Identification of core genes shared by endometrial cancer and ovarian cancer using an integrated approach

Yumin Ke¹, Xuanxuan Zhuang¹, LiuXia You²

¹ Department of Obstetrics and Gynecology, the Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian, China
² Department of Clinical Laboratory, Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian, China

ABSTRACT

Ovarian cancer (OC) is one of the most commonly diagnosed cancers among women. Moreover, Endometrial cancer (EC) is a usual genital tract cancer in females; however, the hub genes and molecular pathways shared by these two cancers have not been surveyed yet. So, this study aimed to identify the common candidate genes or biomarkers and molecular pathways in OC and EC. Differences in the expressed genes between these two microarray data sets were detected. Pathway enrichment analysis and gene ontology (GO) was also performed and protein-protein interactions (PPI) network analysis was done using Cytoscape and the most important genes were identified by the Cytohubba plugin. We found that 154 common DEGs shared by OC and EC were also detected. 10 hub proteins were identified as follows: CDC20, BUB1, CENPF, KIF11, CCNB2, FOXM1, TTK, TOP2A, DEPDC1, and NCAPG. The most important and significant miRNAs were identified to be hsa-mir-186-5p, hsa-mir-192-5p, hsa-mir-215-5p, and hsa-mir-193b-3p regulated expressions of the DEGs. This investigation demonstrated that these hub genes and their miRNA might be key genes with great effects on OC and EC. However, more studies are needed for a better understanding of the role of these hub genes and their function in these two cancers.

Introduction

Among gynecologic cancer prevalent in women, ovarian cancer is known as the leading cause of death among women and is the fifth leading cause of death in the United States. According to global statistics in 2012, 238,700 new female cases of ovarian cancer were identified and 151,900 deaths were reported due to this cancer (1). In 2016, this number decreased and approximately 22280 new cases were identified, of them, 14240 cases died of this disease (2). Unfortunately, due to the lack of sufficient information on the early diagnosis of this disease, 80% of cases are usually diagnosed at the advanced stage of the disease and unfortunately less than 40% of women recover from this disease (3).

The use of serum CA-125 is one of the most common methods applied for diagnosing ovarian cancer in women, which has shown an increase of 80 to 85%. Unfortunately, in most cases, this biomarker is not able to detect this disease at the early stages of the disease, and only 50% of these items are usually detected by this indicator. CA-125 is a useful widely used method for the detection of ovarian cancer; however, it does not have the sensitivity needed for the early diagnosis of ovarian cancer (4).

Another cancer specific to women, which unfortunately causes a large number of deaths every year, is endometrial cancer. Accordingly, this is the sixth most common cancer among women. In 2018, 382069 people were found with this disease and 8,926 people lost their lives (5). The most important reasons for this type of cancer are metabolic syndromes and obesity, which have led to an increase in the number of mortality due to this disease in recent years. Due to the growing trend of this disease among women, the number of deaths is also expected to increase to 20.3% in 2025 (5).

Lack of sufficient knowledge on the early detection of this disease causes tumor recurrence, which was shown to be directly related to mortality. Although the incidence of endometrial cancer is less prevalent in underdeveloped countries compared to other developed countries, other factors cause death in these societies (6,7). For example, in Canada and Europe, the incidence of endometrial cancer is ten times higher than that of less developed countries. In general, this cancer is the fourth most common cancer among women after breast, lung, and intestinal cancers (5,8).

Biomarkers are used either to distinguish biological processes at different stages or to evaluate pathogenic processes and drug responses at different treatment stages (9). Correspondingly, these markers can provide a better understanding of the complexities.

One of the applications of biomarkers is the choice of a systematic treatment for patients. Based on various investigations performed in this regard, especially high-throughput omics platforms give us the opportunity to use microRNAs, RNAs and DNA as well as proteins, as biomarkers of metabolic epigenetic changes, helping to identify their use at different stages of the disease (10).

One of the major disadvantages of this type of indicator is that it is expensive. Given this, researchers mostly de-
decide to use some markers such as immunohistochemistry (IHC) and traditional targeted DNA sequencing, which are available and cost-effective (11).

One of the most important tools currently used to study genetic data on a large scale is gene expression microarray. Accordingly, this is generally used to collect and study gene expression data in many diseases, especially human cancers. This method is used to study tumor genes, molecular targeting, molecular prevention, and treatment of cancer. By creating a database for information on the gene expression in various diseases, it became possible for researchers to provide a better study on the mechanism of diseases (12,13).

In the present study, we have employed a comprehensive bioinformatics approach, in order to decode mutual genes and pathways between OC and EC for clarifying their potential shared mechanisms (Figure 1).

In this investigation, the raw microarray data were obtained from the National Center for Biotechnology Information (NCBI) database. Moreover, we used some bioinformatics software for analyzing and identifying DEGs between OC and EC. Afterward, KEEG and gene ontology analyses of DEGs were performed to understand the underlying molecular mechanisms.

