94.
- 59 L. Wu, Z. H. Ling, H. Wang, X. Y. Wang, J. Gui, Upregulation of SCNN1A promotes cell proliferation, migration, and predicts poor prognosis in ovarian cancer through regulating epithelial-mesenchymal transformation, *Cancer Biother. Radiopharm.*, **34** (2019), 642–649.
- 60 C. Chen, S. Chen, M. Luo, H. Yan, L. Pang, C. Zhu, et al., The role of the CDCA gene family in ovarian cancer, *Ann. Transl. Med.*, **8** (2020), 190.
- 61 Y. Li, M. Xiao, F. Guo, The role of Sox6 and Netrin-1 in ovarian cancer cell growth, invasiveness, and angiogenesis, *Tumour Biol.*, **39** (2017), 1010428317705508.
- 62 R. Aguirre-Gamboa, H. Gomez-Rueda, E. Martínez-Ledesma, A. Martínez-Torteya, R. Chacolla-Huaranga, A. Rodriguez-Barrientos, et al., SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis, *PloS One*, **8** (2013), e74250.
- 63 R. W. Tothill, A. V. Tinker, J. George, R. Brown, S. B. Fox, S. Lade, et al., Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome, *Clin. Cancer Res.*, **14** (2008), 5198–5208.
- 64 Y. Liu, Survival correlation of immune response in human cancers, *Oncotarget*, **10** (2019), 6885–6897.
- 65 J. Wang, F. H. C. Cheng, J. Tedrow, W. Chang, C. Zhang, A. K. Mitra, Modulation of immune infiltration of ovarian cancer tumor microenvironment by specific subpopulations of fibroblasts, *Cancers*, **12** (2020), 3184.
- 66 W. Wang, W. Zou, J. R. Liu, Tumor-infiltrating T cells in epithelial ovarian cancer: predictors of prognosis and biological basis of immunotherapy, *Gynecol. Oncol.*, **151** (2018), 1–3.
- 67 S. Lieber, S. Reinartz, H. Raifer, F. Finkernagel, T. Dreyer, H. Bronger, et al., Prognosis of ovarian cancer is associated with effector memory CD8⁺ T cell accumulation in ascites, CXCL9 levels and activation-triggered signal transduction in T cells, *Oncoimmunology*, **7** (2018), e1424672.
- 68 S. U. Kumar, D. T. Kumar, R. Siva, C. G. P. Doss, H. Zayed, Integrative bioinformatics approaches to map potential novel genes and pathways involved in ovarian cancer, *Front. Bioeng. Biotechnol.*, **7** (2019), 391.
- 69 S. Mishra, M. I. Shah, S. Udhaya Kumar, D. Thirumal Kumar, C. Gopalakrishnan, A. M. Al-Subaie, et al., Chapter Eleven - Network analysis of transcriptomics data for the prediction and prioritization of membrane-associated biomarkers for idiopathic pulmonary fibrosis (IPF) by bioinformatics approach, *123* (2020), 241–273.

- 70 H. Yan, G. Zheng, J. Qu, Y. Liu, X. Huang, E. Zhang, et al., Identification of key candidate genes and pathways in multiple myeloma by integrated bioinformatics analysis, *J. Cell. Physiol.*, **234** (2019), 23785–23797.
- 71 S. Udhaya Kumar, D. Thirumal Kumar, R. Siva, C. George Priya Doss, S. Younes, N. Younes, et al., Dysregulation of signaling pathways due to differentially expressed genes from the B-cell transcriptomes of systemic lupus erythematosus patients-a bioinformatics approach, *Front. Bioeng. Biotechnol.*, **8** (2020), 276.
- 72 B. Charbonneau, E. L. Goode, K. R. Kalli, K. L. Knutson, M. S. DeRycke, The immune system in the pathogenesis of ovarian cancer, *Crit. Rev. Immunol.*, **33** (2013), 137–164.
- 73 R. Nameki, H. Chang, J. Reddy, R. I. Corona, K. Lawrenson, Transcription factors in epithelial ovarian cancer: histotype-specific drivers and novel therapeutic targets, *Pharmacol. Ther.*, **220** (2021), 107722.
- 74 H. Ohtani, Focus on TILs: prognostic significance of tumor infiltrating lymphocytes in human colorectal cancer, *Cancer Immun.*, **7** (2007), 4.
- 75 D. Yang, Y. He, B. Wu, Y. Deng, N. Wang, M. Li, et al., Integrated bioinformatics analysis for the screening of hub genes and therapeutic drugs in ovarian cancer, *J. Ovarian Res.*, **13** (2020), 10.
- 76 H. Feng, Z. Y. Gu, Q. Li, Q. H. Liu, X. Y. Yang, J. J. Zhang, Identification of significant genes with poor prognosis in ovarian cancer via bioinformatical analysis, *J. Ovarian Res.*, **12** (2019), 35.
- 77 X. Yang, S. Zhu, L. Li, L. Zhang, S. Xian, Y. Wang, et al., Identification of differentially expressed genes and signaling pathways in ovarian cancer by integrated bioinformatics analysis, *OncoTargets Ther.*, **11** (2018), 1457–1474.



AIMS Press

©2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)