Identification of microRNA and protein interaction networks in human ovarian cancer

Ponnusamy Nirmaladevi1, Ganapathi Keerthana2, Sri Rajkumar Sanjana2, Ul Husna Asma2 and Arumugam Mohanapriya1*

1. Department of Biotechnology, School of Biosciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu, 632014, INDIA
2. Department of Integrative Biology, School of Biosciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu, 632014, INDIA
*mohanapriyaas@vit.ac.in

Abstract

Ovarian cancer is one of the deadliest tumors in women, with a high mortality rate brought on by the lack of early detection. In this work, our main aim is to find promising biomarkers and pertinent mechanisms. GSE36668 was chosen from the Gene Expression Omnibus (GEO) to identify the differentially expressed genes (DEGs) using the GEO2R tool. To forecast gene ontology (GO) and pathway enrichment, online tools from ToppGene, FunRich and DAVID were employed. The protein-protein interaction (PPI) network is built via STRING v.11.5 and Cytoscape v.3.9.1. Following the detection of the hub genes, a Kaplan-Meier plotter was used to conduct additional validation survival analyses. A total of 1556 DEGs were identified using GEO2R, out of which 697 were upregulated and 859 were downregulated. According to GO analysis, DEGs were much more common in the online tools DAVID and ToppGene for cell adhesion, axoneme assembly and cilium assembly in the biological processes whereas cell surface is an essential component of the plasma membrane and extracellular matrix in the cellular component.

In contrast, the plasma membranes are present in DAVID and FunRich. The DEGs are mostly linked to the MAPK, PI3K-Akt and RAP1 signaling pathways in KEGG and in the Reactome pathway, they are involved in cell-cell communication, cell and cell-cell junction organization. The PPI network was constructed to find the gene clusters and to identify the hub genes MAPK1, CDH1, CBL and CCND1 by Cytoscape. The survival analysis of this hub gene CBL showed high expression in ovarian cancer which led to fewer survival chances. According to this study, ovarian cancer biomarkers are crucial to understand the molecular causes of the disease.

Keywords: Ovarian Cancer, miRNAs, Survival analysis, Differential expressed genes, FunRich.

Introduction

Ovarian cancer (OC) in recent years is one of the most common cancer-related diseases. Cancer is a leading cause of death globally and ovarian cancer is women's eighth most common cancer. OC is an uncontrollable growth of cells in the ovary. Ovaries are reproductive glands that produce eggs for reproduction in women. Usually, ovaries are made up of three types of cells that can develop into tumors and cause cancer namely epithelial tumors, germ cell tumors and stromal tumors. According to World Cancer Research Fund International, nearly 3,13,000 new OC cases were found in 20201. The risk factors2 of ovarian cancer may include a family history of having breast or ovarian cancer, obesity3, hormonal replacement therapy4 and woman who had late pregnancy or no children.

The recommended diagnoses are blood tests5, imaging tests6, laparoscopy7 and biopsy8. However, a biopsy can only confirm the presence of cancer. The treatment may depend on factors like stage and grade of cancer, the patient’s age and health state and the affordability of treatment. The treatment options are surgery9, chemotherapy10, targeted therapy11, radiation therapy and immunotherapy12. Though we have some detection methods, there is a lack of early and effective diagnosis. So, we must focus on understanding the disease and develop new effective diagnostic and therapeutic strategies.

Recently, biomarkers have played a prominent role in identifying cancer through molecular or other testing. It gives information about the cause of growth and allows the therapist to proceed with the treatment. Nowadays, differentially expressed genes (DEGs) are used to screen which may give a new vision of ovarian cancer13. MicroRNAs (miRNAs) are non-coding RNAs that regulate gene expression. It may act as an oncogene or tumor suppressor according to certain conditions. There are insufficient studies on miRNA, so focused research is needed to identify miRNA in ovarian cancer.

In this study, we aim to determine the effects of OC by assessing gene expression and we identified DEGs from GEO2R. We sought to functionally annotate genes using Gene Ontology (GO) and pathway enrichment analysis. From this, miRNA and protein interactions were analyzed. For further validation, a survival analysis was performed.

Material and Methods

Selection of the Ovarian Cancer Dataset: We acquired the microarray dataset GSE36668 of ovarian cancer from the Gene Expression Omnibus (GEO) database. The tissue samples were chosen from the serous ovarian borderline tumor and superficial scrapings from the normal ovary. Genes that exhibit differential expression under various conditions are considered significant.

https://doi.org/10.25303/1810rjbt1480153
experimental conditions can be found using the web-based interactive tool GEO2R from the NCBI-GEO database. A dataset of normal tissue (superficial scraping from the normal ovary) and ovarian cancer tissue (serous ovarian borderline tumor) was taken using GEO2R. Then they were adjusted to the P-value <0.05 and |log fold change (FC)|>2/<2 and these were considered as DEG's.14.

