

Meta-analysis based gene expressional profiling reveals functional genes in ovarian cancer

Lin Zhao

Shandong Academy of Medical Sciences

Yuhui Li

Shandong First Medical University and Shandong Academy of Medical Sciences

Zhen Zhang

Shandong First Medical University and Shandong Academy of Medical Sciences

Hong Guo

Shandong Desheng Biological Engineering CO.Ltd

Jianfu Li

Houston Methodist Research Institute

Ran Wei

Shandong First Medical University & Shandong Academy of Medical Sciences

Qiang Guo

Shandong First Medical University & Shandong Academy of Medical Sciences

Xiaoxiao Zhu

Shandong First Medical University & Shandong Academy of Medical Sciences

Chu Chu

School of Medicine and Life Sciences, University of Jinan Shandong Academy of Medical Sciences

Yunhong Zhang

School of Medicine and Life Sciences, University of Jinan. Shandong Academy of Medical Sciences

Bin Wang

Affiliated Hospital of Shandong University of Traditional Chinese Medicine

Jinbo Yue

Shandong Tumor Hospital and Institute

Xia Li (✉ 786735868@qq.com)

Shandong Academy of medical sciences

Research article

Keywords: Ovarian cancer, Gene expression omnibus, Microarray analysis, Differentially expressed genes

Posted Date: March 5th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-16156/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Bioscience Reports on November 1st, 2020.

See the published version at <https://doi.org/10.1042/BSR20202911>.

Abstract

Background Ovarian cancer causes high mortality rate worldwide, and despite numerous attempts, the outcome for patients with ovarian cancer are still not well improved. Microarray based gene expression analysis provides with valuable information for discriminating functional genes in ovarian cancer development and progression. However, due to the differences in experimental design, the results varied significantly across individual datasets. **Methods** In the present study, the data of gene expression in ovarian cancer was downloaded from Gene Expression Omnibus and 16 studies were included. A meta-analysis based gene expression analysis was performed to identify differentially expressed genes (DEGs). The most differentially expressed genes in our meta-analysis were selected for gene expression and gene function validation. **Results** A total of 972 DEGs with P-value<0.001 were identified in ovarian cancer, including 541 up-regulated genes and 431 down-regulated genes, among which 92 additional DEGs were found as gained DEGs. Top five up-and down-regulated genes were selected for the validation of gene expression profiling. Among these genes, up-regulated CD24 , SOX17 , WFDC2 , EPCAM , INAVA , and down-regulated AOX1 were revealed to be with consistent expressional patterns in clinical patient samples of ovarian cancer. Gene functional analysis demonstrated that up-regulated WFCD2 and INAVA promoted ovarian cancer cell migration, WFDC2 enhanced cell proliferation, while down-regulated AOX1 was functional in inducing cell apoptosis of ovarian cancer. **Conclusion** Our study shed light on the molecular mechanisms underlying the development of ovarian cancer, and facilitated the understanding of novel diagnostic and therapeutic targets in ovarian cancer.

Background

Ovarian cancer is the second mostly encountered gynecologic cancer worldwide with high mortality and low rate of five-year survival due to the difficulty of early diagnosis and high rates of recurrence and metastasis [1]. If ovarian cancer could be diagnosed at the early stage, 90% patients could survive for more than five years with appropriate treatment [2]. Despite numerous attempts, the common molecular events for early diagnosis of ovarian cancer remain to be established. Gene expression profile analysis is a powerful research strategy, which integrates data in genetics, molecular transcription, and functional genomics to reveal dysregulated genes between patients and healthy donors. Microarray provides increasing body of gene-wide transcriptional data regarding ovarian cancer. For example, Moreno et al. proposed that cellular proliferation, cell cycle, DNA damage, and apoptosis were up-regulated in ovarian cancer by using microarrays [3]. Bowen et al. identified alterations in expression levels of gene products functioning in several signaling pathways including Wnt, Notch, TGF β /BMP and canonical cell cycle in ovarian cancer by microarray analysis [4]. However, results vary between studies due to the diversity in cohort selection, specimen source, and experimental designs. Combining different microarray datasets is advantageous to enhance statistical power to detect the dysregulated genes that might be functional in ovarian cancer development and progression.

Microarray data integration-based meta-analyses depend on efficient in silico tools. With the advances of ever-growing theories and bio-informatics tools, we can now employ in silico tools to efficiently combine

multiple microarray datasets ignoring different populations, experimental designs, and different specimen sources [5–7]. NetworkAnalyst is a useful web-based tool, which functions in many aspects such as preliminary data processing, sample annotation, batch effect adjustment, dataset integration, and results visualization [8–12]. To maximally overcome the impact caused by the differences in study design and platform usage among different datasets, “Combining Effect Size (ES)” analysis and Random Effect Modeling (REM) were applied to achieve more consistent and accurate results by taking into consideration of both direction and magnitude of gene expression changes. In this study, 16 eligible microarray datasets for ovarian cancer were selected from publicly available dataset repositories. Totally, 972 differentially expressed genes (DEGs) with P -value < 0.001 were identified, including 541 up- and 431 down-regulated genes, respectively. Interestingly, by comparing the list of meta-based DEGs to that of individual microarray dataset, 92 additional DEGs were demonstrated as gained DEGs. Expression profiles for the CD24 molecule (CD24), Sex determining region Y-box 17 (SOX17), WAP four-disulfide core domain 2 (WFDC2), the epithelial cell adhesion molecule (EPCAM), innate immunity activator (INAVA), and Aldehyde oxidase 1 (AOX1) were validated in clinical cancer samples of ovarian cancer, which are among top five dysregulated DEGs revealed by our meta-analysis. Functional analysis uncovered that WFDC2, INAVA, and AOX1 affect cell proliferation, migration, and apoptosis respectively. Our results provided with novel molecular mechanisms for the development and progression of ovarian cancer and propose potential targets for the prevention and treatment for clinics [13].

Materials Identification and Selection of Eligible Gene Expression Datasets for Meta-analysis

Microarray-based gene expression profiling studies for ovarian cancer were identified in the PubMed database (<http://www.pubmed.gov>), Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/gds/>) and ArrayExpress dataset of the European Molecular Biology Laboratory-European Bioinformatics Institute (<http://www.ebi.ac.uk/arrayexpress/>). The following keywords were used: (“Ovarian neoplasms”[Mesh]) and (microarray or gene profile or gene profiling). Eligible studies and datasets should follow these inclusive criteria: (1) case and control studies of human; (2) analysis of gene expression profiling; (3) comparable experimental conditions and untreated; (4) available complete raw and processed microarray data. Studies were excluded if they were: (1) letters, abstracts, meta-analysis, review articles and human case report; (2) cell lines were used in experimental design; (3) RT-PCR only for profiling studies; (4) studies without healthy control. All the datasets and references, which are conformed to the above-mentioned criteria, were manually screened. The complete workflow was shown in Fig. 1 for eligible dataset selection. The latest search was performed on Mar. 15, 2019.

Data Extraction and Processing

Full text and supplementary materials of selected articles were extracted from each identified study: GEO series accession number, number of patients and healthy donors, specimen sources, platform of

microarray and references. The series matrix files were downloaded from GEO datasets for all studies, and common Entrez IDs were used to substitute all the gene probes in accordance with the corresponding microarray platforms, with the exception of GSE17308, in which the gene probe was substituted by Genbank ID. Before integrative meta-analysis, individual dataset was normalized by R-mediated mean, log₂ transformation and quantile normalization. Expression data of patients with ovarian cancer and healthy donors were defined as class 2 and class 1 respectively, according to the guidelines of NetworkAnalyst [6].

Batch Effect Adjustment and Individual Data Analysis

Batch effect correction option in NetworkAnalyst was used to reduce potential study-specific batch effects, by which the normalized individual datasets were subjected to the well-established ComBat procedures [11]. Empirical Bayes methods were used to stabilize the expression ratios for genes with very high or very low ratios, stabilize gene variances by shrinking variances across all other genes, and protect their inference from artifacts in the data. To compare the sample clustering patterns with and without applying the ComBat procedures, the results were visually examined using the principal component analysis (PCA).

