

Identification and validation of shared genes and key pathways in endometriosis and endometriosis-associated ovarian cancer by weighted gene co-expression network analysis and machine learning algorithms

Yongming Du

Ningbo First Hospital

Yichao Hu

Ningbo First Hospital

Fubin Zhang

Ningbo First Hospital

Tianhong Zhu

Ningbo First Hospital

Yutao Guan (✉ guanyutao@nbdyyy.com)

Ningbo First Hospital

Research Article

Keywords: endometriosis, endometriosis-associated ovarian cancer, WGCNA, machine learning algorithms

Posted Date: February 7th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2542861/v1>

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

[Read Full License](#)

Abstract

Background: Endometriosis is a widespread disease in reproductive age. Epidemiological studies reported that patients with endometriosis had an increased risk of developing endometriosis-associated ovarian cancer (EAOC). The present study aimed to identify shared genes and key pathways that commonly interacted between EAOC and endometriosis.

Methods: The expression matrix of ovarian cancer and endometriosis were collected from the Gene Expression Omnibus database. The weighted gene co-expression network analysis (WGCNA) was used to construct co-expression gene network. Functional enrichment analyses were conducted to clarify the potential regulatory mechanisms. Protein-protein interaction (PPI) network and machine learning algorithms were applied to identify characteristic genes. CIBERSORT deconvolution algorithm was used to explore the difference in tumor immune microenvironment. Receiver operating characteristic curves were utilized to assess the clinical diagnostic ability of hub genes. Furthermore, diagnostic nomogram was constructed and evaluated for supporting clinical practicality.

Results: We identified 262 shared genes between EAOC and endometriosis via WGCNA analysis. They were mainly enriched in cytokine-cytokine receptor interaction, which may be considered a common mechanism between EAOC and endometriosis. After PPI network and machine learning algorithms, we recognized two characteristic genes (EDNRA, OCLN) and established a nomogram that presented an outstanding predictive performance. The hub genes demonstrated remarkable associations with immunological functions. OCLN were highly upregulated in ovarian cancer compared to non-tumor tissues, while expression levels of EDNRA were significantly downregulated in ovarian cancer samples. Survival analysis indicated that dysregulated expressions of EDNRA and OCLN were closely correlated with prognosis of ovarian cancer patients. GSEA analyses revealed that the two characteristic genes were mainly enriched in the cancer- and immune-related pathways. Gene drug interaction analysis found 15 drugs compound that interacted with the hub genes.

Conclusion: We identified two hub genes (EDNRA, OCLN) and constructed a nomogram to predict the risk of EAOC based on WGCNA analyses and machine learning algorithms. They can be used as effective predictive biomarkers for detecting EAOC. Our findings pave the way for further investigation of potential candidate genes and will aid in improving the diagnosis and treatment of EAOC in endometriosis patients.

Introduction

Endometriosis is a disease caused by the implantation of active and growing ectopic endometrium in the ovary, commonly in women of childbearing age, which is benign but with invasive and recurrent malignancy-like behavior [1, 2]. Epidemiological studies reported that patients with endometriosis, especially ovarian endometriosis cysts, had an increased risk of developing epithelial ovarian cancer, significantly higher than the general population, and 0.7%-2.5% of endometriosis was associated with

malignant transformation to endometriosis-associated ovarian cancer (EAOC) [3, 4]. The main clinical features of patients at risk of endometriosis malignancy are: (i) a long history of endometriosis (> 10 years); (ii) having endometriosis at an early age; (iii) a history of endometriosis-related infertility and/or infertility treatment; and (iv) ovarian endometriosis cysts. Sampson originally proposed the criteria of EAOC caused by endometriosis included that the tumor needed to be situated next to unequivocal foci of endometriosis, no other primary tumor present, and the presence of tissues similar to endometrial stroma surrounding epithelial glands [5]. Growing number of evidences demonstrated that atypical endometriosis was a precancer lesion for ovarian cancer [6–8]. EAOC patients have a relatively early age of onset and early clinical stage, with the proportion of stage I patients accounting for approximately 54.2%. Age factors, economic status, depression, pelvic inflammatory disease, and endometriosis-related infertility are high-risk factors for EAOC.

The development of sequencing technologies and bioinformatics has made it possible to investigate the genetic pathogenesis of disease-disease interactions. This study aimed to identify the co-expression genes modules and shared genes of EAOC and endometriosis via weighted gene co-expression network analysis (WGCNA) [9] and machine learning algorithms. WGCNA enables the reconstruction of gene regulatory networks and the identification of modules of genes with similar expression patterns. By integrating gene expression data with other omics data, WGCNA provides a comprehensive view of gene interactions and can help to uncover the underlying biological pathways involved in complex diseases [10, 11]. Machine learning algorithm has been utilized to identify characteristic genes for several diseases [12, 13]. Support vector machine - recursive feature elimination (SVM-RFE) [14] algorithm eliminates superfluous variables and retains only those variables that are relevant to the results. Least absolute shrinkage and selection operator (LASSO) [15] logistic regression is commonly utilized to narrow down variables and identify the most appropriate λ for categorization. Random forest (RF) [16] is used to work with high-dimensional data, construct predictive model and predict the importance of each variable. It is generally accepted that bioinformatics and machine learning algorithms can be effectively employed to identify shared genes and biomarkers for forecasting EAOC.

Our study revealed that EAOC and endometriosis shared 262 genes. They were mainly enriched in cytokine-cytokine receptor interaction, which may be considered a common mechanism between EAOC and endometriosis. After machine learning algorithm in the confirmation cohort, we identified two hub genes (EDNRA, OCLN) and established a nomogram. They can be used as effective predictive biomarkers for detecting occult EAOC in endometriosis patients. Additionally, the dysregulated degree of immune cell infiltration and association of hub genes with immune cells were initially discussed.

Material And Methods

Data sources and preprocessing

The expression profiles of ovarian carcinoma were collected from the Gene Expression Omnibus (GEO) database. In total, four available datasets were collected via R package “GEOquery”. The high throughput

sequencing dataset GSE157153 including 29 ovarian carcinoma tissues and 37 endometriosis tissues was used to explore the shared genes between endometriosis and endometriosis-associated ovarian cancer (EAOC). GSE5108 contained 11 ectopic endometrial samples and 11 eutopic endometrial tissues, GSE18520 consisted of 53 ovarian adenocarcinoma tissues and 10 normal ovary samples. Both of them were utilized to confirm the shared genes. Furthermore, the expression matrix of malignant ovarian samples from The Cancer Genome Atlas (TCGA) dataset and expression profile of normal ovarian tissues from the Genotype-Tissue Expression (GTEx) also served as a validation dataset. The study flowchart is shown in Fig. 1.

WGCNA

We constructed co-expression gene networks of endometriosis and EAOC using WGCNA in order to investigate the underlying biological processes and common genes involved in disease progression. The data were preprocessed by calling the "WGCNA" function package, and the "goodsamplegenes()" function was used to check whether there were abnormal genes in the data and select appropriate thresholds to reject outliers. The clinical feature data were loaded, and the gene-feature clustering tree was drawn to complete the construction of the co-expression network and the selection of modules. The "pickSoftThreshold" algorithm chose an appropriate "soft" threshold (β) to calculate the adjacency matrix and to obtain a biologically significant scale-free network. The clustering dendrogram was drawn for module selection, and the minimum module capacity and module shear height were set to merge modules with high similarity. The EAOC gene set in the modules that were highly correlated with ovarian carcinoma. Scatter plots of gene significance (GS) and module membership (MM) in the significant co-expression modules were plotted to clarify the significance of genes in the selected modules.

Functional enrichment analysis

Gene set enrichment analysis (GSEA) was performed to clarify the potential regulatory mechanisms responsible for the disease progression of endometriosis to EAOC, and the gene set ("c2.cp.kegg.v7.4.symbols.gmt") were obtained from the Broad Institute database. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment were also conducted. Gene Set Variation Analysis (GSVA) was utilized to reveal the differences in signaling pathways between the endometriosis and EAOC by "GSVA" R package. The selected gene set "h.all.v7.4.symbols.gmt" was got from the Molecular Signature Database (MSigDB). All these analyses were performed via R packages "limma", "clusterProfiler", "gseaplot2", and "ggplot2".

Confirmation of shared genes in endometriosis and EAOC

To confirm the shared genes in endometriosis and EAOC, we applied the differentially expressed gene (DEGs) analysis in confirmation dataset GSE157153 including both endometriosis and EAOC tissues with the cutoff value of adjusted $p < 0.05$ and $|\text{Log}_2(\text{fold change})| > 1$ via "limma" package. The overlap of above shared genes from WGCNA analyses of GSE5108 and GSE18520 and DEGs in the confirmation

cohort GSE157153 were presented by a Venn diagram using EVenn (<http://www.ehbio.com/test/venn/#/>) online tool.

Identification of hub genes from protein-protein interaction network

The STRING database was used for protein interaction analysis for a set of shared genes and the results were imported into the Cytoscape software to investigate the target modules and hub genes from the protein-protein interaction (PPI) network. An interaction score with a confidence level of 0.4 was used as the cutoff criterion for construction and unconnected nodes were hidden. Then we utilized four different algorithms (Closeness, Betweenness, MCC, and Degree) via the CytoHubba plugin to identify the top 30 hub shared genes.

Machine learning algorithm

Three machine learning algorithms including SVM-RFE, RF, and LASSO regression analysis were used to investigate the effectiveness of 23 shared DEGs via packages of “msvmRFE.R”, “caret”, and “glmnet”. These algorithms were performed to carry out the binary classification of the data through supervised learning method. Receiver operating characteristic (ROC) curves were calculated based on 23 common genes from the GSE157153 transcriptome dataset, and area under the ROC curves (AUCs) were used to evaluate the diagnostic ability of commonly shared genes from the three classification algorithms. The dataset was randomly divided into training cohort and test cohort in a ratio of 4:1.