### Materials and Methods

#### Microarray data and Pre-processing and statistical analysis in EC and OC

A publicly available dataset for Ovarian Cancer was downloaded from the Gene Expression Omnibus (GEO) repository (accessed using the number of GSE18521) microarray expression dataset. Notably, this dataset is based on the Affymetrix GPL570 platform (Human Genome U133 Plus 2.0 Array), submitted by Mok (14). For Endometrial Cancer, a dataset (GEO accession number GSE17025) was used, which is based on the Affymetrix GPL570 platform (Human Genome U133 Plus 2.0 Array) submitted by Day (15). The GSE18521 dataset contained 91 samples, including 79 tumor samples and 12 normal samples. The GSE17025 dataset contained 63 samples, including 53 tumor samples and 10 normal samples. In this investigation, for identifying DEGs between the tumor and normal samples, GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) was applied. Accordingly, GEO2R is an online web tool that compares the two groups of samples under the same experimental condition. Correspondingly, this can also analyze most of the GEO series (16). This dataset was initially filtered to include only the measurements with the signal power p-value < 0.05 and absolute log fold-change greater than 2.

#### Gene ontology and pathway enrichment analysis

The DEGs of the present investigation were analyzed to detect their biological functions. As well, GO and KEGG pathways were analyzed using the Integrated Annotation and Visualization Database (DAVID) (http://david.ncifcrf.gov/), and online tools were also considered. P < 0.05 was considered as the statistically significant level. For obtaining the best results, those interactions with the highest confidence scores were selected in the STRING database (17). The PPIs were then analyzed by Cytoscape software (version 3.8.2) (18), in order to analyze and visualize the PPI network.

#### Protein-protein interaction analysis

Cytoscape is software developed by an international consortium of open-source developers (18). Moreover, there are some plugins used to find the best network as well as some genes with a high interaction like cytoHubba (19).

#### Hub genes validation by SurvExpress

After identifying the hub genes from these microarray expression profile datasets of OC and EC cross-validated from the cancer genome atlas (TCGA) was used via SurvExpress (20). We used a box plot to analyze DEG between high-risk and low-risk groups as well as sureExpress to analyze Cox regression. For considering any statistically significant difference, we employed Kaplan–Meier plots and a log-rank p-value < 0.05. Moreover, the network analyst 3.0 visualization web tool was utilized to extract and analyze the miRNA-target gene interaction networks (21).

### Results

#### Identification of DEGs

In this study, for identifying DEG in OC and EC statistically analyzed using Limma, 8277 genes for OC and 11874 genes for EC were measured. We compared the DEGs of both the OC and BC datasets to identify some genes expressed between them. We identified 154 genes as the common gene between the two datasets. Thereafter, we considered these 154 DEGs for further studies. The Venn plot of each gene expression profile data is shown in Figure 2.

#### Gene ontology and KEGG pathway analysis

DEGs data were uploaded to DAVID for identifying...
GO pathways. The significant enriched biological processes (BP) were as follows: ‘mitotic nuclear division’, ‘positive regulation of transcription from RNA polymerase II promoter’, ‘positive regulation of cell proliferation, ‘cell proliferation, and ‘negative regulation of transcription from RNA polymerase II promoter’. Moreover, DEGs data were significantly enriched in MF, including ‘heparin binding’, ‘dipeptidyl-peptidase activity’, ‘microtubule binding’, ‘RNA polymerase II core promoter proximal region sequence-specific DNA binding’, and ‘sequence-specific DNA binding’. Thereafter, for cellular components (CC), the common DEGs between OC and BC were enriched in ‘spindle’, ‘midbody’, ‘interstitial matrix’, ‘centrosome’, and ‘extracellular space’ (Table 1).

Table 2. indicates the result of the KEGG analysis stating that the most significantly enriched pathways in the common DEGs between OC and BC were in ‘Oocyte meiosis’ and ‘Cell cycle’.

### Identification of hub genes

Working by interacting with each other is known as one of the main properties of biomolecules in biological systems; therefore, based on the PPI network, the common DEGs between OC and BC were analyzed using the Cytohubba plugin in Cytoscape software. (Figure. 3) in addition, key hub proteins, namely CDC20, BUB1, CENPE, KIF11, CCNB2, FOXM1, TTK, TOP2A, DEPDC1, and NCAPG were detected by topological analysis (Table 3).

### Hub genes validation and identification MicRNA

For validating the expressions of these key potential genes, the online tool, cross-validated with TCGA datasets was used for OC and EC. These results show that the selected hub genes have statistically significant differential expression between these two cancers, as shown in figure 4 and Figure 5.

The validation of hub genes/proteins was performed for the assessment of the selected genes as a biomarker in cancers and enhancing treatment plans in OC and EC.