Meta-analysis

We conducted the meta-analysis using NetworkAnalyst, a web interface for integrative statistical analysis and visualization. In the option of “multiple gene expression data”, all the datasets were uploaded to the “multiple gene expression data” input area and analyzed in a streamlined manner, including data processing for Entrez ID/Genbank ID, annotation check, and PCA plot examining, confirmation of data normalization, individual DEG analysis and data summary. For NetworkAnalyst based DEGs discrimination, the cut-off of P-value was adjusted to 0.001. Furthermore, all datasets were subjected to integrity check to ensure that the merged data could be carried out by ES combination, which would generate more biologically consistent meta-based DEGs by incorporation of both the magnitude and direction of gene fold change. While performing Cochran’s Q test on the integration data, the random effects model was suggested to be used when Q values deviates significantly from a chi-squared distribution; otherwise, the fixed-effects model was selected. The Q values deviates significantly from a chi-squared distribution of the current study, so this study used REM which assumes that each study contains a random effect size that could incorporate unknown cross-study heterogeneities, as demonstrated by Cochran’s Q tests [10, 12]. “Define custom signature” tool from NetworkAnalyst was used to produce the heatmap visualization for both top 25 up-and down-regulated genes.

Patients and Specimens

Thirty epithelial ovarian carcinoma samples and thirty unpaired normal ovary samples were collected at the Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences. No local or systemic neo-adjuvant radiotherapy, or/and chemotherapy and targeted therapy had been received. Written informed consent was obtained from each patient. Our study was

approved by the Ethics Committee of Shandong First Medical University and Shandong Academy of Medical Sciences.

Cell Culture and siRNA Transfection

OVCAR-3 cell line was obtained from Procell Life Science & Technology Co., Ltd (Wuhan, China). HOSEpiC, and CAOV-3 cell was purchased from BeNa culture collection (Beijing, China). CAOV-3 cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) (Bioind, Kibbuiz, Israel). OVCAR-3 cells were cultured in Roswell Park Memorial Institute (RPMI) 1640 Medium (Bioind, Kibbuiz, Israel). All culture medium was supplemented with 10% Fetal Bovine Serum (FBS) (Bioind, Kibbuiz, Israel) and 1% penicillin/streptomycin. Cells were cultured in a humidified incubator at 37 °C and 5% CO₂. siRNAs for CD24, SOX17, WFDC2, EPCAM, INAVA, basonuclin 1 (BNC1), AOX1, gasdermin E (GSDME), receptor accessory protein 1 (REEP1) and heart and neural crest derivatives expressed 2 (HAND2) were transfected into OVCAR-3 and CAOV-3 cells by HiperFect at the concentration of 50 nM according to the manufacturer's instructions. The sequences of siRNAs used for CD24, SOX17, WFDC2, EPCAM, INAVA, BNC1, AOX1, GSDME, REEP1, and HAND2 knockdown are listed in Table S1.

Real-time quantitative RT-PCR (qRT-PCR)

Total RNA was extracted by using TRIzol Reagent (Invitrogen, Carlsbad, USA) according to the manufacturer's instructions. RNA was reverse transcribed with the mRNA 1st Strand cDNA Synthesis Kit (Vazyme, Nanjing, China) for mRNA. qRT-PCR assay was performed on an Applied Biosystems 7500 instrument (Applied Biosystems, Foster, USA) by using SYBR Green (Invitrogen, Carlsbad, USA). Relative RNA quantification was calculated via the comparative $2^{-\Delta\Delta Ct}$ method. The relative expression of indicated genes was normalized to that of GAPDH in each sample.

Cell Apoptosis Assay

For apoptosis assay, cells were seeded into 12-well plates and allowed to adhere overnight. Then, the cells were digested with EDTA-free trypsin and left to dissociate into single cells. After being washed with PBS, the harvested cells were resuspended at a density of 1×10^6 cells/ml in $1 \times$ binding buffer and double stained with FITC Annexin V and Propidium Iodide using FITC Annexin V Apoptosis Detection Kit I (BD Biosciences, USA). A minimum of 1×10^4 single-cell events were acquired by flow cytometry (BD FACSVerse, USA), and cell apoptosis was analysed by using FlowJo software.

Cell Proliferation Assay

OVCAR-3 and CAOV-3 cells were seeded at 2000 cells per well in 96-well plates after transfection. CCK8 (Cell Counting Kit-8) assay was used to measure cell number and viability, and the absorbance was determined using a spectrophotometric plate reader at 450 nm (SPECTRAMAX 190, USA). The data were represented as mean \pm SD. Each experiment was repeated three times.

Cell Migration Assays

Cell migration was assessed by transwell experiments. For cell migration assay, indicated cells were trypsinized, re-suspended in serum-free DMEM medium, and placed in the upper chamber (4×10^4 cells/well). Then, DMEM medium supplemented with 10% FBS was added to the lower chamber. The plates were incubated for 48 h. Cells in the upper chamber were completely removed with a cotton swab. Cells migrating into the lower chamber were washed with PBS, fixed in 4% paraformaldehyde, and stained with 0.5% crystal violet. Finally, the cells were counted under a microscope in five random fields.

Statistical Analysis

The web-server NetworkAnalyst analyzed the data and performed meta-analysis by REM-based combined ES statistics. This study adjusted P-value < 0.001 to discriminate the DEGs. All experiments were performed at least three independent times, unless otherwise stated. Values are presented as the mean \pm SD. The 2-tailed Student's t-test was used for comparisons between two groups. Graphpad 8.0 was used for statistical analysis of data and data plotting. P-value ≤ 0.05 was considered significant.

Results

Selection of eligible microarray datasets

A total of eligible 16 microarray datasets were selected for ovarian cancer (GSE26712, GSE40595, GSE66957, GSE18520, GSE27651, GSE6008, GSE38666, GSE17308, GSE14407, GSE14001, GSE29220, GSE29450, GSE52037, GSR15578, GSE36668 and GSE23391) [1, 4, 14–25], respectively, as described in methods (Fig. 1A). All the 16 datasets were screened according to the inclusion and exclusion criteria. A total of 135/505 (healthy control/ patient) samples were included in this study. Four microarray platforms were applied in the selected datasets, including Affymetrix Human Genome U133 Plus 2.0 Array, Affymetrix Human Genome U133A Array, Rosetta/Merck Human RSTA Custom Affymetrix 2.0 microarray and PC Human Operon 21 k v2. In addition, various tissues, such as stroma, epithelium, mucinous, clear cell, and cell-free saliva, were used for the microarray analysis. In the process of data integration, patient samples from different sources were not distinguished, for the purpose to reveal the common gene signature in ovarian cancer. Table 1 presents the detailed information for each dataset, including GEO accession number, sample composition, sample source, and the corresponding references.

Table 1
Characteristics of each selected microarray dataset for the meta- analysis

GEO accession no.	Sample(Ctl/Pt)	Sample source	Platform	Reference
GSE26712	n = 195(10/185)	Ovarian tissue	Affymetrix Human Genome U133A Array	19
GSE40595	n = 77(14/63)	Ovarian tissue	Affymetrix Human Genome U133 Plus 2.0 Array	14
GSE66957	n = 69(12/57)	Ovarian tissue	Rosetta/Merck Human RSTA Custom Affymetrix 2.0 microarray	Unavailable
GSE18520	n = 63(10/53)	Ovarian tissue	Affymetrix Human Genome U133 Plus 2.0 Array	15
GSE27651	n = 49(6/43)	Ovarian tissue	Affymetrix Human Genome U133 Plus 2.0 Array	16
GSE6008	n = 103(4/99)	Ovarian tissue	Affymetrix Human Genome U133A Array	20
GSE38666	n = 45(20/25)	Ovarian tissue	Affymetrix Human Genome U133 Plus 2.0 Array	1
GSE17308	n = 32(4/28)	Ovarian tissue	PC Human Operon 21 k v2	17
GSE14407	n = 24(12/12)	Ovarian tissue	Affymetrix Human Genome U133 Plus 2.0 Array	4
GSE14001	n = 23(3/20)	Ovarian tissue	Affymetrix Human Genome U133 Plus 2.0 Array	21
GSE29220	n = 22(11/11)	Cell-free saliva	Affymetrix Human Genome U133 Plus 2.0 Array	18
GSE29450	n = 20(10/10)	Ovarian tissue	Affymetrix Human Genome U133 Plus 2.0 Array	22
GSE52037	n = 20(10/10)	Ovarian tissue	Affymetrix Human Genome U133 Plus 2.0 Array	23
GSE15578	n = 17(6/11)	Ovarian tissue	Affymetrix Human Genome U133 Plus 2.0 Array	Unavailable
GSE36668	n = 12(4/8)	Ovarian tissue	Affymetrix Human Genome U133 Plus 2.0 Array	24
GSE23391	n = 4(1/3)	Ovarian tissue	Affymetrix Human Genome U133 Plus 2.0 Array	25