Nomogram construction

The common genes obtained as a result of the three classification algorithms were combined to construct the nomogram via “rms” package. The consistence between the nomogram – predicted probability of nonadherence and the actual results were evaluated by calibration curves. We also used decision curve analysis (DCA) to summarize the performance of our shared gene constructed model and to evaluate whether this model has utility in supporting clinical decisions. All these analyses were performed via R packages “VRPM”, “rmda”, “ggDCA”, “rms”, and “ggplot2”.

Expression profile and survival analysis of shared genes between endometriosis and EAOC

We calculated the different expression of shared genes between endometriosis and EAOC in the GSE157153 dataset. Then, these expression profiles were evaluated in the GSE18520 and TCGA-OV database. Boxplot was utilized to present the differential expression of these genes. Furthermore, the overall survival of these shared genes was compared by Kaplan-Meier method according to the package “survminer” and “survival”. An adjusted p-value < 0.05 was considered statistically significant.

Tumor microenvironment cell infiltration analyses

To quantify the relative proportion of immune cells in endometriosis cohort and EAOc cohort, CIBERSORT deconvolution algorithm was used to evaluate the immune infiltration proportion of 22 immune cell subtypes in accordance to the gene expression values. Spearman correlation analysis was used to display the association between immune infiltration proportion of immune cells and the expression of hub genes.

Predicting regulatory targets of transcription factor

The JASPAR TF binding site profile database was used to identify transcription factors (TFs) that bound to shared genes. Additionally, shared genes regulation-related miRNA that negatively correlated with protein expression were predicted by TarBase (v8.0). The topology networks of TFs-gene and miRNA-gene were constructed based on the NetworkAnalyst platform.

Evaluation of clinical diagnostic ability of shared genes

ROC curves and AUCs were utilized to assess whether shared genes could well distinguish EAOc patients from control group via “pROC” package in GSE157153. Additionally, GSE18520 and TCGA-OV datasets were used to assess and explore the shared genes with the most clinical diagnostic ability.

The protein expressions and shared genes between endometriosis and EAOc

The Human Protein Atlas (HPA) database (<https://www.proteinatlas.org/>) mapped all the human proteins in cells, tissues and organs by integrating multiple omics technologies. We searched the representative images of immunohistochemistry showing shared genes expression in endometriosis and EAOc tissues.

Gene–Drug interaction analysis

Although there is promising development of novel therapeutic drugs for ovarian cancer, the prognosis of ovarian cancer is often poor. In the present study, we searched the drug-gene interaction database (DGIdb) for potential drugs that interacted with common genes between endometriosis and EAOc.

Results

Identification of significant modules via WGCNA

We executed WGCNA to identify co-expressed gene modules in endometriosis- and EAOc-related datasets. Based on scale independence of > 0.8 , the best soft threshold (β) was 4 for GSE5108 to ensure biologically significant scalefree topology and we found the same soft threshold for GSE18520 (Fig. 2A, 2B). GSE5108 and GSE18520 co-expression networks were separated into 14 and 6 modules, respectively. Figure 2C and Fig. 2D presented the co-expression modules clustering in endometriosis and ovarian cancer. Each branch represented each gene, and genes clustered into the same module were assigned the same color. As depicted in Fig. 2F, correlation heatmaps of module eigengene values with clinical traits revealed that the Blue and Brown modules were highly correlated with EAOc, while Black

and Grey modules were predominantly associated with endometriosis (Fig. 2E). Therefore, we identified and intersected the hub modules to obtain the shared genes between endometriosis and EAOC patients. Finally, 362 common genes were selected for further investigation.

GO and KEGG pathway enrichment analysis

GO and KEGG pathway enrichment were used based on these 362 common genes between endometriosis and ovarian cancer. GO enrichment analysis consisted of biological process (BP), cell composition (CC) and molecular function (MF) analysis. BP mainly correlated with leukocyte cell-cell adhesion and regulation of leukocyte proliferation. CC was mainly associated with apical part of cell and collagen-containing extracellular matrix. The MF was mainly related to peptidase regulator activity and endopeptidase regulator activity (Fig. 3A). The KEGG pathway showed that these shared genes were mainly enriched in the cytokine-cytokine receptor interaction, intestinal immune network for IgA production, and IL-17 signaling, indicating that these common genes may be involved in tumorigenesis and tumor progression in ovarian cancer (Fig. 3B).

Confirmation of shared genes between endometriosis and EAOC

To confirm shared genes between endometriosis and EAOC, we performed the differentially expressed gene (DEGs) analysis in confirmation dataset GSE157153, which included both endometriosis and EAOC tissues, using the cutoff value of adjusted $p < 0.05$ and $|\text{Log}_2(\text{fold change})| > 1$. We identified 1316 upregulated and 1662 downregulated DEGs from these data (Fig. 3C). Heatmap showed the top 30 up-regulated and down-regulated DEGs in endometriosis and EAOC (Fig. 3D). The shared genes in endometriosis and EAOC were then intersected with these DEGs. Finally, we obtained a total of 116 common genes, which may play crucial role in both endometriosis and EAOC (Fig. 3E).

Protein-protein interaction network construction

A PPI network of potential targets was constructed via the STRING online database, and a total of 116 common genes were visualized by Cytoscape software. Using the CytoHubba plugin and four distinct algorithms (Closeness, Betweenness, MCC, and Degree), we identified the top 30 shared genes (Fig. 3F). Finally, 23 node genes were selected from this network for subsequent analysis.

Machine learning algorithms for hub genes selection

Three classification algorithms were applied in this study to further investigate shared genes between endometriosis and EAOC for potential biomarkers from 23 potential genes. Twelve genes were found to have the lowest binomial deviation using the LASSO regression algorithm (Fig. 4A, 4B). SVM-RFE analysis revealed that the model incorporating four genes owing the lowest error (Fig. 4D). Based on the importance score provided by the RF algorithm, the top ten genes were chosen and ordered. (Fig. 4C, 4E). Finally, two hub genes (EDNRA and OCLN) were selected for the nomogram construction and diagnostic ability evaluation. A Venn diagram was used to display the intersected genes of three machine learning algorithms (Fig. 4F).

Nomogram construction and assessment of diagnostic ability of hub genes

The diagnostic nomogram was constructed and presented that the expression of two characteristic genes contributed to the clinical diagnosis of EAOC. By integrating the scores corresponding to the relative expression of each gene, the probability of a patient being diagnosed with EAOC was calculated (Fig. 5A). Calibration plots demonstrated the diagnostic nomogram's strong prognostic value (Fig. 5B). To evaluate the clinical practicality of our characteristic genes, DCA was performed, which integrated the preferences of the patients or clinicians into consideration (Fig. 5C). In this analysis, clinical judgment of the benefits and harms related to diagnostic model will be made. We compared EDNRA, OCLN and nomogram and found that nomogram demonstrated the best overall net benefit across a wide and practical range of threshold probabilities. This study may help clinicians make more accurate diagnoses of patients with EAOC.

ROC curves and AUCs were utilized in GSE157153 to evaluate the diagnostic performance of EDNRA and OCLN. The results showed each gene to have exceptional predictive performance as shown below.: OCLN (AUC: 0.991, 95%CI: 0.976–1.000, Fig. 5D); EDNRA (AUC: 0.922, 95%CI: 0.850–0.993, Fig. 5E); Nomogram (AUC: 0.993, 95%CI: 0.982–1.000, Fig. 5F). The result showed that the nomogram consisting of EDNRA and OCLN presenting the best predictive performance and acting a potential diagnostic biomarker for EAOC. GSE18520 was utilized to validate the diagnostic value and the results showed that the AUCs of OCLN, EDNRA, and Nomogram were 0.789, 0.860, and 0.900 respectively via ROC curve analyses (Fig. 5J-L). Furthermore, TCGA-OV was also used to evaluate the predictive performance of EDNRA and OCLN. The ROC analyses yielded the AUCs of 0.995, 0.867, and 0.996 for OCLN, EDNRA, and Nomogram respectively, demonstrating good clinical diagnostic ability (Fig. 5G-I).

Immune infiltration analysis

To compare the difference in immune infiltration landscape between endometriosis and EAOC patients, we evaluated the infiltration levels of the immune cells, immune checkpoint molecular. CIBERSORT deconvolution algorithm demonstrated that the immune infiltration levels of plasma cells, T cells follicular helper, Macrophages M1, and Mast cells activated were significantly elevated while CD8⁺ T cells, NK cells activated, macrophages M2, and mast cells resting were remarkably decreased in EAOC group (Fig. 6A). Correlation analysis showed that T cells CD4 memory resting had the greatest positive correlation with CD8⁺ T cells, while B cells memory displayed the highest negative correlation with T cells regulatory (Tregs) (Fig. 6D). Furthermore, spearman correlation analyses were conducted to investigate the expression values of hub genes with immune cell infiltration degree and the results showed that OCLN was positively correlated with Tfh, macrophages M1, and mast cells activated, while negatively correlated with CD8⁺ T cells, mast cells resting, and macrophages M2 (Fig. 6B). EDNRA was positively correlated with mast cells resting, NK cells activated, and CD8⁺ T cells, and negatively correlated with plasma cells, macrophages M1, T cells follicular helper (Tfh), and mast cells activated (Fig. 6C).