**Gene ontologies and Pathway enrichment:** Gene ontology (GO) analysis was done by FunRich (http://www.funrich.org/), DAVID (https://david.ncifcrf.gov/) and ToppGene (https://toppgene.cchmc.org/). The functional enrichments of cellular component (CC), biological process (BP) and molecular function (MF) from each of the databases were done. Using DAVID, pathway enrichment analysis (KEGG and Reactome pathways) was used to predict the DEGs involved in the pathways.15.

**Protein-protein interaction network construction and cluster analysis:** The protein-protein interaction (PPI) network of the identified DEGs was constructed by an online tool, STRING v.11.5 (the Search Tool for the Retrieval of Interacting Genes/Proteins) with a high confidence of 0.7. The STRING (https://string-db.org/) contains information from various sources including experimental data, computational prediction methods and public text mining. The clusters of the PPI network were presented using the molecular complex detection (MCODE) plugin by Cytoscape v.3.9.1 with the parameters of degree cutoff = 2, node score cutoff = 0.2, k-score = 2 and max. Depth = 100.

Using the cytoHubba plugin, the top 10 genes were considered as hub genes based on the algorithms of degree, closeness and radiality.16. The selected hub gene was further analyzed for interaction with miRNAs using Network Analyst v.3.0 (https://www.networkanalyst.ca/).

**Survival analysis and hub gene validation:** The prognostic effect of genes on the survival analysis of ovarian cancer can be retrieved using the Kaplan-Meier plotter.14 According to the ovarian cancer expression level, they were divided into two groups: a high-expression group and a low-expression group. The general inclusion criteria for the survival analysis were set as follows: (1) user-selected probe set (2) using a 2022 version dataset and (3) excluded biased arrays. The hub gene must be considered when the p-value is <0.05.

**Results**
The workflow and approaches towards our current study are depicted in figure 1.

**Differentially Expressed Genes (DEGs) Identification:** The expression profile GSE36668 was obtained from the GEO database. It consisted of four cases- serous ovarian borderline tumor and four control of superficial scraping from the normal ovary. This dataset identified a total of 1556 DEGs containing 697 upregulated and 859 down-regulated genes (Figure 2).

**Functional annotation:** The ontologies of DEGs were analyzed using FunRich, DAVID and ToppGene (Table 1). DAVID and ToppGene show similar functions in CC regarding integral component of plasma membrane, cell surface and apical plasma membrane. Compared to CC, the BP term had similar functions in cilium movement, cell-cell adhesion, axoneme assembly and cilium assembly.

---

[Diagram of Workflow](http://example.com/figure1.png)

![Figure 1: Workflow of this study](http://example.com/figure1.png)
In contrast, MF terms of serine-type endopeptidase activity and calcium ion binding were similar. In DAVID and FunRich, similar functions were denoted only in the CC term of the plasma membrane. Pathway enrichment, the KEGG pathway showed ECM-receptor interaction, pathways in cancer, PI3K-Akt signaling pathway and cell adhesion molecules whereas the Reactome pathway showed diseases of glycosylation.

Protein-protein interaction network construction and cluster analysis: A total of 1556 DEGs were imported into the STRING database. Subsequently, all the genes were mapped into the PPI network. The PPI network consists of 1197 nodes and 1133 edges and the PPI enrichment p-value was lower than $< 1.0e^{-16}$. The significant clusters were identified through the MCODE plugin by the Cytoscape. The top four clusters were selected and validated. The cluster 1, MCODE score =11 (KIF4A, BIRC5, BUB1B, DTL, KIAA0101, CCNB2, CDC20, AURKB, RRM2, CDC45, PBK) had 11 nodes and 55 edges. The cluster 2, MCODE score =10 (DRC1, RSPH4A, RSPH9, DNAAF3, DNAH5, CCDC40, CCDC114, DNAI2, DNAAF1, DNAI1) had 10 nodes and 45 edges.

The cluster 3, MCODE score =6 (STAC3, CACNA1E, C1orf101, CACNA2D1, CACNA1D, CACNB2) had 6 nodes and 15 edges. Cluster 4, MCODE score =6 (HLA-DQA1, HLA-DQB1, HLA-DOA, HLA-DPB1, B2M, HLA-DM) had 6 nodes and 15 edges. We selected the hub genes by using the CytoHubba plugin based on methods namely degree, closeness and radiality (Figure 3).

Survival analysis and hub gene validations: The prophecy details of these four hub genes for ovarian cancer and serous ovarian borderline tumor at non-identical stages were analysed by the Kaplan-Meier plotter (Figure 5). The hub gene CBL has HR=1.56 (1.28-1.89) p=7.6e-06. It is involved