Identification of common DEGs in ovarian cancer

The workflow for meta-analysis in this study was illustrated in Fig. 1B. To identify the transcriptional signature between healthy donors and patients, a total of 16 datasets were simultaneously analyzed by NetworkAnalyst. The 16 datasets were uploaded in succession, and each dataset was processed by gene ID matching, sample (control/patient) annotation and individual DEGs identification. To remove batch effects among different datasets, “ComBat”-based batch effects adjustment were performed [26], and the results with and without the adjustment were visualized by PCA plots respectively (Fig. 2A-B). All the 16 gene expressional microarray data were then integrated and merged. Meta-analysis was conducted following the Cochran’s Q test REM and ES statistical methods, which facilitated to reveal the DEGs between healthy donors and patients across different microarray datasets by permitting variable true effect size and integrating unknown cross-study heterogeneities. We found a total of 972 DEGs, including 541 up- and 431 down-regulated genes across the 16 datasets with significance threshold of adjusted P-value < 0.001. Besides, by taking the advantage of meta-analysis, there were 92 gained genes with consistent expressional profiles and 12638 lost genes with inconsistent expressional changes across datasets acquired by comparing meta-based DEGs with that of individual gene expression analysis (Fig. 2C). Expression profile of the top 25 up- and down-regulated genes among the 972 identified DEGs was visualized by heatmap (Fig. 2D). CD24, SOX17, and EPCAM were among the most significantly up-regulated genes, while BNC1, AOX1 and GSDME were among the most significantly down-regulated genes (Table 2).

Table 2
Top 20 DEGs identified by Meta- analysis across different datasets

EntrezID	Gene symbol	Gene Name	Combined ES	Adjusted p-value
Top10 Over-expressed Genes				
100133941	CD24	CD24 molecule	2.8049	7.02E-08
64321	SOX17	SRY (sex determining region Y)-box 17	2.6537	5.12E-06
4072	EPCAM	epithelial cell adhesion molecule	2.5631	4.70E-07
10406	WFDC2	WAP four-disulfide core domain 2	2.4890	1.08E-06
55765	INAVA	innate immunity activator	2.3726	9.61E-07
54845	ESRP1	epithelial splicing regulatory protein 1	2.2399	6.31E-06
2122	MECOM	MDS1 and EVI1 complex locus	2.2230	7.31E-07
1999	ELF3	E74 like ETS transcription factor 3	2.2060	7.98E-08
1356	CP	ceruloplasmin	2.0245	5.12E-06
4582	MUC1	mucin 1, cell surface associated	1.8852	1.33E-05
Top10 Under-expressed Genes				
646	BNC1	basonuclin 1	-3.3816	1.37E-04
316	AOX1	aldehyde oxidase 1	-3.1022	3.44E-05
1687	GSDME	gasdermin E	-2.9762	6.78E-05
65055	REEP1	receptor accessory protein 1	-2.6894	6.83 E-04
79804	HAND2	heart and neural crest derivatives expressed 2	-2.5998	1.37E-04
10332	CLEC4M	C-type lectin domain family 4 member M	-2.5927	2.50E-04
56245	C21orf62	chromosome 21 open reading frame 62	-2.5673	4.02E-05
80310	PDGFD	platelet derived growth factor D	-2.5513	6.75E-07
54810	GIPC2	GIPC PDZ domain containing family member 2	-2.5123	3.15E-09
8622	PDE8B	phosphodiesterase 8B	-2.5070	3.50E-04

Gene expressional confirmation by clinical epithelial ovarian cancer samples

In order to validate the results of differentially expressed genes for our meta-analysis, we collected thirty ovarian cancer samples and unpaired thirty normal ovarian tissues for control. RNA was extracted for each sample and the mRNA levels of each gene were quantified by real-time PCR analysis. The top five up- and down-regulated genes in ovarian cancers compared with normal control were selected for the analysis, including CD24, SOX17, WFDC2, EPCAM, INAVA, AOX1, REEP1, BNC1, HAND2, and GSDME. It has been revealed by our results that the expressional profile for the top five up-regulated genes of CD24, SOX17, WFDC2, EPCAM, and INAVA were validated to be consistent in the clinical samples with our meta-analysis (Fig. 3A-E). However, only AOX1 of top five down-regulated genes was verified to be decreased in expression for expanding sample of ovarian cancers (Fig. 3F), with that of REEP1, BNC1, HAND2, and GSDME unchanged (Fig. S1A-D). The results indicated that although there are some artificial effects for the results of our meta-analysis, the assay is still of high reliability providing with genes that are potentially functional in ovarian cancer development and progress.

WFDC2 knockdown inhibited proliferation of ovarian cancer cells

Then, we tried to figure out whether these verified differentially expressed genes participated in the pathogenesis of ovarian cancer. Two ovarian cancer cell lines of OVCAR-3 and CAOV-3 were selected for the functional analysis. siRNAs targeting each gene were designed and synthesized by genepharma. The knockdown efficiency for each gene was confirmed by qPCR analysis in both cells (Fig. S2A-B). So we transfected the cells with siRNAs targeting CD24, SOX17, WFDC2, EPCAM, INAVA, and AOX1, and detected cell proliferation changes by CCK8 analysis. The results showed that WFDC2 knockdown inhibited cell proliferation in both OVCAR-3 and CAOV-3 cells (Fig. 4A-B), while knockdown of CD24, EPCAM, INAVA, and AOX1 leave cell proliferation largely unchanged. However, SOX17 knockdown slightly inhibited cell proliferation of CAOV-3 cells (Fig. S3). Collectively, our data demonstrated that WFDC2 promoted cell proliferation of ovarian cancer cells; SOX17 might play marginal role in cell proliferation and the function is likely dependent on cellular context; CD24, EPCAM, INAVA, and AOX1 have no effect on ovarian cancer proliferation.

AOX1 knockdown decreases cell death ratio for ovarian cancer cells

Deregulated cell death is an essential characteristic of cancer cells, including ovarian cancer. Screening for genes functional in cell death regulation is always a prospective way to find targets for cancer treatment. With the hypothesis that these validated differentially expressed genes might modulate cell death of ovarian cancer cells, we detected the ratio of apoptotic cell by flow cytometry analysis. OVCAR-3 and CAOV-3 cells were transfected with siRNAs specific for these six genes, and only when AOX1

expression was decreased, the cell death ratio will be lowered for both early (Annexin V⁺/PI⁻) and late (Annexin V⁺/PI⁺) apoptotic cells compared with NC transfected cells (Fig. 5A-B). The data of cell death affected by CD24, SOX17, WFDC2, EPCAM, and INAVA knockdown were provided in Fig. S4. Collectively, these data suggested a role for AOX1 in modulating cell death of ovarian cancer.

INAVA and WFDC2 promote cell migration for ovarian cancer cells

In addition to cell proliferation and cell death, the property of cancer cell metastasis is the most important reason for cancer recurrence and low survival rate. Cell migration assay is designed to detect whether certain genes are functional in cancer metastasis and provided with potential targets for cancer intervention. We knockdown the validated six genes in OVCAR-3 and CAOV-3 cells, and check whether cell migration ability was affected. According to our results, WFDC2 and INAVA knockdown dramatically inhibited cell migration for both cell lines (Fig. 6), while knockdown of CD24, EPCAM, SOX17, AOX1 did not affect cell migration (Fig. S5). These data suggested that WFDC2 and INAVA promote cell migration and might be valuable targets for cancer therapy and worth further study for the functional mechanism.