GSVA of hallmark gene sets between EAOc and endometriosis and GSEA functional analysis of hub genes

GSVA of hallmark gene sets between EAOc and endometriosis patients was performed in the GSE157153 dataset (Fig. 7C). The results showed that the MYC_TARGETS_V1, PI3K_AKT_MTOR_SIGNALING, and P53_PATHWAY were activated, while NOTCH_SIGNALING, and TGF_BETA_SIGNALING were suppressed in the progression from endometriosis to EAOc. Furthermore, GSEA analyses were utilized to evaluate signaling pathways involved in the hub genes. We found that EDNRA and OCLN were mainly enriched in the tumor-related pathways, including P53 signaling pathway, calcium signaling pathway, cell cycle, focal adhesion, and MAPK signaling pathway (Fig. 7A, 7B). While OCLN were also associated with enrichment in immune-related pathways, like systemic lupus erythematosus, cytokine–cytokine receptor interaction, chemokine signaling pathway, and autoimmune thyroid disease (Fig. 7B).

Identification of TFs and miRNAs interacting with shared genes

To investigate the regulatory mechanisms of shared genes OCLN and EDNRA in endometriosis and EAOc, we used NetworkAnalyst tool to explore the TFs and miRNAs. In the TF-hub genes network, we found that OCLN may be regulated by HNF4A, JUND, and SRF. EDNRA may be regulated by STAT3, FOXL1, and CREB1 (Additional file 1A). Additionally, miRNA-hub genes network demonstrated that has-mir-129-2-3p may play core regulatory role. Through network analysis, the TFs and miRNAs connected to two characteristic genes were shown in Additional file 1B.

The expression profiles and protein expression of shared genes

We compared the expression profiles of the hub genes EDNRA and OCLN in patients with endometriosis and EAOc, the results revealed that the expression values of EDNRA were profoundly downregulated while the expression profiles of OCLN were significantly upregulated in ovarian cancer (Fig. 8A, 8B). The above findings were also evaluated and validated in the GSE18520 (Fig. 8C, 8D) and TCGA-OV cohorts (Fig. 8E, 8F). To further investigate the protein expression of EDNRA and OCLN, we searched the representative images of immunohistochemistry stainings from HPA and found that OCLN proteins expression levels were significantly elevated in ovarian carcinoma tissues when compared to normal ovarian tissues, while EDNRA were remarkably overexpressed in normal ovarian samples when compared to ovarian carcinoma samples (Fig. 8I), which were consistent with the expression patterns of mRNA. To evaluate the correlation of expression profiles of hub genes with prognosis of EAOc, Kaplan-Meier curves were plotted to compare the overall survival difference between high- and low-expression groups. The results demonstrated that EAOc patients in TCGA-OV dataset with upregulated expression of OCLN and EDNRA showed significant worse prognosis than those with downregulated expression of EDNRA and OCLN (Fig. 8G, 8H).

Prediction of potential therapeutic drugs

Drug-gene interaction analysis was performed to investigate the potential drugs that interact with two shared genes between endometriosis and EAOC via the DGldb database. We constructed drug-corresponding hub genes networks and found 15 potential drugs, including sitaxentan, ambrisentan, and bosentan. These potential drugs derived from hub genes may play a crucial role in elucidating the mechanism of genes and cell function in relation to receptor sensitivity (Additional file 2).

Discussion

Endometriosis is a widespread disease in reproductive age. It causes chronic inflammatory reaction that may result in the formation of scar tissue within the pelvis. The expression of cytokines, which were responsible for the inflammatory response and its regulation, was shown to be affected by the dysfunction of the immune system. The activation of immune cells triggered signaling pathways, caused the release of inflammatory cytokines and accumulated multiple cell types at the site of inflammation [17–19]. In the endometriotic environment, higher levels of IL-17, which modulates immune cells and perpetuates chronic inflammation, have been detected [20–22]. Recent research demonstrated that the immune system actively contributed to the pathogenesis of ovarian cancer, and that local and systemic immune responses determined tumor growth and clinical outcomes. Immunotherapy for gynecological cancers may assist in reversing immunosuppression and lymphocyte depletion brought on by previous treatments. Interestingly, the KEGG pathway showed that these shared genes were mainly correlated with the cytokine-cytokine receptor interaction, intestinal immune network for IgA production, and IL-17 signaling pathway. The GO enrichment analysis demonstrated the biological functions of these common genes were enriched in regulation of immune cells proliferation. These results indicated that the immune system played a crucial role in the progression of endometriosis and EAOC.

WGCNA is a systematic biological method that uses a topological overlap measure to construct a network of genes based on their expression profiles. WGCNA has been widely used to identify gene-gene interactions, study gene expression patterns, and identify gene regulatory networks. It has also been applied to a variety of biological problems, from cancer to cardiovascular disease. Herein, there were 14 endometriosis-related and 6 EAOC-related gene modules via WGCNA. The blue and brown modules exhibited the highest correlation and were deemed EAOC-related modules, whereas the black and grey modules were applied as endometriosis-related key modules. By intersecting the shared genes of endometriosis-related and EAOC-related gene clusters, a total of 362 genes were obtained. Limma analysis was utilized to identify DEGs in the confirmation cohort consisting both endometriosis and EAOC patients. Then, we selected 23 DEGs after PPI network construction and analysis. After intersection of three machine learning methods (SVM-RFE, LASSO, and RF), two hub DEGs (EDNRA and OCLN) were chosen for nomogram construction and diagnostic ability evaluation for endometriosis patients with EAOC. First, we identified two diagnostic biomarkers (EDNRA and OCLN) by integrating the scores corresponding to the expression of hub genes. Then, we performed external validation using two additional datasets, which revealed that the two hub genes had exceptional predictive performance for EAOC in endometriosis patients, indicating their potential diagnostic roles during the tumorigenesis and progression of EAOC.

Endothelin (EDN) was initially isolated from cultured porcine aortic endothelial cells [23, 24]. EDN receptors, namely endothelin A receptor (EDNRA) and endothelin B receptor (EDNRB), which belonged to the G protein-coupled receptor family [25], mediated the biological actions of EDN. EDNRA was expressed in both vascular and non-vascular tissues and regulated a wide range of physiological processes [26, 27]. A study on chemotherapeutic resistance of ovarian cancer found that the activation of EDNRA by EDN1 promoted a direct interaction between β -arr1 and β -catenin to regulate epigenetic modifications driving cancer chemoresistance onset. It also reported that the overexpression of EDNRA correlated with poor prognosis, indicating EDNRA as a potential predictive marker of chemoresistance [28]. The survival rate of EAOc patients with a high expression of EDNRA was also found to be lower in the present study.

Tight junctions are protein structures that regulate the transcellular transport of water, ions, and macromolecules. OCLN, the first tight junctions protein identified, is considered a necessary integral protein for the structure and function of tight junctions [29]. Pakuła et al. demonstrated that the expression of junctional proteins, including OCLN, E-cadherin, and desmoglein, in peritoneal mesothelial cells determined the integrity of the peritoneal mesothelium, which played critical role for the invasion of ovarian cancer [30]. Tobioka et al. reported that in human endometrial carcinoma, OCLN expression decreased proportionally with tumor grade and correlated with myometrial invasion and lymph node metastases [31]. Forced expression of OCLN in human cervical carcinoma cells enhanced the sensitivity of the cells to differently triggered apoptotic stimuli and thus inhibited the tumorigenic potencies of the transformed cells and suppressed tumor development [32]. Martin et al. found that knockdown of OCLN in human breast cancer cells lead to the profoundly increased invasive potential cells. In addition, the absence of OCLN expression was linked to metastatic disease. And they concluded that OCLN was necessary for the proper function of tight junctions in human breast epithelial cells [33].

Local and systemic immune responses determined tumor growth and clinical outcomes and had a significant impact on the pathogenesis of ovarian cancer. As a result, we compared the proportion of immune cell infiltrating in EAOc patients and found that the immune infiltration levels of plasma cells, Tfh, macrophages M1, and mast cells activated were significantly elevated, while CD8⁺ T cells, NK cells, macrophages M2, and mast cells resting were remarkably decreased. Previous research demonstrated that CD8⁺ T cells were essential for tumor suppression and linked to better prognoses in a variety of cancers [34]. NK cells are granular lymphocytes that play critical roles in the immune surveillance of tumors. In patients with cancer, the degree of NK cell infiltration in tumor tissues was prognostic, and a reduced NK cell function was associated with worse outcome [35, 36]. The recent advancements in ovarian cancer research showed that macrophage could enhance ovarian tumor progression, but due to its high plasticity, it can also inhibit tumor growth by remodeling the M2-like phenotype to M1-like phenotype, and the plasticity of macrophage can be exploited in ovarian cancer treatment [37, 38]. Tfh are a subset of CD4⁺ T cells identified for the first time in the human tonsil. They play an essential role in protective immunity by assisting B cells in the production of antibodies against foreign pathogens. Recent studies find that mast cells have both tumor-promoting and tumor-inhibiting effects, and compartmentalized by the microenvironment responsible for this effect [39]. Chan et al. demonstrated

that high degree of mast cell infiltration was correlated with better prognosis than low degree of infiltration in women with advanced ovarian cancer [40]. Moreover, our study revealed that there was positive correlation between EDNRA expression level and activated CD8⁺ T cells and NK cells, indicating that EDNRA was associated with enhancing the antitumor immune response. Combined with the downregulated expression levels of EDNRA in EAOC patients, our study indicated that decreased EDNRA levels were correlated with a suppressive immunophenotype in patients with EAOC. OCLN expression level showed positive correlations with immune infiltration degree of Tfh, macrophages M1, and mast cells activated, Meanwhile, it was negatively correlated macrophages M2. Similar to EDNRA, the above results also suggest that OCLN is associated with an activated immunophenotype. It is partially confirmed that the increased OCLN expression level may be related to the enhanced activation of the immune system when considering the expression pattern of OCLN in EAOC patients. However, the functions of these hub genes in EAOC need further investigate by vitro and vivo experiments.