**Figure 3.** Protein network of the common genes in OC and EC. The ten red large nodes indicate the hub proteins.

<table>
<thead>
<tr>
<th>Pathway ID</th>
<th>Name</th>
<th>Count</th>
<th>P-value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa04114</td>
<td>Oocyte meiosis</td>
<td>4</td>
<td>0/019</td>
<td>CDC20, CCNB2, PGR, BUB1</td>
</tr>
<tr>
<td>hsa04110</td>
<td>Cell cycle</td>
<td>4</td>
<td>0/026</td>
<td>CDC20, CCNB2, TTK, BUB1</td>
</tr>
</tbody>
</table>
The prognostic evaluation revealed that the hub genes, e.g. CDC20, BUB1, CENPF, KIF11, CCNB2, FOXM1, TTK, TOP2A, DEPDC1, and NCAPG were statistically significant in the prognosis of OC, with a log-rank p-value of 0.00039 and hazard ratio of 1.82 (figure 6a). These hub proteins showed a significance in the prognosis of EC, with a log-rank p-value of 0.00028 and hazard ratio of 5.15 (Figure 6b).

Discussion

The present investigation provided a survey of DEGs in both ovarian cancer (OC) and endometrial cancer (EC). Biomarkers were studied by analyzing microarray data, which are widely used today to identify complex biological interactions in various diseases (22,23). By analyzing the expression data of ovarian and endometrial cancers’ gene expressions, 154 transcripts multiplying these two...
diseases were found to be significantly different, which are known as the common genes between these two diseases. To better understand the effects of these common genes between the OC and EC, the protein interactions of these genes were examined. After performing network and PPI analyses, 10 hub genes were identified, namely CDC20, BUB1, CENPF, KIF11, CCNB2, FOXM1, TTK, TOP2A, DEPDC1, and NCAPG. Moreover, some miRNAs were identified to have interactions with hub genes such as hsa-mir-186-5p, hsa-mir-192-5p, hsa-mir-215-5p, and hsa-mir-193b-3p that had the highest scores.

The results of this study are in agreement with the findings of a great deal of the previous works performed in this field. CDC20 is one of the genes that can play important roles in the development and spread of cancer in the human body. Therefore, one of the ways to treat cancer can be controlling the expression of the CDC20 gene, which can be done by targeting several upstream genes that control the expression of this gene such as sp53, RASSF1A, EMI1, and USP44 (24). Recent evidences showed that four genes BUB1B, BUB1, TTK, and CCNB1 are significantly enriched in the cell cycle pathway via re-analyzing DAVID in Ovarian cancer (25). In addition, according to the statistical analyses performed between the expressions of Bub1/Mad2 genes, a significant difference was found between the expressions of these two genes and the survival of endometrial cancer, which can be known as a valuable factor in both the diagnosis and prevention of this cancer (26).

As mentioned previously in the article research by Liu (2019), three overlapping genes between the differentially expressed genes and miRNAs targets, BIRC5, CENPF, and HJURP, were found to be associated with significantly worse overall survival of patients with EC (27). Moreover, among the upregulated genes, CENPF and UHRF1 were found to be involved in regulating the cell cycle by regulating transcription factors as well as miRNAs (28). This gene leads to tyrosine kinase 1 degradation by overexpression in ovarian cancer cells (29).

These findings further supported the idea that KIF11 gene expression is higher in cancerous tissues compared to healthy tissues, and is associated with tumor grade, stage TNM, and lymph node invasion (30).

The result of the analysis of the gene expression data (RNA-seq) showed a biological pathway for FOXM1-SLC27A2 that may have promoted endometrial cancer. Additionally, the loss of the FOXM1 gene also has a direct effect on cancer cells. Thus, FOXM1 can affect the biological activity of cancer cells, so it can be studied as an important factor in regulating cancer cells for treatment and diagnosis (25). As mentioned in the article research by Barger (2015), a role was demonstrated for FOXM1 in cell cycle progression using primary and immortalized human OSE cells as well as an HGSOC cell line (31).

One of the main biomarkers that can help researchers in the detection of various stages of cancers is miRNAs. However, some of these markers could not clearly represent their independent clinical potentials; therefore, these biomarkers failed to enhance the accuracy of the current monograms. After the first discovery of miRNAs by Lee in 1993, many studies investigated whether the dysregulation of miRNA expression in cancer has a correlation with both the risk and progression of this disease (32). One of these miRNAs that have been extensively studied in various cancers is miR-186. Accordingly, this miRNA is used as one of the biomarkers for the early detection of various cancers, because it affects various biological processes (33).

In our study, we identified CDC20, BUB1, CENPF, KIF11, CCNB2, FOXM1, TTK, TOP2A, DEPDC1, and NCAPG as hub genes in OC and EC. As well, the functions of these major genes in this cancer are needed to be deeply analyzed.

The current study provided a bioinformatics analysis of DEGs, which might contribute either to tumorigenesis or to the progression of OC and EC patients. Further studies are needed for having a better understanding of the role of these hub genes and their function in these cancers.

Acknowledgments
There is nothing to be declared.

Interest conflict
The authors declare that they have no competing interests.

Funding source
The author(s) received no specific funding for this work.

Compliance with ethical standards
The authors declare that there is no conflict of interest. This article does not contain any research involving humans or animals as subjects of research.

Additional information
The text was submitted by the authors in English.

References