Discussion

Despite advances in surgical and medical therapy, the overall survival rate for patients with ovarian cancer remains largely unchanged over the past decades. The lethality of ovarian cancer is mainly due to the difficulties in diagnosis at early stage and the lack of effective treatments for patients with an advanced status [27]. Since different ovarian cancer cells are biologically, genetically, and molecularly similar, identification of common molecular markers may contribute to reveal the appropriate risk factors, as well as make preventive, diagnostic and therapeutic decisions for these patients. Recently, microarray technology has been developing rapidly and widely used to uncover the general genetic alterations in progression of diseases, which enables the identification of targets for diagnosis, therapeutic, and prognosis of tumors [28]. Despite significant amounts of studies using microarray-based technology to identify molecular markers in different ovarian cancer cells, contradictory results have been reported due to the diversity in patient selection, tissue source, and study designs [29]. Therefore, in order to recognize the common gene signatures underlying ovarian cancer, we performed meta-analysis by using 16 publically available microarray datasets. In the present study, 972 DEGs were found at with P-value < 0.001, including 541 up- and 431 down-regulated respectively. Interestingly, by comparing the list of meta-based DEGs and that of individual microarray dataset, 92 additional DEGs were found as gained DEGs, which means our meta-based analysis lead to discovery of new DEGs compared with individual dataset. By uncovering shared gene expressional profiles, this study highlighted potential diagnostic and therapeutic biomarkers in ovarian cancer, and might aid in understanding the molecular mechanisms of its development and progression.

Among the up-regulated DEGs, the expression for top five genes were further validated in clinical samples, including CD24, EPCAM, SOX17, WFDC2, and INAVA. The protein encoded by CD24 is a small, heavily glycosylated mucin-like cell surface protein [30]. Accumulating evidences had demonstrated that CD24 was involved in epithelial-mesenchymal transition (EMT) and its increased expression was correlated with poor prognosis in patients of epithelial ovarian cancer [31, 32]. EPCAM, which encodes an epithelia cell localized type I transmembrane glycoprotein, was initially identified as a tumor-associated antigen, which highly expressed in ovarian cancer [33, 34]. Collectively, these reported data proved the accuracy for our meta-analysis based DEGs. SOX17, an important member of SOX family, which encodes an antagonist of Wnt signaling pathway, was found to be frequently methylated in high grade serous ovarian cancer [35, 36]. It is well known that promoter methylation typically represses gene expression. Contradictory, SOX17 expression was revealed to be up-regulated in our meta-analysis and clinical samples. So the expressional profiles and relevance of SOX17 in ovarian cancer remains to be further validated. WFDC2 is located on human chromosome 20q12-13.1, the protein product for which is also called human epididymis protein 4 (HE4). The gene has been reported to highly express in epithelial ovarian cancer comparing with normal epithelium [37]. Besides, our functional study on WFDC2 effects on cell proliferation consolidated the previous discovery that WFDC2 gene overexpression could promote ovarian tumor growth [38]. Thus, targeted screening for WFDC2 protein inhibition might benefits for the treatment of ovarian cancer. INAVA gene, also called C1orf106, has been recently reported essential for optimal MAPK and NF- κ B activation in primary human cells by recruiting 14-3-3 τ [39]. And inflammatory bowel disease (IBD) associated polymorphisms in INAVA is revealed to play important roles in intestinal immune homeostasis by regulating pattern recognition receptor (PRR) mediated signaling transduction and bacterial clearance. In our analysis, INAVA was found to promote migration of ovarian cancer cells, which unmask a novel role for INAVA in ovarian cancer progression and pointed potential implication for INAVA in targeted ovarian cancer treatment.

For the down-regulated DEGs, expressional confirmation was also performed in the clinical samples for the top five dysregulated genes of AOX1, REEP1, BNC1, HAND2, and GSDME. However, only AOX1 was verified with consistent expressional profile with the result of our meta-analysis, which implied that artificial effects for meta-analysis existed and could be eliminated by further validation. AOX1 is a xenobiotic enzyme, which is a family member of xanthine oxidase of cytosolic molybdoenzymes [40]. The enzyme catalyzes the oxidation of aldehydes and aromatic N-heterocycles [41]. Recently, AOX1 has been uncovered to participate in the development and progression of a series of cancers. For example, loss of AOX1 expression by epigenetic regulation could promote bladder cancer via metabolic deregulations; the SNP rs73055188 of AOX1 locus is related to prostate cancer survival time, and AOX1 gene expression level is correlated with recurrence of prostate cancer [42]. However, the roles for AOX1 in ovarian cancer have not been addressed. In this study, we found that knockdown of AOX1 in ovarian cancer cells inhibited cell apoptosis. Although the mechanism underlying which AOX1 functions in ovarian cancer has not been elucidated, the therapeutic value for AOX1 is still worth further exploration.

Conclusions

In our study, several differentially expressed genes, including *WFDC2*, *INAVA*, and *AOX1*, in ovarian cancer revealed by our meta-analysis have been validated to potentially participate in ovarian cancer development and progression. Importantly, our study will shed light on the novel molecular mechanisms underlying the pathogenesis of ovarian cancer, and facilitated the understanding of novel diagnostic and therapeutic targets in ovarian cancer.

Abbreviations

GEO-Gene Expression Omnibus, DEG-differentially expressed gene, CD24-CD24 molecule, SOX17-SRY-box transcription factor 17, WFDC2-WAP four-disulfide core domain 2, EPCAM-epithelial cell adhesion molecule, INAVA-innate immunity activator, BNC1-basonuclin 1, AOX1-aldehyde oxidase 1, GSDME-gasdermin E, REEP1-receptor accessory protein 1, and HAND2-heart and neural crest derivatives expressed 2.

Declarations

Acknowledgements

Not applicable.

Funding

This work was supported by the Natural Science Foundation of China (81873337, 81601442, 81704116), the Primary Research and Development Plan of Shandong Province (2016GSF202016, 2017GSF218013, 2019GSF107036, 2019GSF108176), the Project of Transformation in High-tech Achievements (2013ZHZX2A0405), the Natural Science Foundation of Shandong Province (ZR2019MH039, ZR2018PH042), Taishan Scholars (Tsqn201812125), the Innovation Project of Shandong Academy of Medical Sciences and the Academic promotion program of Shandong First Medical University.

Availability of data and materials

Not applicable.

Authors' contributions

J.B.Y., and X.L. designed this study; L.Z., Z.Z., and C.C. searched databases and collected full-text papers; L.Z., Z.Z., Q.G., J.F.L. and X.X.Z. extracted the data; R.W. X.Q.Y., and Y.H.Z., performed the quality assessment; L.Z., Y.H.L., Z.Z., and H.G. operated related experiments and analyzed the data; L.Z., Y.H.L. and J.F.L. contributed resources to facilitate the analysis; L.Z., Z.Z. and X.L. wrote the manuscript. All authors reviewed the manuscript.

Ethics approval and consent to participate

The present study was approved by the Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences. Written informed consent was obtained from each patient. Our study was approved by the Ethics Committee of Shandong First Medical University and Shandong Academy of Medical Sciences.

Patient consent for publication

Not applicable.

Competing interests

The authors declare no conflicts of interest in this work.