There were some limitations in this study. Firstly, the diagnostic biomarkers were constructed only on the basis of public databases without real-world data validation. Therefore, further verification with improvement in sample size, sequencing data, and clinical information is needed in the future. Secondly, due to the distinct experimental designs for each dataset, the combination of various gene expression datasets may be biased.

Conclusion

We identified two shared genes (EDNRA, OCLN) and developed a diagnostic nomogram to predict the risk of EAOC patients with endometriosis via WGCNA analysis and machine learning algorithms. Cytokine-cytokine receptor interaction was found to be a common signaling pathway between EAOC and endometriosis. The correlation of hub genes with immune cells as well as the dysregulated level of immune cell infiltration were initially discussed. Collectively, the results of our research lay the groundwork for further exploration of possible candidate genes and aid in the improvement of EAOC diagnosis and treatment. However, more research is required to determine the exact mechanisms by which these two hub genes affecting the development and progression of EAOC from endometriosis.

Abbreviations

AUCs: Area under the ROC curves

EAOC: Endometriosis-associated ovarian cancer

GSEA: Gene set enrichment analysis

GO: Gene Ontology

GSVA: Gene set variation analysis

GEO, Gene Expression Omnibus

GTEX: Genotype-Tissue Expression

KEGG: Kyoto Encyclopedia of Genes and Genomes

Limma: Linear models for microarray data

LASSO: Least absolute shrinkage and selection operator

RF: Random forest

SVM-RFE: Support vector machine-recursive feature elimination

TCGA: The Cancer Genome Atlas

WGCNA: Weighted gene co-expression network analysis

TFs: Transcription factors

ROC: Receiver operating characteristic curve

Declarations

Ethics approval

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data during the present study is available from the corresponding authors on reasonable request.

Funding

This work was supported by grants from Medical Health Science and Technology Project of Zhejiang Provincial Health Commission (2022KY303).

Author contributions

YM Du, YC Hu and FB Zhang conducted the data analyses; YM Du and TH Zhu drafted the manuscript; YM Du, YC Hu, and YT Guan contributed to manuscript review. All the authors approved the final manuscript.

Acknowledgments

None.

Conflict of Interest

The authors have no conflict of interest to disclose.

References

1. Taylor HS, Osteen KG, Bruner-Tran KL, Lockwood CJ, Krikun G, Sokalska A, Duleba AJ. Novel therapies targeting endometriosis. *Reprod Sci.* 2011;18:814-823.
2. Burney RO, Giudice LC. Pathogenesis and pathophysiology of endometriosis. *Fertil Steril.* 2012;98:511-519.
3. Pearce CL, Templeman C, Rossing MA, Lee A, Near AM, Webb PM et al. Berchuck A. Association between endometriosis and risk of histological subtypes of ovarian cancer: a pooled analysis of case-control studies. *Lancet Oncol.* 2012;13:385-394.
4. Hermens M, van Altena AM, Nieboer TE, Schoot BC, van Vliet H, Siebers AG, Bekkers RLM. Incidence of endometrioid and clear-cell ovarian cancer in histological proven endometriosis: the ENOCA population-based cohort study. *Am J Obstet Gynecol.* 2020;223:107 e101-107 e111.
5. SAMPSON JA. ENDOMETRIAL CARCINOMA OF THE OVARY, ARISING IN ENDOMETRIAL TISSUE IN THAT ORGAN. *Archives of Surgery.* 1925;10:1-72.
6. Wiegand KC, Shah SP, Al-Agha OM, Zhao Y, Tse K, Zeng Tet et al. Huntsman DG. ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med.* 2010;363:1532-1543.
7. Fukunaga M, Nomura K, Ishikawa E, Ushigome S. Ovarian atypical endometriosis: its close association with malignant epithelial tumours. *Histopathology.* 1997;30:249-255.
8. Ogawa S, Kaku T, Amada S, Kobayashi H, Hirakawa T, Ariyoshi Ket et al. Nakano H. Ovarian endometriosis associated with ovarian carcinoma: a clinicopathological and immunohistochemical study. *Gynecol Oncol.* 2000;77:298-304.
9. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
10. Guo X, Xiao H, Guo S, Dong L, Chen J. Identification of breast cancer mechanism based on weighted gene coexpression network analysis. *Cancer Gene Ther.* 2017;24:333-341.
11. Feng S, Xu Y, Dai Z, Yin H, Zhang K, Shen Y. Integrative Analysis From Multicenter Studies Identifies a WGCNA-Derived Cancer-Associated Fibroblast Signature for Ovarian Cancer. *Front Immunol.* 2022;13:951582.
12. Yu R, Zhang J, Zhuo Y, Hong X, Ye J, Tang S, Zhang Y. Identification of Diagnostic Signatures and Immune Cell Infiltration Characteristics in Rheumatoid Arthritis by Integrating Bioinformatic Analysis and Machine-Learning Strategies. *Front Immunol.* 2021;12:724934.

13. Zheng PF, Chen LZ, Liu P, Pan HW, Fan WJ, Liu ZY. Identification of immune-related key genes in the peripheral blood of ischaemic stroke patients using a weighted gene coexpression network analysis and machine learning. *J Transl Med.* 2022;20:361.
14. Sanz H, Valim C, Vegas E, Oller JM, Reverter F. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinformatics.* 2018;19:432.
15. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 2011;73:273-282.
16. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. *Brief Bioinform.* 2023;
17. Vallvé-Juanico J, Houshdaran S, Giudice LC. The endometrial immune environment of women with endometriosis. *Hum Reprod Update.* 2019;25:564-591.
18. Zhang T, De Carolis C, Man GCW, Wang CC. The link between immunity, autoimmunity and endometriosis: a literature update. *Autoimmun Rev.* 2018;17:945-955.
19. Harada T, Iwabe T, Terakawa N. Role of cytokines in endometriosis. *Fertil Steril.* 2001;76:1-10.
20. Hirata T, Osuga Y, Takamura M, Saito A, Hasegawa A, Koga K et al. Taketani Y. Interleukin-17F increases the secretion of interleukin-8 and the expression of cyclooxygenase 2 in endometriosis. *Fertil Steril.* 2011;96:113-117.
21. Ahn SH, Edwards AK, Singh SS, Young SL, Lessey BA, Tayade C. IL-17A Contributes to the Pathogenesis of Endometriosis by Triggering Proinflammatory Cytokines and Angiogenic Growth Factors. *J Immunol.* 2015;195:2591-2600.
22. Shi FD, Ljunggren HG, La Cava A, Van Kaer L. Organ-specific features of natural killer cells. *Nat Rev Immunol.* 2011;11:658-671.
23. Inoue A, Yanagisawa M, Kimura S, Kasuya Y, Miyauchi T, Goto K, Masaki T. The human endothelin family: three structurally and pharmacologically distinct isopeptides predicted by three separate genes. *Proc Natl Acad Sci U S A.* 1989;86:2863-2867.
24. Yanagisawa M, Kurihara H, Kimura S, Tomobe Y, Kobayashi M, Mitsui Y et al. Masaki T. A novel potent vasoconstrictor peptide produced by vascular endothelial cells. *Nature.* 1988;332:411-415.
25. Davenport AP, Hyndman KA, Dhaun N, Southan C, Kohan DE, Pollock J et al. Maguire JJ. Endothelin. *Pharmacol Rev.* 2016;68:357-418.
26. Bondurand N, Dufour S, Pingault V. News from the endothelin-3/EDNRB signaling pathway: Role during enteric nervous system development and involvement in neural crest-associated disorders. *Dev Biol.* 2018;444 Suppl 1:S156-S169.
27. Kohan DE, Rossi NF, Inscho EW, Pollock DM. Regulation of blood pressure and salt homeostasis by endothelin. *Physiol Rev.* 2011;91:1-77.
28. Rosanò L, Cianfrocca R, Tocci P, Spinella F, Di Castro V, Caprara V et al. Bagnato A. Endothelin A receptor/ β -arrestin signaling to the Wnt pathway renders ovarian cancer cells resistant to chemotherapy. *Cancer Res.* 2014;74:7453-7464.

29. Furuse M, Hirase T, Itoh M, Nagafuchi A, Yonemura S, Tsukita S, Tsukita S. Occludin: a novel integral membrane protein localizing at tight junctions. *J Cell Biol.* 1993;123:1777-1788.
30. Pakuła M, Witucka A, Uruski P, Radziemski A, Moszyński R, Szpurek Det al.Książek K. Senescence-related deterioration of intercellular junctions in the peritoneal mesothelium promotes the transmesothelial invasion of ovarian cancer cells. *Sci Rep.* 2019;9:7587.
31. Tobioka H, Isomura H, Kokai Y, Tokunaga Y, Yamaguchi J, Sawada N. Occludin expression decreases with the progression of human endometrial carcinoma. *Hum Pathol.* 2004;35:159-164.
32. Osanai M, Murata M, Nishikiori N, Chiba H, Kojima T, Sawada N. Epigenetic silencing of occludin promotes tumorigenic and metastatic properties of cancer cells via modulations of unique sets of apoptosis-associated genes. *Cancer Res.* 2006;66:9125-9133.
33. Martin TA, Mansel RE, Jiang WG. Loss of occludin leads to the progression of human breast cancer. *Int J Mol Med.* 2010;26:723-734.
34. St Paul M, Ohashi PS. The Roles of CD8(+) T Cell Subsets in Antitumor Immunity. *Trends Cell Biol.* 2020;30:695-704.
35. Albertsson PA, Basse PH, Hokland M, Goldfarb RH, Nagelkerke JF, Nannmark U, Kuppen PJ. NK cells and the tumour microenvironment: implications for NK-cell function and anti-tumour activity. *Trends Immunol.* 2003;24:603-609.
36. Morvan MG, Lanier LL. NK cells and cancer: you can teach innate cells new tricks. *Nat Rev Cancer.* 2016;16:7-19.
37. Zhang Z, Huang L, Brayboy L. Macrophages: an indispensable piece of ovarian health. *Biol Reprod.* 2021;104:527-538.
38. Puthenveetil A, Dubey S. Metabolic reprogramming of tumor-associated macrophages. *Ann Transl Med.* 2020;8:1030.
39. Ch'ng S, Wallis RA, Yuan L, Davis PF, Tan ST. Mast cells and cutaneous malignancies. *Mod Pathol.* 2006;19:149-159.
40. Chan JK, Magistris A, Loizzi V, Lin F, Rutgers J, Osann Ket al.Samoszuk M. Mast cell density, angiogenesis, blood clotting, and prognosis in women with advanced ovarian cancer. *Gynecol Oncol.* 2005;99:20-25.