Table 1

<table>
<thead>
<tr>
<th>Dataset</th>
<th>FunRich</th>
<th>DAVID</th>
<th>ToppGene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extracellular</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasma membrane</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extracellular region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytoplasm</td>
<td></td>
<td></td>
<td>Axoneme</td>
</tr>
<tr>
<td>Plasma membrane</td>
<td></td>
<td></td>
<td>Ciliary plasm</td>
</tr>
<tr>
<td>Integral component of membrane</td>
<td></td>
<td></td>
<td>Plasma membrane bounded</td>
</tr>
<tr>
<td>Extracellular region</td>
<td></td>
<td></td>
<td>Cell projection cytoplasm</td>
</tr>
<tr>
<td>Membrane</td>
<td></td>
<td></td>
<td>Cytoplasmic region</td>
</tr>
<tr>
<td>Membrane</td>
<td></td>
<td></td>
<td>axonemal microtubule</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Biological process</th>
<th>FunRich</th>
<th>DAVID</th>
<th>ToppGene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signal transduction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signal transduction</td>
<td></td>
<td></td>
<td>Cilium movement</td>
</tr>
<tr>
<td>Positive regulation of transcription from RNA polymerase II promoter</td>
<td></td>
<td></td>
<td>Epithelial cell differentiation</td>
</tr>
<tr>
<td>Cell differentiation</td>
<td></td>
<td></td>
<td>Axoneme assembly</td>
</tr>
<tr>
<td>Cell adhesion</td>
<td></td>
<td></td>
<td>Microtubule bundle formation</td>
</tr>
<tr>
<td>Positive regulation of gene expression</td>
<td></td>
<td></td>
<td>Regulation of morphogenesis of an epithelium</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Molecular function</th>
<th>FunRich</th>
<th>DAVID</th>
<th>ToppGene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth factor activity</td>
<td></td>
<td></td>
<td>Troponin T binding</td>
</tr>
<tr>
<td>Calcium ion binding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth factor activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP-dependent microtubule motor activity, minus-end-directed</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
in negatively regulating receptor tyrosine kinase (RTKs) and also it acts as adaptor proteins which in turn activate RTKs.

**Discussion**

One of the reasons for ovarian cancer deaths, which increase the risk of metastasis and chemotherapy resistance, is the lack of effective detection methods. Therefore, finding precise molecular mechanisms and relevant biomarkers for the early detection, management and prognosis of ovarian cancer can shed light on cancer research. Malignant ovarian lesions comprise of primary lesions that develop from normal ovarian functions and secondary lesions that develop from cancers in other parts of the body.

---

**Figure 3:** The PPI network of the OC using the Cytoscape-cytoHubba plugin (A) Degree (B) Closeness and (C) Radiality

**Figure 4:** The hub genes interacted with miRNAs using Network analyst
A convenient and all-encompassing platform for exploring general genetic alterations identifying DEGs and elucidating molecular mechanisms for the identification, therapy and prognosis of tumours has recently been made available due to the development of bioinformatics. GSE536668, a GEO dataset, was chosen for this study's DEG analysis. The intersecting 1556 DEGs, which include 697 upregulated and 859 downregulated DEGs, were found using the online tools GEO2R and Funrich. The DEGs were then subjected to gene ontology and KEGG pathway analyses using the online tool DAVID. They were primarily involved in the calcium signalling pathway, RAS signalling pathway, RAP1 signalling pathway and MAPK signalling pathway, according to the KEGG pathway.

Additionally, these results provided crucial pointers for studying how molecules interact during the development of OC. Numerous investigations have revealed a significant relationship between metabolic pathways, the cell cycle and the emergence and growth of OC tumours. For instance, it is well known that autophagy affects several processes including survival in challenging metabolic conditions and that cancer-associated fibroblasts may also impact the actual autophagy level in OC cells. Studies have identified many antibiotics including minocycline, salinomycin, monensin and pegylated liposomal doxorubicin, that may effectively treat ovarian cancer. Here, a PPI network is built using the online tool STRING.

1197 nodes made up the constructed PPI network string and 1133 of its edges were built using DEGs. The DEG interaction network was built using FunRich. The remarkable clusters were then discovered using molecular complex detection (MCODE) in the Cytoscape app. Scores for MCODE were noted. Using the highest degree scores found in the cytoHubba plugin, the top four genes were determined to be hub genes. Using survival analysis, the four hub genes MAPK1, CDH1, CBL and CCND1 are discovered.

However, compared to patients without MAPK1 amplification, patients with MAPK1 amplification had significantly worse progression-free survival. The gene CDH1 is responsible for producing E-cadherin. Although somatic E-cadherin mutations have been linked to increased cell invasion and metastasis in ovarian cancer, germline CDH1 mutation carriers have not been shown to have an increased risk of epithelial ovarian cancer. The expression of the CCND1 gene was discovered to be significantly higher in ovarian cancer tissue than in healthy ovarian tissue by analyzing the Adib Ovarian, Bonome Ovarian and Hendrix Ovarian microarrays.

**Conclusion**

In summary, an integrated bioinformatics analysis may be critical to accelerating progress in cancer research. As a result of establishing the miRNA-hub gene network and potential targeted medications linked to the hub genes, it may be possible to learn more about the mechanisms underlying ovarian cancer.

**Acknowledgement**

We would like to thank Vellore Institute of Technology (VIT) for providing computational facility.

**References**


(Received 23rd May 2023, accepted 25th June 2023)