References

1. Lili LN, Matyunina LV, Walker LD, Benigno BB, McDonald JF: Molecular profiling predicts the existence of two functionally distinct classes of ovarian cancer stroma. *Biomed Res Int* 2013, 2013:846387.
2. Urban N, Drescher C: Current and future developments in screening for ovarian cancer. *Womens Health (Lond)* 2006, 2(5):733-42.
3. Moreno CS, Matyunina L, Dickerson EB, Schubert N, Bowen NJ, Logani S, Benigno BB, McDonald JF: Evidence that p53-mediated cell-cycle-arrest inhibits chemotherapeutic treatment of ovarian carcinomas. *PLoS One* 2007, 2(5):e441.
4. Bowen NJ, Walker LD, Matyunina LV, Logani S, Totten KA, Benigno BB, McDonald JF: Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells. *BMC Med Genomics* 2009, 2:71.
5. Tseng GC, Ghosh D, Feingold E: Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Res* 2012, 40(9):3785-99.
6. O'Donoghue SI, Gavin AC, Gehlenborg N, Goodsell DS, Heriche JK, Nielsen CB, North C, Olson AJ, Procter JB, Shattuck DW et al: Visualizing biological data-now and in the future. *Nat Methods* 2010, 7(3 Suppl):S2-4.
7. Gomez-Cabrero D, Abugessaisa I, Maier D, Teschendorff A, Merkschlager M, Gisel A, Ballestar E, Bongcam-Rudloff E, Conesa A, Tegner J: Data integration in the era of omics: current and future challenges. *BMC Syst Biol* 2014, 8 Suppl 2:11.
8. Tuller T, Atar S, Ruppin E, Gurevich M, Achiron A: Common and specific signatures of gene expression and protein-protein interactions in autoimmune diseases. *Genes Immun* 2013, 14(2):67-82.
9. Silva GL, Junta CM, Mello SS, Garcia PS, Rassi DM, Sakamoto-Hojo ET, Donadi EA, Passos GA: Profiling meta-analysis reveals primarily gene coexpression concordance between systemic lupus erythematosus and rheumatoid arthritis. *Ann N Y Acad Sci* 2007, 1110:33-46.

10. Higgs BW, Zhu W, Richman L, Fiorentino DF, Greenberg SA, Jallal B, Yao Y: Identification of activated cytokine pathways in the blood of systemic lupus erythematosus, myositis, rheumatoid arthritis, and scleroderma patients. *Int J Rheum Dis* 2012, 15(1):25-35.
11. Jha PK, Vijay A, Sahu A, Ashraf MZ: Comprehensive Gene expression meta-analysis and integrated bioinformatic approaches reveal shared signatures between thrombosis and myeloproliferative disorders. *Sci Rep* 2016, 6:37099.
12. Xia J, Gill EE, Hancock RE: NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat Protoc* 2015, 10(6):823-44.
13. Haidich AB: Meta-analysis in medical research. *Hippokratia* 2010, 14(Suppl 1):29-37.
14. Yeung TL, Leung CS, Wong KK, Samimi G, Thompson MS, Liu J, Zaid TM, Ghosh S, Birrer MJ, Mok SC: TGF-beta modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment. *Cancer Res* 2013, 73(16):5016-28.
15. Siamakpour-Reihani S, Owzar K, Jiang C, Turner T, Deng Y, Bean SM, Horton JK, Berchuck A, Marks JR, Dewhirst MW et al: Prognostic significance of differential expression of angiogenic genes in women with high-grade serous ovarian carcinoma. *Gynecol Oncol* 2015, 139(1):23-9.
16. King ER, Tung CS, Tsang YT, Zu Z, Lok GT, Deavers MT, Malpica A, Wolf JK, Lu KH, Birrer MJ et al: The anterior gradient homolog 3 (AGR3) gene is associated with differentiation and survival in ovarian cancer. *Am J Surg Pathol* 2011, 35(6):904-12.
17. Merritt MA, Parsons PG, Newton TR, Martyn AC, Webb PM, Green AC, Papadimos DJ, Boyle GM: Expression profiling identifies genes involved in neoplastic transformation of serous ovarian cancer. *BMC Cancer* 2009, 9:378.
18. Lee YH, Kim JH, Zhou H, Kim BW, Wong DT: Salivary transcriptomic biomarkers for detection of ovarian cancer: for serous papillary adenocarcinoma. *J Mol Med (Berl)* 2012, 90(4):427-34.
19. Vathipadiekal V, Wang V, Wei W, Waldron L, Drapkin R, Gillette M, Skates S, Birrer M: Creation of a Human Secretome: A Novel Composite Library of Human Secreted Proteins: Validation Using Ovarian Cancer Gene Expression Data and a Virtual Secretome Array. *Clin Cancer Res* 2015, 21(21):4960-9.
20. Chandler RL, Raab JR, Vernon M, Magnuson T, Schisler JC: Global gene expression profiling of a mouse model of ovarian clear cell carcinoma caused by ARID1A and PIK3CA mutations implicates a role for inflammatory cytokine signaling. *Genom Data* 2015, 5:329-32.
21. Tung CS, Mok SC, Tsang YT, Zu Z, Song H, Liu J, Deavers MT, Malpica A, Wolf JK, Lu KH et al: PAX2 expression in low malignant potential ovarian tumors and low-grade ovarian serous carcinomas. *Mod Pathol* 2009, 22(9):1243-50.
22. Stany MP, Vathipadiekal V, Ozbun L, Stone RL, Mok SC, Xue H, Kagami T, Wang Y, McAlpine JN, Bowtell D et al: Identification of novel therapeutic targets in microdissected clear cell ovarian cancers. *PLoS One* 2011, 6(7):e21121.
23. Hill CG, Matyunina LV, Walker D, Benigno BB, McDonald JF: Transcriptional override: a regulatory network model of indirect responses to modulations in microRNA expression. *BMC Syst Biol* 2014,

8:36.

24. 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(9):e46317.
25. Shahab SW, Matyunina LV, Mezencev R, Walker LD, Bowen NJ, Benigno BB, McDonald JF: Evidence for the complexity of microRNA-mediated regulation in ovarian cancer: a systems approach. *PLoS One* 2011, 6(7):e22508.
26. Johnson WE, Li C, Rabinovic A: Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007, 8(1):118-27.
27. Pliarchopoulou K, Pectasides D: Epithelial ovarian cancer: focus on targeted therapy. *Crit Rev Oncol Hematol* 2011, 79(1):17-23.
28. Shi C, Zhang Z: Screening of potentially crucial genes and regulatory factors involved in epithelial ovarian cancer using microarray analysis. *Oncol Lett* 2017, 14(1):725-32.
29. Chatterjee M, Hurley LC, Tainsky MA: Paraneoplastic antigens as biomarkers for early diagnosis of ovarian cancer. *Gynecol Oncol Rep* 2017, 21:37-44.
30. Kay R, Rosten PM, Humphries RK: CD24, a signal transducer modulating B cell activation responses, is a very short peptide with a glycosyl phosphatidylinositol membrane anchor. *J Immunol* 1991, 147(4):1412-6.
31. Zhu J, Zhang G, Lu H: CD24, COX-2, and p53 in epithelial ovarian cancer and its clinical significance. *Front Biosci (Elite Ed)* 2012, 4:2645-51.
32. Nakamura K, Terai Y, Tanabe A, Ono YJ, Hayashi M, Maeda K, Fujiwara S, Ashihara K, Nakamura M, Tanaka Y et al: CD24 expression is a marker for predicting clinical outcome and regulates the epithelial-mesenchymal transition in ovarian cancer via both the Akt and ERK pathways. *Oncol Rep* 2017, 37(6):3189-200.
33. Visvader JE, Lindeman GJ: Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nat Rev Cancer* 2008, 8(10):755-68.
34. Balzar M, Winter MJ, de Boer CJ, Litvinov SV: The biology of the 17-1A antigen (Ep-CAM). *J Mol Med (Berl)* 1999, 77(10):699-712.
35. Ma K, Cao B, Guo M: The detective, prognostic, and predictive value of DNA methylation in human esophageal squamous cell carcinoma. *Clin Epigenetics* 2016, 8:43.
36. Du L, Qian X, Dai C, Wang L, Huang D, Wang S, Shen X: Screening the molecular targets of ovarian cancer based on bioinformatics analysis. *Tumori* 2015, 101(4):384-9.
37. Hellstrom I, Raycraft J, Hayden-Ledbetter M, Ledbetter JA, Schummer M, McIntosh M, Drescher C, Urban N, Hellstrom KE: The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. *Cancer Res* 2003, 63(13):3695-700.
38. Moore RG, Hill EK, Horan T, Yano N, Kim K, MacLaughlan S, Lambert-Messerlian G, Tseng YD, Padbury JF, Miller MC et al: HE4 (WFDC2) gene overexpression promotes ovarian tumor growth. *Sci*

Rep 2014, 4:3574.