Figures

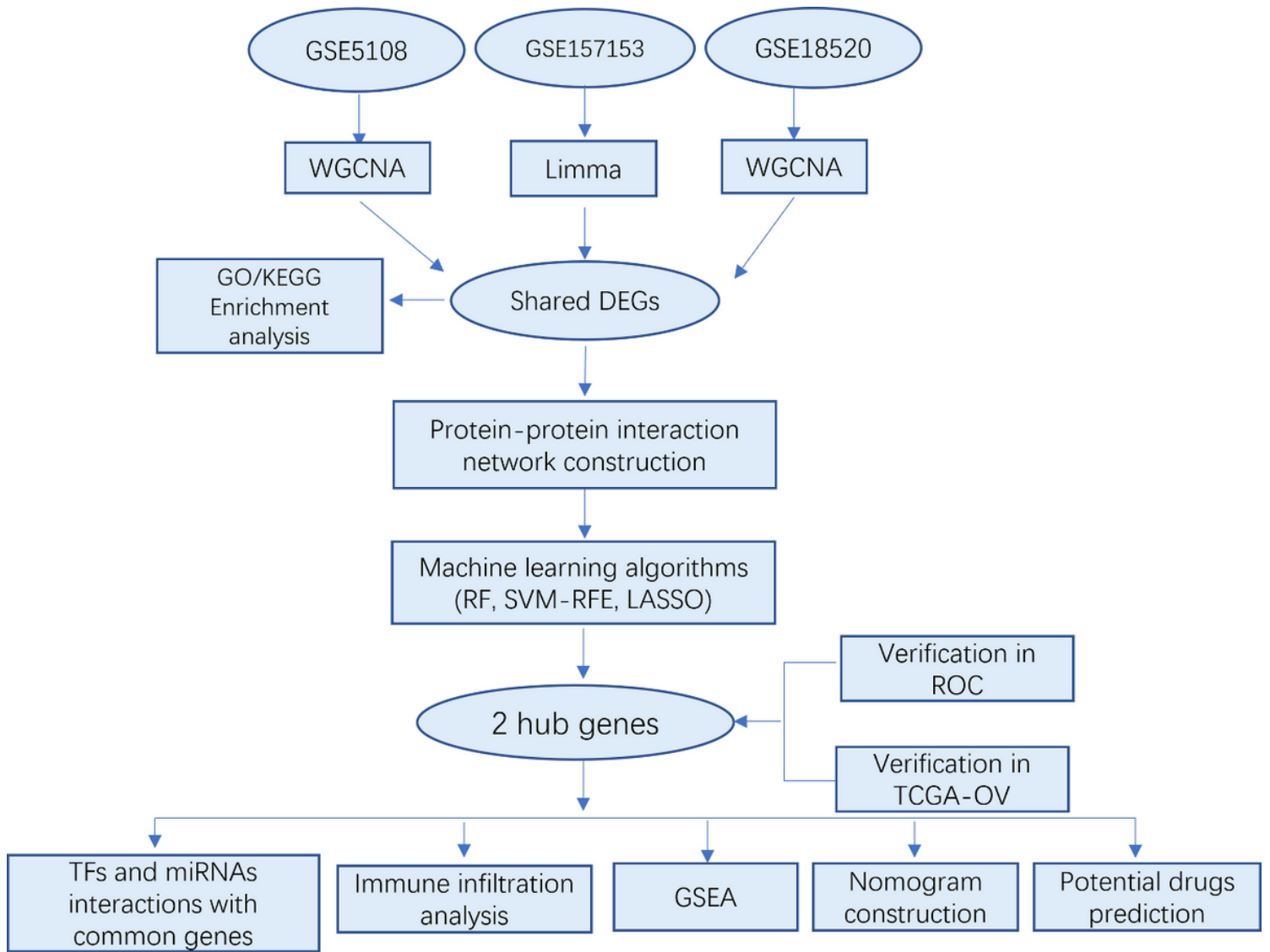


Figure 1

Study flowchart.

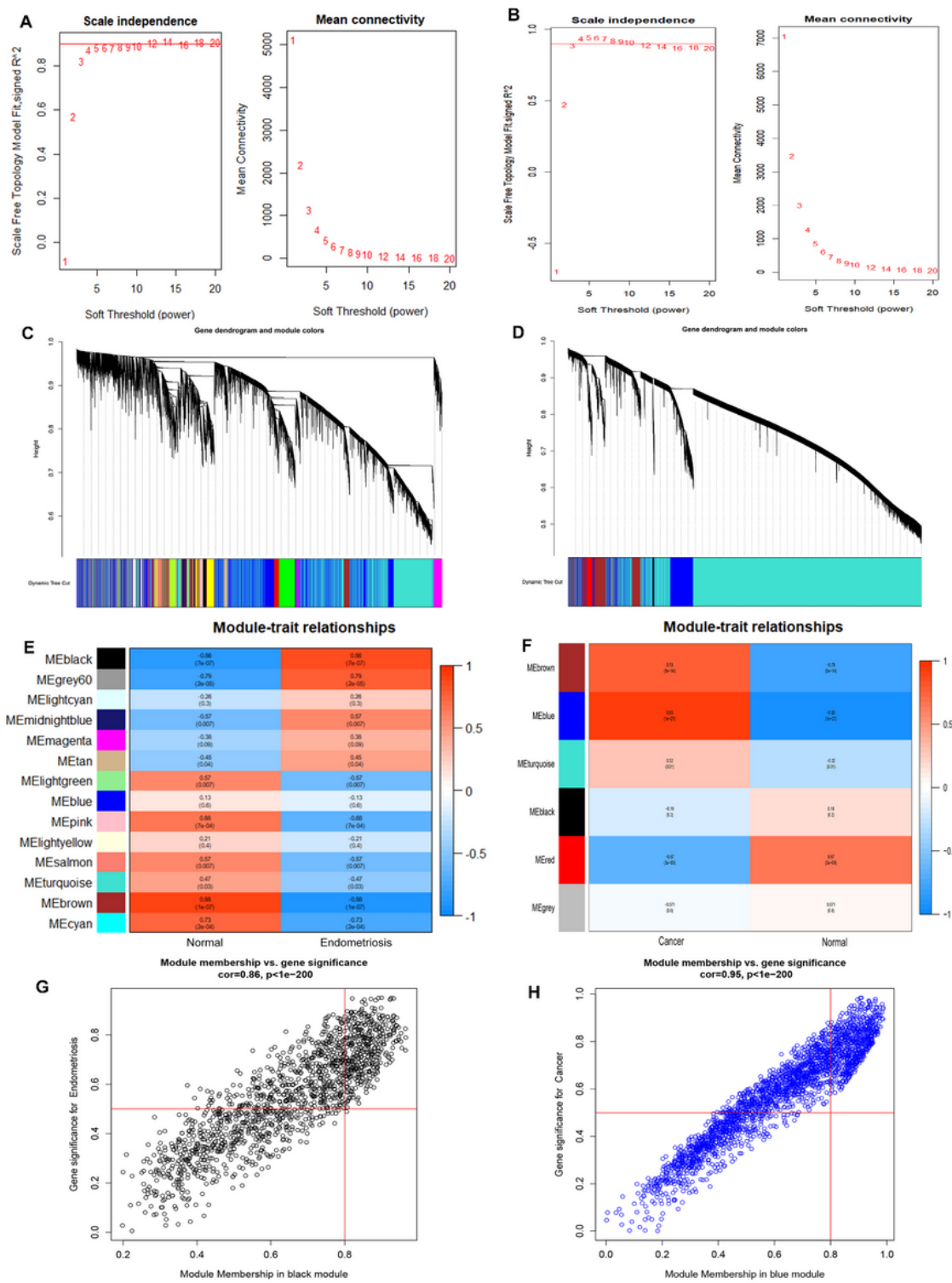


Figure 2

Identification of modules linked to clinical features of endometriosis in the GSE5108 dataset and ovarian cancer in the GSE18520 dataset by WGCNA. (A, B) Determining the appropriate “soft” threshold (β) that satisfied a scale-free network. (C, D) Co-expression module clustering in endometriosis (C) and ovarian cancer (D). Each branch represented each gene, and genes clustered into the same module were assigned the same color. (E, F) Heap of module–trait relationships in endometriosis (E) and ovarian cancer (F). (G,

H) Scatter plots for the relationships of module membership in the black module and blue module with gene significance for (G) endometriosis and (H) ovarian cancer.