39. Yan J, Hedl M, Abraham C: An inflammatory bowel disease-risk variant in INAVA decreases pattern recognition receptor-induced outcomes. *J Clin Invest* 2017, 127(6):2192-205.
40. Garattini E, Fratelli M, Terao M: The mammalian aldehyde oxidase gene family. *Hum Genomics* 2009, 4(2):119-30.
41. Kitamura S, Sugihara K, Ohta S: Drug-metabolizing ability of molybdenum hydroxylases. *Drug Metab Pharmacokinet* 2006, 21(2):83-98.
42. Li W, Middha M, Bicak M, Sjoberg DD, Vertosick E, Dahlin A, Haggstrom C, Hallmans G, Ronn AC, Stattin P et al: Genome-wide Scan Identifies Role for AOX1 in Prostate Cancer Survival. *Eur Urol* 2018, 74(6):710-19.

Supplementary File Legends

Supplementary Figure 1. Four Down-regulated genes fail to be validated by extended samples. Relative gene levels of ovarian cancer and normal tissues measured by qPCR were shown using waterfall plot. (A-D) Represent the levels of *REEP1*, *BNC1*, *HAND2*, *GSDME* separately.

Supplementary Figure 2. Gene knockdown efficiency. (A-B) QPCR analysis for the expression of each gene in OVCAR-3 and CAOV-3 cells upon NC, si-*CD24*, si-*SOX17*, si-*WFDC2*, si-*EPCAM*, si-*INAVA* and si-*AOX1* transfection.

Supplementary Figure 3. Genes that have no effects on proliferation for ovarian cancer cells. Growth curve of OVCAR-3 and CAOV-3 cells upon transfection with NC and si-*CD24*, si-*SOX17*, si-*EPCAM*, si-*INAVA* and si-*AOX1*. Transfected cells were counted and plated in 96-well plates.

Supplementary Figure 4. Genes that have no effects on apoptosis for ovarian cancer cells. NC, si-*CD24*, si-*SOX17*, si-*WFDC2*, si-*EPCAM* and si-*INAVA* were transfected into OVCAR-3 and CAOV-3 cells, cell apoptosis analysis was performed as indicated in methods.

Supplementary Figure 5. Genes that have no effects on migration for ovarian cancer cells. Transwell assay analyzed the migration of OVCAR-3 and CAOV-3 upon NC, si-*CD24*, si-*SOX17*, si-*EPCAM*, si-*INAVA* and si-*AOX1* transfection (magnification, ×100). Scale bar= 100 μm.

Figures

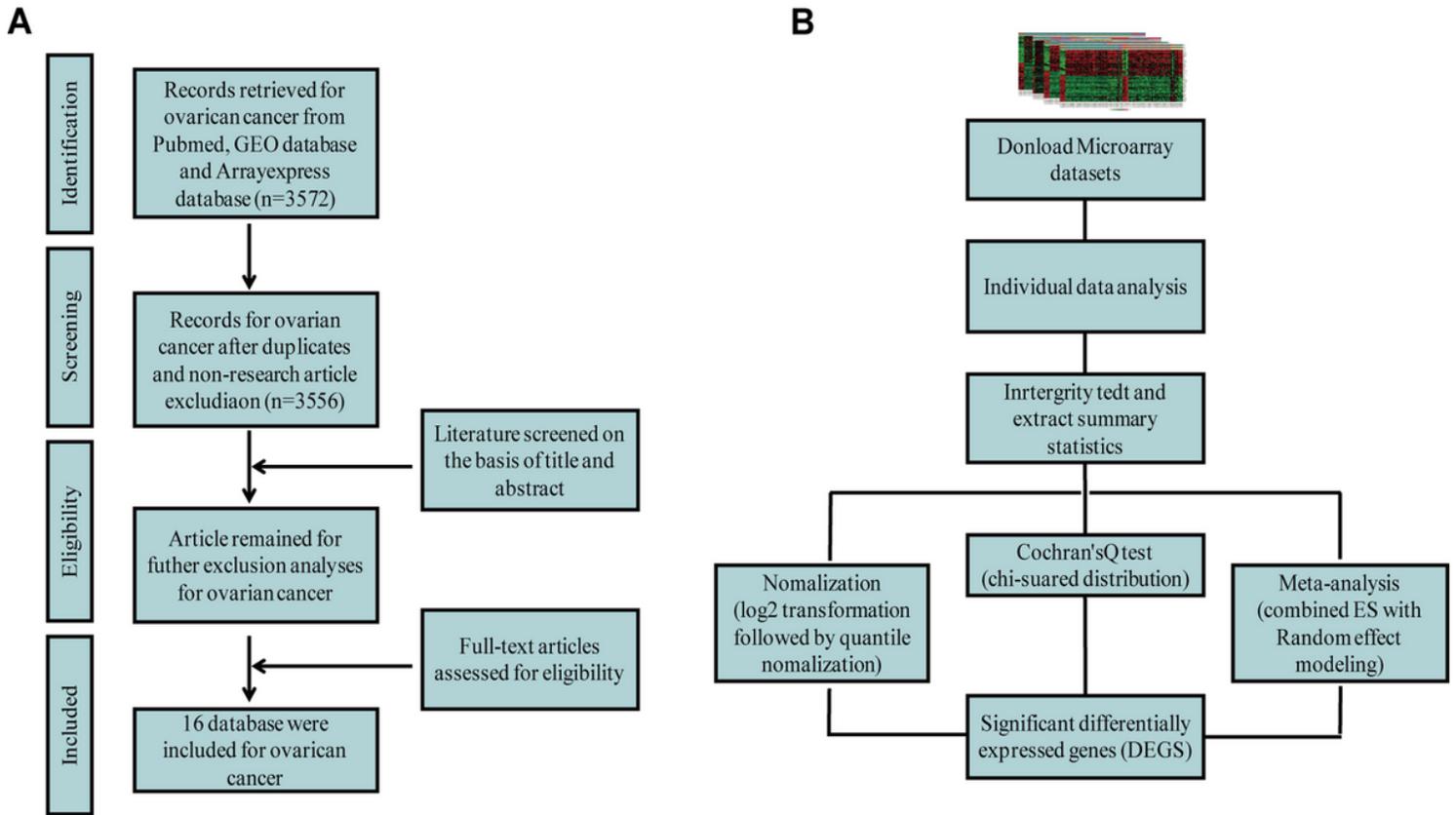


Figure 1

Flowcharts for microarray datasets selection and meta-analysis. (A) Selection process of microarray datasets for meta-analysis of gene expressional signature in ovarian cancer. (B) Process of meta-analysis based data exploration.

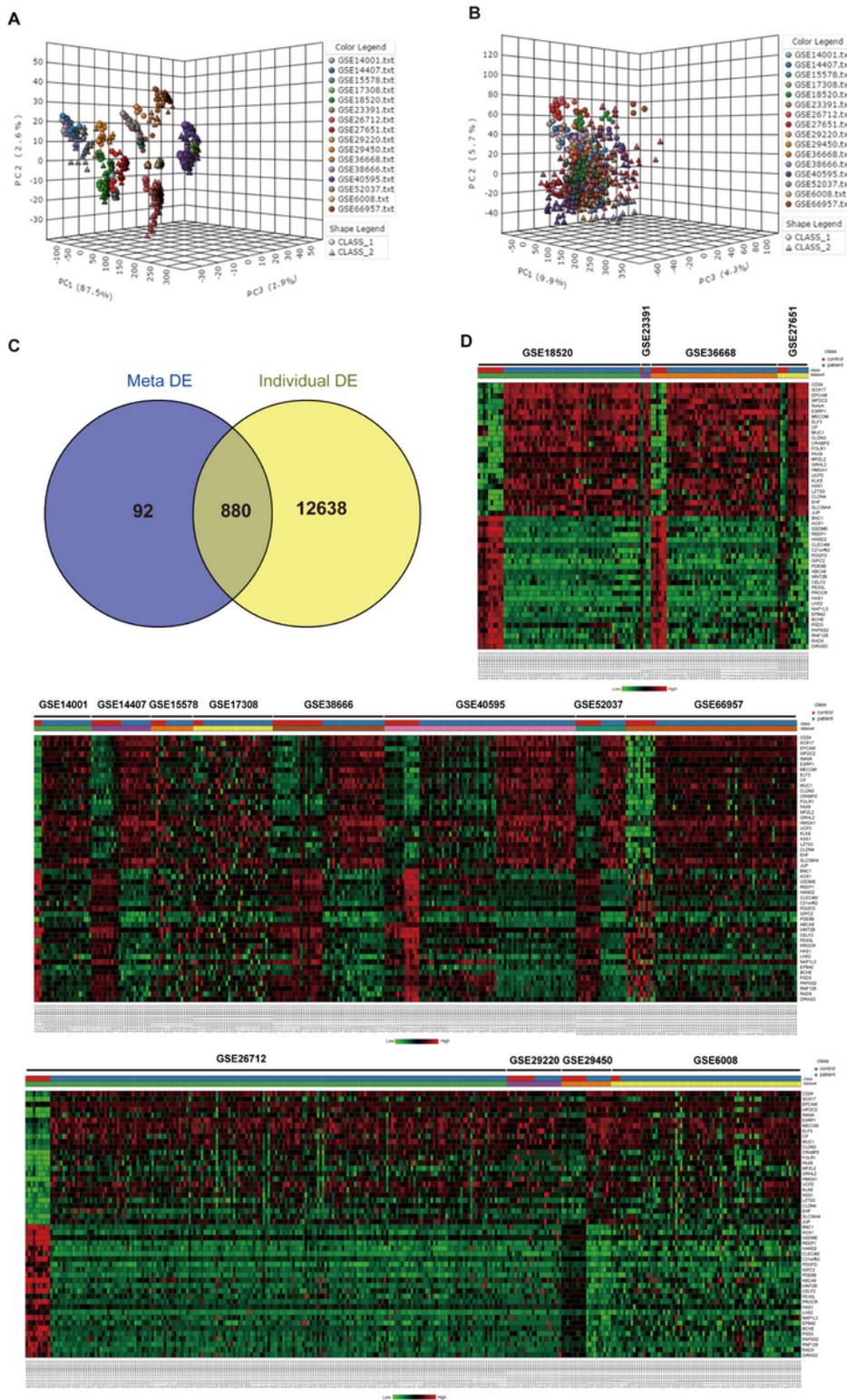


Figure 2

Differentially expressed genes identified by meta-analysis. (A) PCA-3D plot for sample clustering of microarray datasets without batch effect adjustment. (B) PCA-3D plot for sample clustering of microarray datasets after batch effect adjustment. (C) Venn diagram of DEGs by meta-analysis (Meta-DE) and individual microarray dataset analysis (Individual-DE). (D) Heat-map visualization of expressional profiles for top 25 up-and down-regulated DEGs identified by meta-analysis. Genes were ranked by combined ES

value, and the representative heat-map was generated by “Define Custom Signatures” tool of NetworkAnalyst. Class of red: healthy donor; class of green: patient.

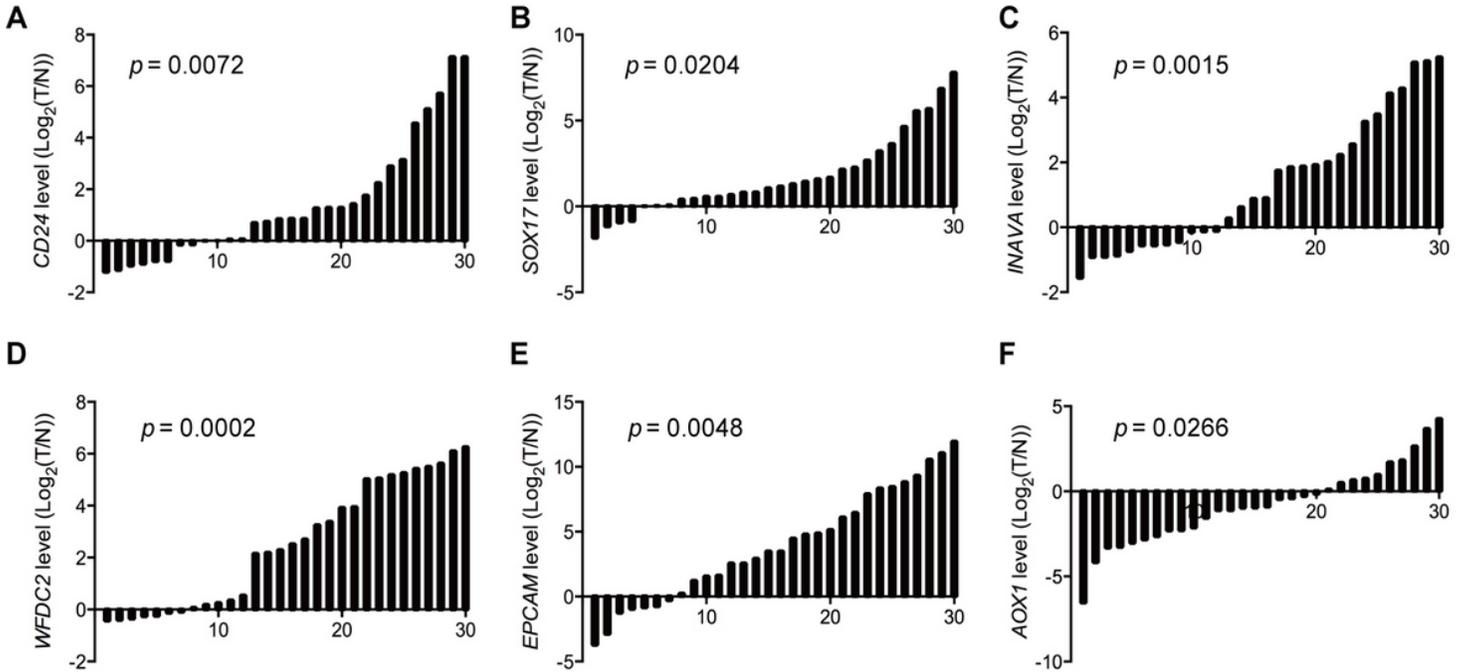


Figure 3

Gene expressional validation for top 5 up and down-regulated genes. Relative genes levels of ovarian cancer and normal tissues measured by qPCR were shown using waterfall plot. (A-F) Represent the expression of CD24, SOX17, WFDC2, EPCAM, INAVA, AOX1 respectively. The fold change of relative genes expression ($\log_2[\text{T/N}] > 1$ or < -1) was defined as significant. T: tumor of ovarian cancer, N: normal control.

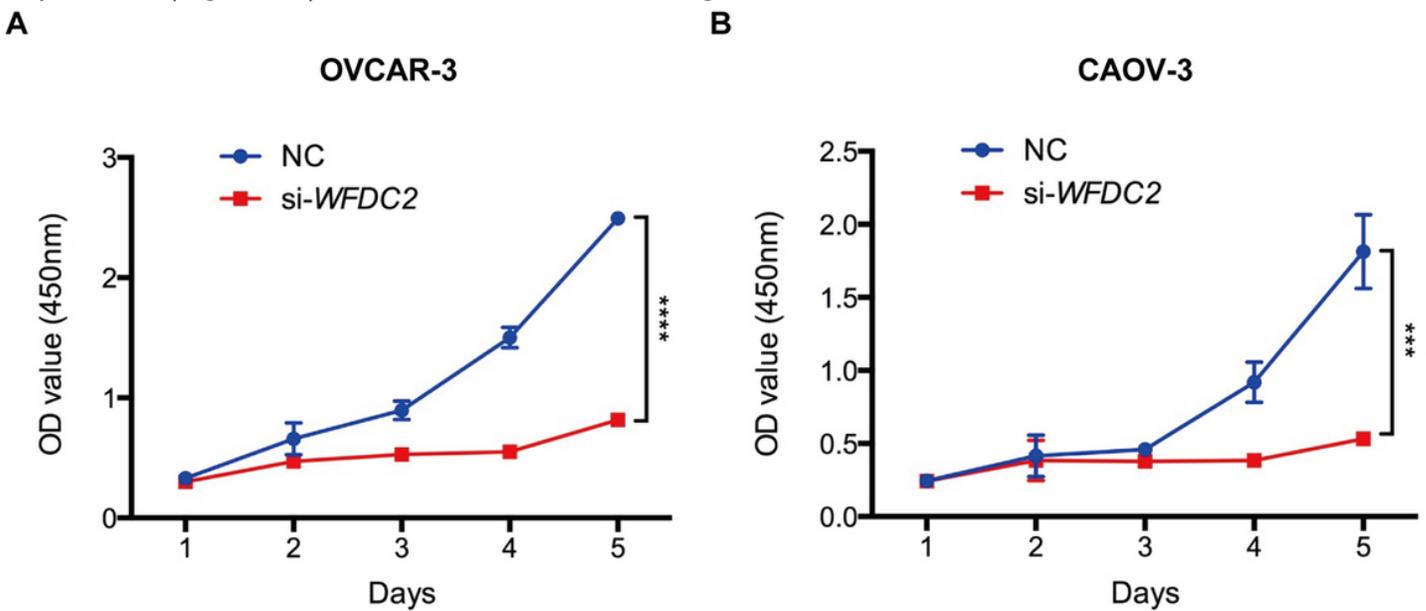


Figure 4

WFDC2 knockdown inhibits proliferation of ovarian cancer cells. (A-B) Growth curve of OVCAR-3 and CAOV-3 cells upon transfection with NC and si-WFDC2. Transfected cells were counted and plated in 96-well plates.