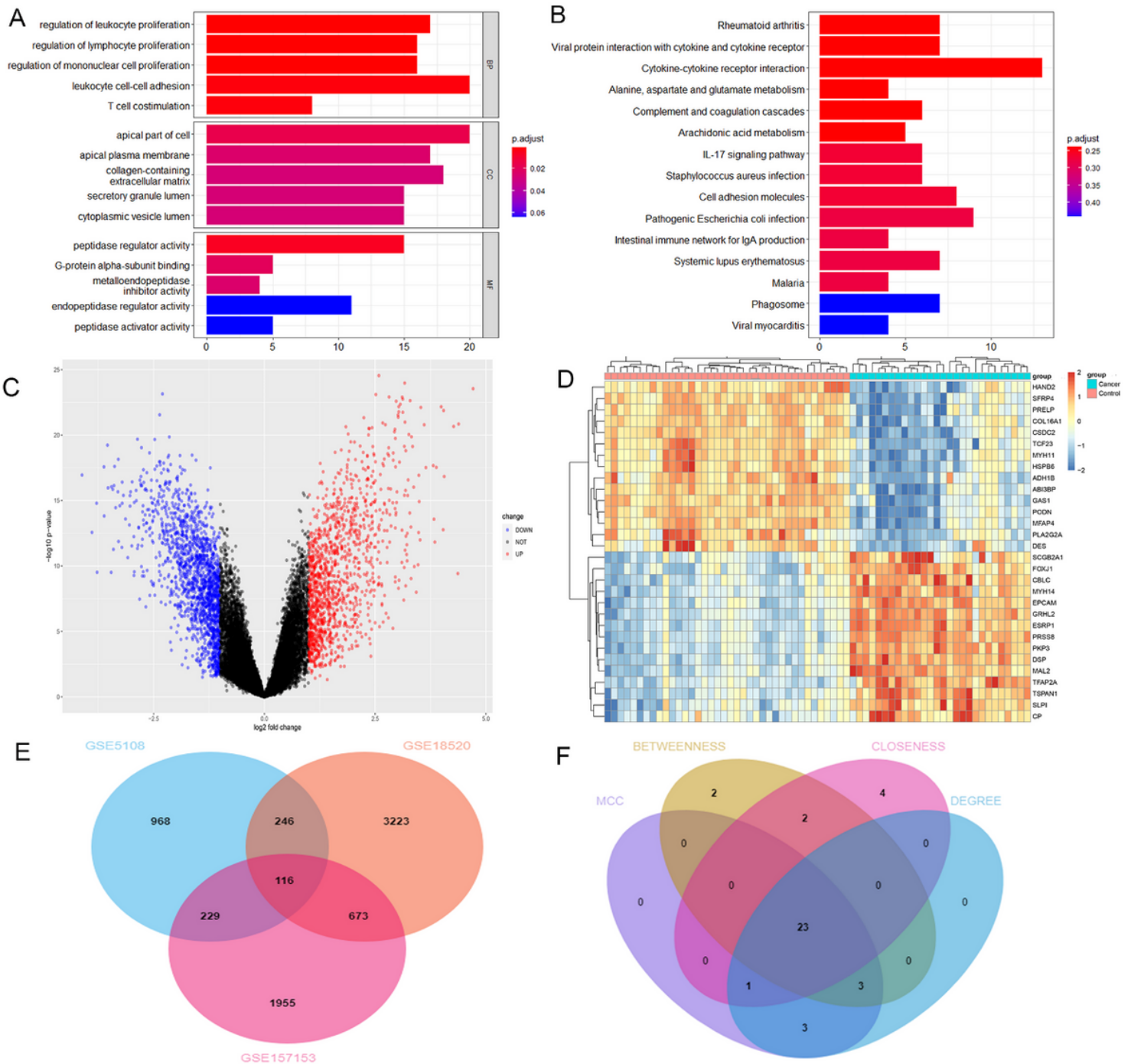


Figure 3

Functional enrichment analysis and identification of DEGs between endometriosis and EAO. (A) GO enrichment analysis. (B) KEGG pathways enrichment analysis. (C) Volcano diagram shows the DEGs in endometriosis and EAO. (D) Heatmap shows the top 30 up-regulated and down-regulated DEGs in endometriosis and EAO. (E) A total of 116 common genes were identified after intersection of genes

from three datasets. (F) 23 DEGs were selected for further analysis based on the intersection of genes from four algorithms.

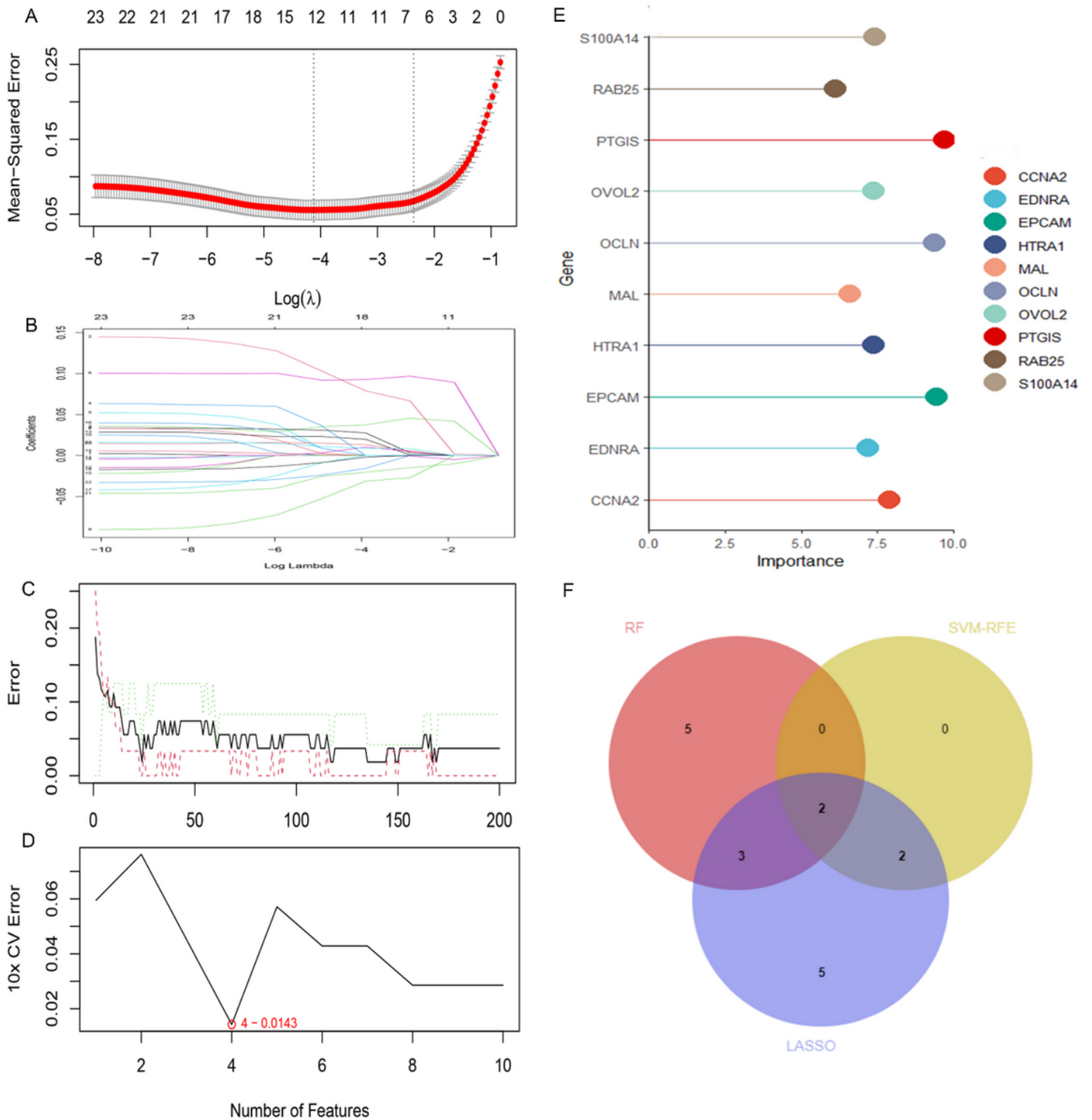


Figure 4

Machine learning algorithms for hub genes selection. (A) In the LASSO regression model, minimum standard was adopted to obtain the value of the super parameter λ by 10-fold cross-validation. Twelve genes were selected as the potential biomarkers with the lowest binominal deviation. (B) The overview of

LASSO coefficients. (C) Random forest construction. The number of trees used to build the random forest was specified as 200 before making the prediction. (D) Four genes were identified based on SVM-RFE with the lowest error. (E) The top 10 genes were selected based on the importance score of RF algorithm. (F) Two hub genes (OCLN, EDNRA) were finally identified via intersection of three machine learning algorithms.