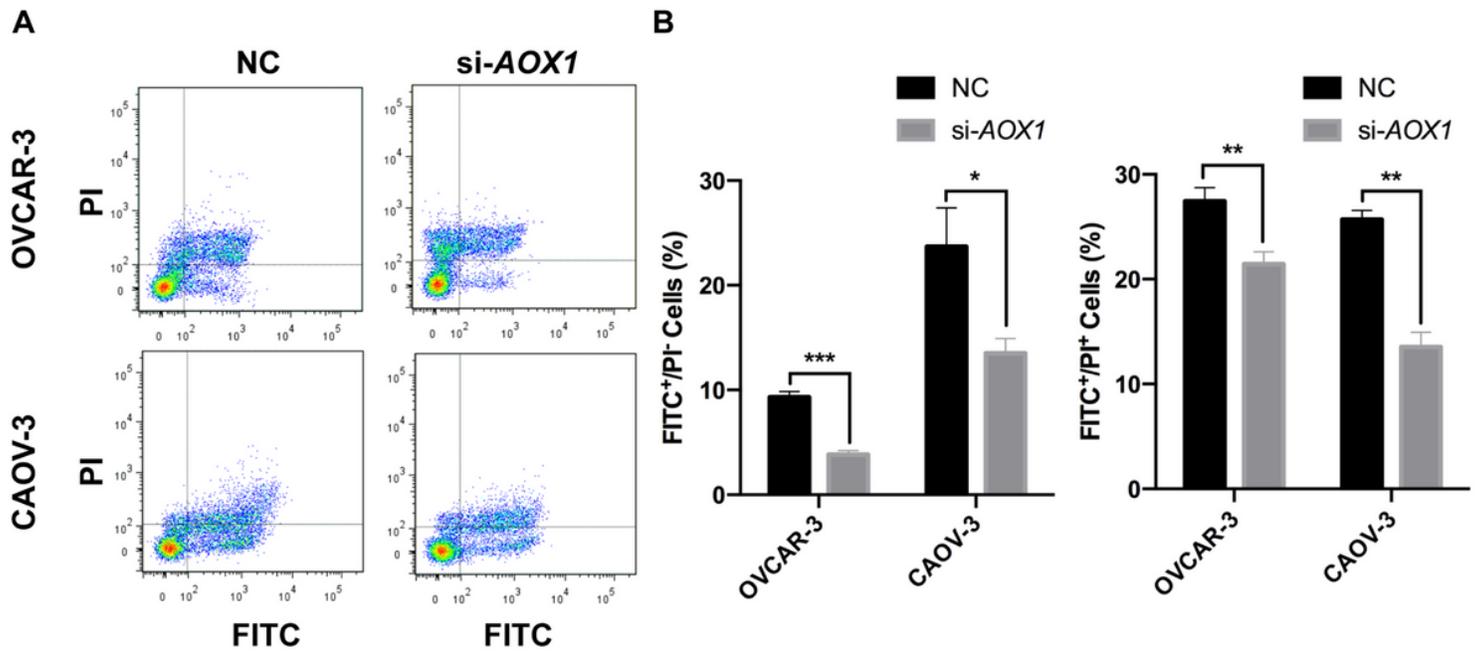


Figure 5

AOX1 knockdown inhibits apoptosis of ovarian cancer cells. Flow cytometry analysis for apoptotic cells affected by si-AOX1 transfection in OVCAR-3 and CAOV-3 cells. Annexin V was labeled by FITC conjugated antibodies, and the DNA was stained by Propidium Iodide. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

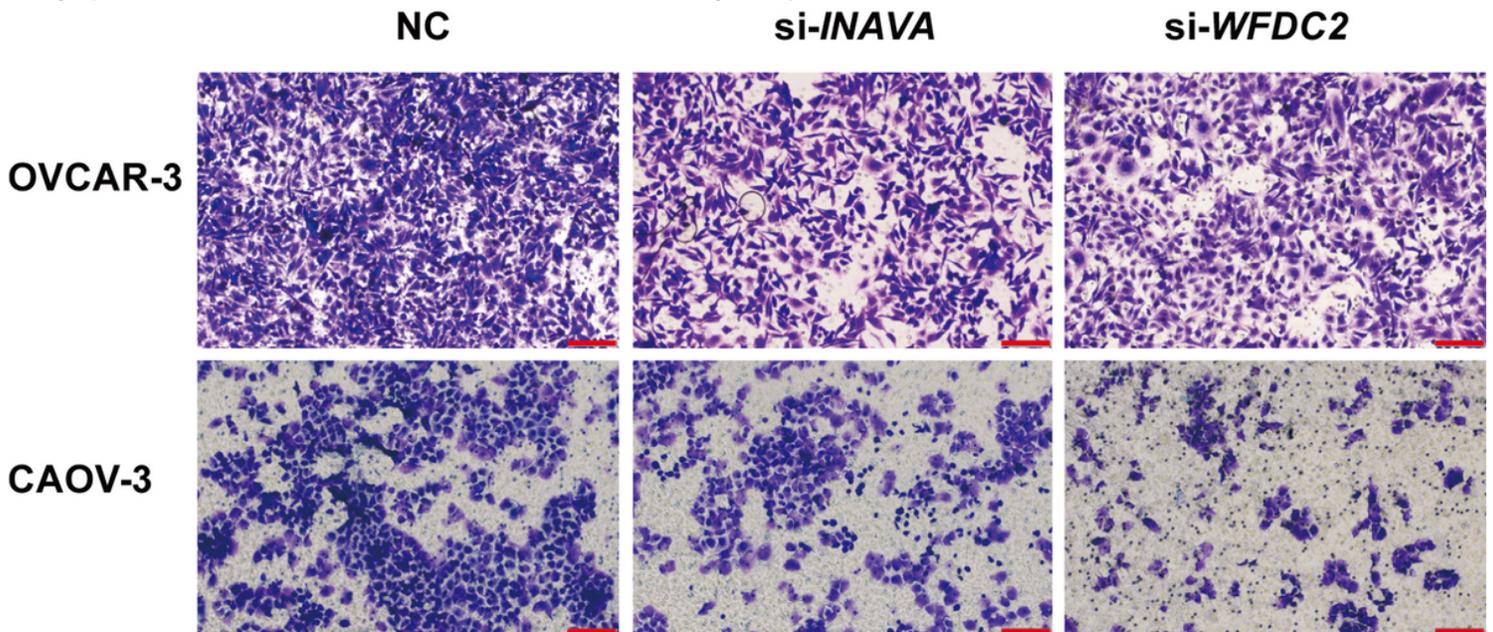


Figure 6

INAVA and WFDC2 knockdown inhibit migration of ovarian cancer cells. Down regulation of INAVA and WFDC2 in OVCAR-3 and CAOV-3 exhibited inhibition of migration (magnification, ×100). Scale bar= 100 μm. The results are representative of three separate experiments.

Supplementary Files

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

- [SupplementaryTable1.doc](#)
- [SupplementaryFigure5.tif](#)
- [SupplementaryFigure2.tif](#)
- [SupplementaryFigure3.tif](#)
- [SupplementaryFigure4.tif](#)
- [SupplementaryFigure1.tif](#)