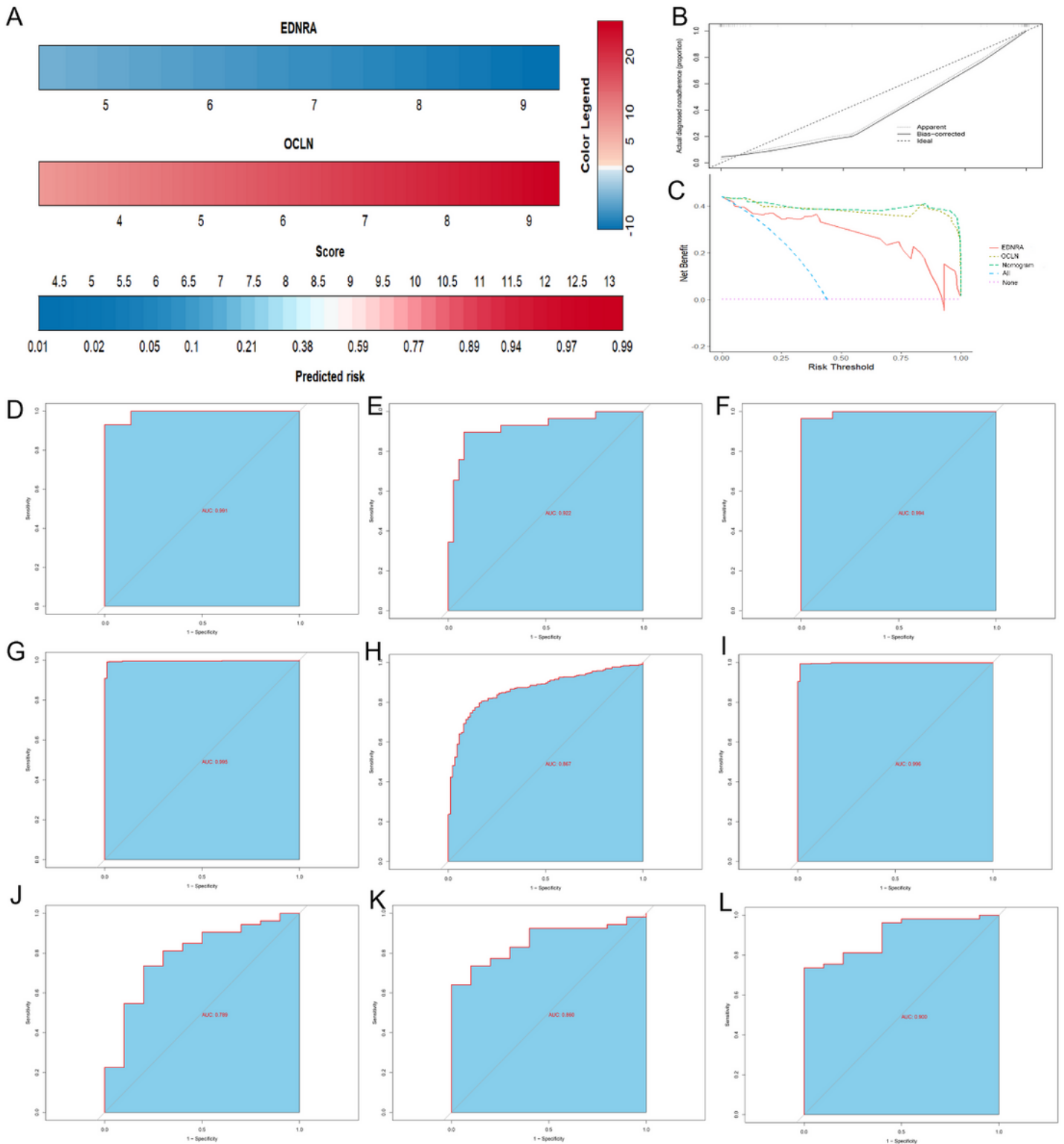


Figure 5

Nomogram construction and evaluation of diagnostic ability of hub genes. (A) Nomogram construction for diagnosing EAOC based on two characteristic genes. (B) Calibration curve. (C) Decision curve analysis (DCA). (D-F) ROC curves and AUCs were utilized in GSE157153 to evaluate the diagnostic performances of OCLN (D), EDNRA (E), and Nomogram (F). (G-I) ROC curves and AUCs were applied in TCGA-OV to evaluate the diagnostic performances of OCLN (G), EDNRA (H), and Nomogram (I). (J-L) ROC curves and AUCs were used in GSE18520 to assess the diagnostic performances of OCLN (J), EDNRA (K), and Nomogram (L).

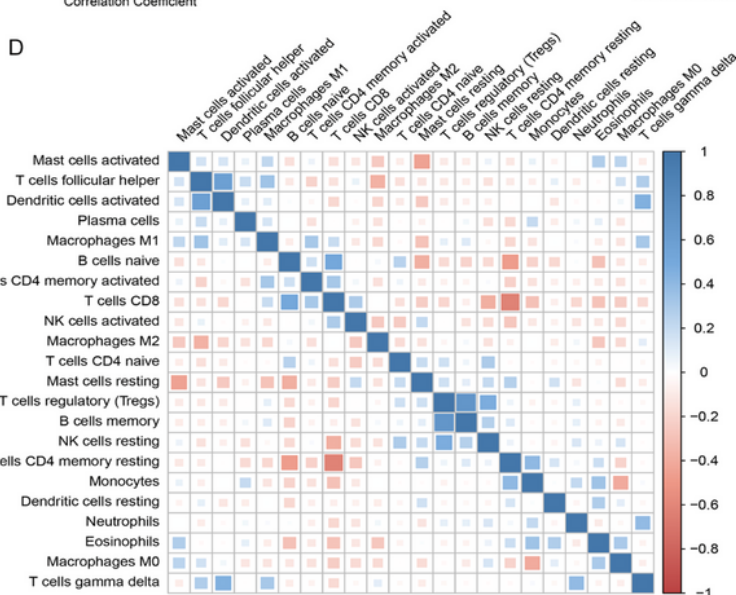
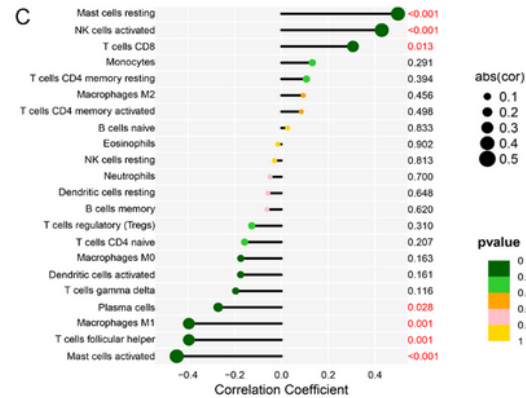
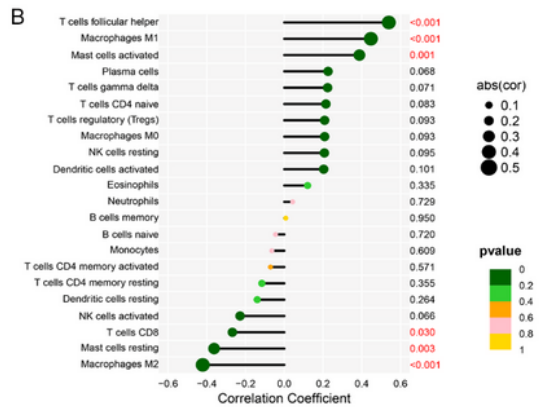
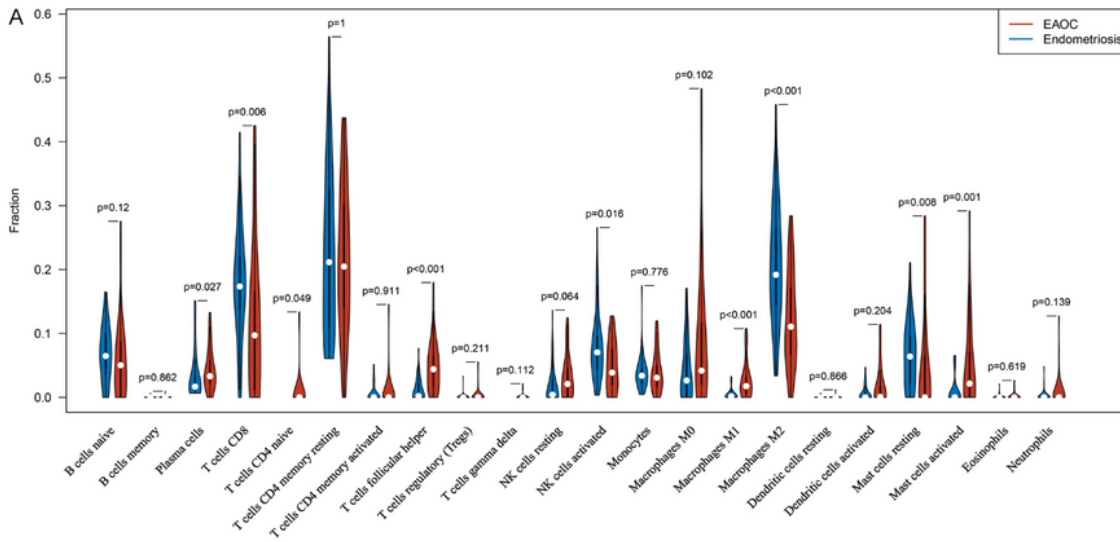


Figure 6

Immune cell infiltration between endometriosis and EAOc groups. (A) The infiltrating levels of 22 immune cell types between endometriosis and EAOc groups via CIBERSORT deconvolution algorithm. (B-C) Spearman analyses were conducted to investigate the correlation of OCLN (B) and ENDRA (C) with immune cell infiltration degree. (D) Correlation among 22 immune cells types, where red and blue denote positive and negative correlations, respectively, and white denotes no correlation between the designated immune cell populations.

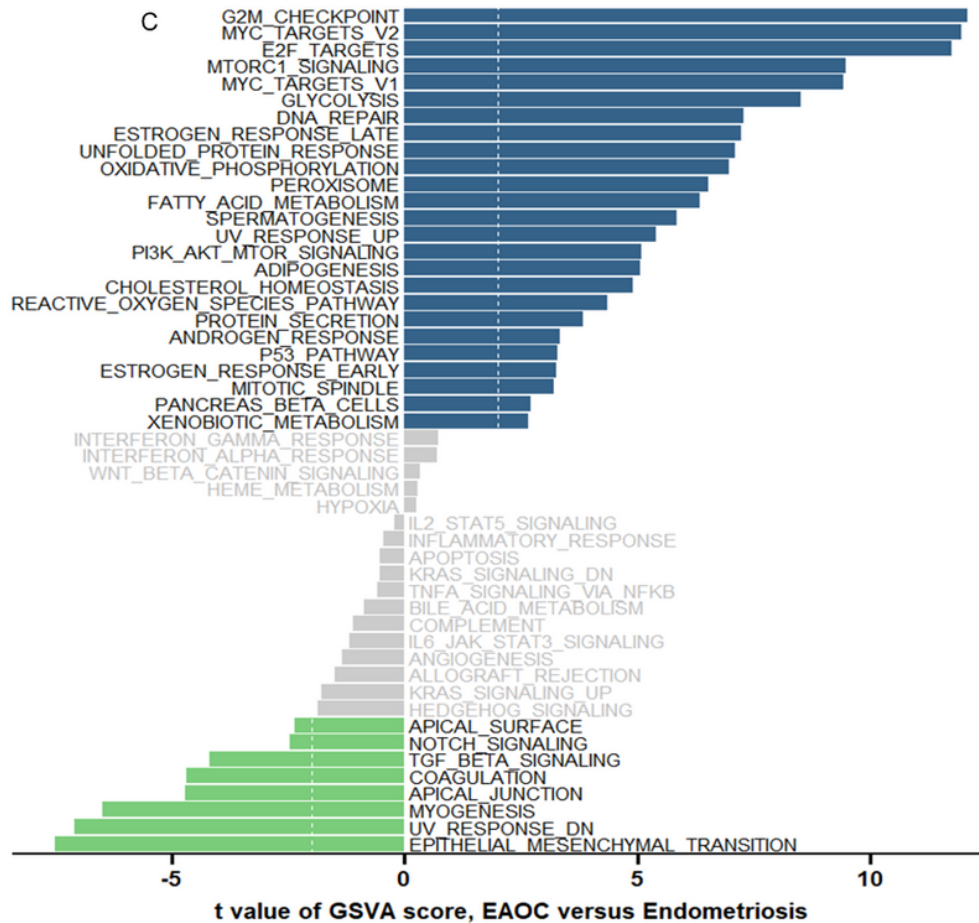
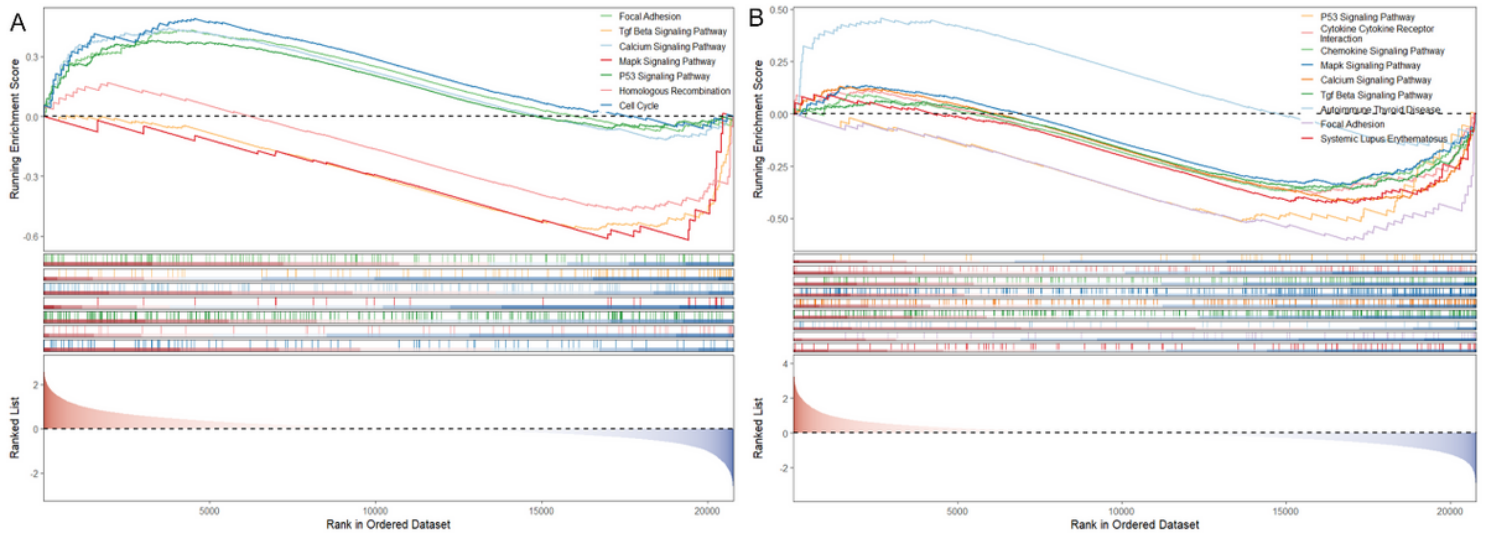


Figure 7

GSVA of hallmark gene sets between EAOc and endometriosis and GSEA functional analysis of hub genes. (A, B) GSEA identified signaling pathways involved in the characteristic genes of ENDRA (A) and OCLN (B). (C) GSVA of hallmark gene sets was conducted to reveal the differences in signaling pathways between EAOc and endometriosis patients.

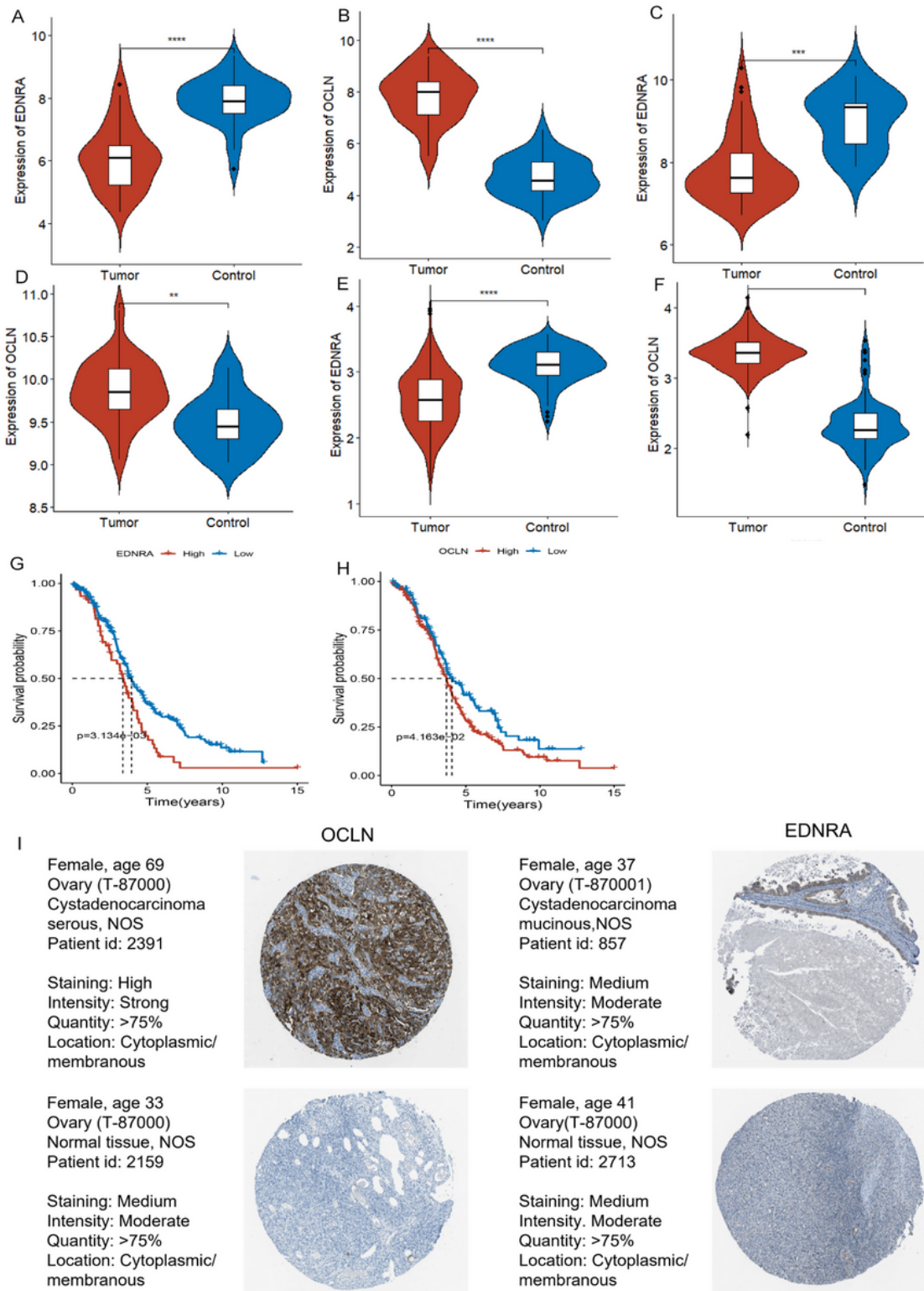


Figure 8

The expression profiles and validation of the two hub genes by immunohistochemistry analysis.

(A, B) The expression profiles of EDNRA (A) and OCLN (B) in EAOC and endometriosis patients in GSE157153 dataset. (C, D) The expression profiles of EDNRA (C) and OCLN (D) in EAOC and control groups in GSE18520 dataset. (E, F) The expression profiles of EDNRA (E) and OCLN (F) in EAOC and control groups in TCGA-OV cohort. (G, H) Kaplan-Meier curves were plotted to compare the overall survival difference between high- and low-expression groups with regard to EDNRA (G) and OCLN (H). (I) Immunohistochemistry of two hub genes in ovarian cancer tissues and control tissues from the Human Protein Atlas database.

Supplementary Files

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

- [Additionalfile1.tif](#)
- [Additionalfile2.tif](#)