Verifying the markers of ovarian cancer using RNA-seq data

TIANFENG LIU1, NINA YU1, FENG DING1, SURONG WANG1, SHIHONG LI1, XIAOFEI ZHANG1, XIANGXIU SUN1, YING CHEN1 and PEISHU LIU2

1Department of Gynecology, Linyi People’s Hospital, Linyi, Shandong 276000; 2Department of Gynecology, Qilu Hospital of Shandong University, Jinan, Shandong 250000, P.R. China

Received March 13, 2014; Accepted December 12, 2014

Abstract. Markers associated with diagnosis, presentation and potential therapeutic targets have received widespread attention in ovarian cancer research in the past few years. However, the majority of these markers have been investigated individually, and the changes in expression and the association between them are rarely documented. Next-generation sequencing, also termed RNA-seq when the sequencing targets are cDNAs, can provide a whole blueprint of the transcriptome of a specific tissue. In the present study, RNA-seq data of human ovarian cancer samples were used to verify the expression of known markers and to identify the association between them. A total of 563 markers associated with ovarian cancer were retrieved from the database of the National Center of Biotechnology Information, and used as the target markers. The transcriptome of the ovarian tissue of four different tumors, containing tumor presentation and recurrence stages, were sequenced using the Illumina GAII platform. Approximately 85.97% markers were expressed of the total 563 markers, and the majority of them were involved in pathways associated with cancer, signaling and infection. In total, 85 markers were found to be aberrantly expressed in tumor cells from patients with ovarian cancer who had recurrences, including 33 upregulated markers at the recurrence stage. Therefore, they may have roles ovarian tumor due to their aberrant expression. Differentially expressed markers and the associations between them can be assessed by examining the RNA-seq data. These findings may provide novel information for further studies on ovarian cancer.

Introduction

Ovarian cancer is one of the most common types of carcinoma in females, usually associated with a high mortality as signs and symptoms are frequently absent (1). The exact causes of ovarian cancer remain to be elucidated. Genetic factors have been identified as important in certain ovarian cancer patients, such as mutations in the BRCA1 and BRCA2 genes, which are also risk factors for breast cancer (2). Mutations in BRCA1 confer a high risk of ovarian cancer and can reduce lifespan (3). In addition, marker genes of ovarian cancer can be classified into several types: Presentation (4,5), recurrence (6), inheritance (7), prognosis (8,9) and therapy targets (10). Although a large number of markers have been observed to be associated with ovarian carcinoma, the incidence of and the associations and interactions between these markers have not been extensively investigated.

In the present study, all markers associated with ovarian cancer were determined by analyzing the whole transcriptome of ovarian cancer cell lines. Tumor samples were selected at two stages (presentation and recurrence) from two patients as the focus of the present study. RNAs were extracted from tumor cell lines and sequenced with Illumina GAII. By aligning RNA-seq data to marker sequences, differentially expressed markers (DEMs) were filtered out in different stages, and the interactions between them were assessed.

Materials and methods

Data sources and processing. A total of four ovarian carcinoma cell lines at different tumor stages: PEO1 (first recurrence, 22 month post chemotherapy; Patient ID, 3), PEO4 (second recurrence, 10 months post first recurrence; Patient ID, 3), PEO23 (recurrence, 7 months post chemotherapy; Patient ID, 2), and PEO14 (presentation, prior to chemotherapy; Patient ID, 2) (11), were grown in vitro. Original tissue materials were collected in a study which was approved by the Cambridge Local Research Ethics Committee (LREC), all patients gave written informed consent prior to participation (12). mRNA of the cells was extracted and then sequenced on Illumina GAII genome analyzer (Illumina Inc., San Diego, CA, USA). Each sample was subject to three experimental repeats for statistical calculation. The raw sequence data were downloaded from the European Molecular Biology Laboratory, European Bioinformatics Institute database (http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-691/). High-quality data were obtained by filtering raw reads with the fastx-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) (13) (Table I). The quality of the data was controlled using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) (14).

Correspondence to: Dr Peishu Liu, Department of Gynecology, Qilu Hospital of Shandong University, 107 West Culture Road, Jinan, Shandong 250000, P.R. China
E-mail: peishuliu@126.com

Key words: ovarian cancer, RNA-seq data, markers
Marker selection. All the markers of ovarian carcinoma were selected from the Nucleotide database in National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). Markers were selected according to the following criteria: The nucleic acid sequence existed, the sequence type was RNA and there were supporting documents (listed in the Nucleotide database with their corresponding marker). Normal human genes were downloaded from University of California Santa Cruz (UCSC) Genome Bioinformatics (http://genome.ucsc.edu/, version hg19). Gene family information was downloaded from the HUGO Gene Nomenclature database (http://www.genenames.org/).

Interaction network of markers pathways. Pathways enrichment analysis for all the selected markers was performed by the Kyoto Encyclopedia of Genes and Genomes (KEGG) data-base (15), then the enrichment pathways were linked together to form an interaction network.

Marker expression evaluation and DEMs detection. Bioinformatics protocols were applied to analyze the sequencing data. Firstly, sequencing data were mapped to marker sequences using bowtie (16) with a maximum of two mismatches in the read, at most three mismatches in a whole read, and only the best alignment reported of each read, while the rest parameters were left as default. Secondly, expression levels of markers were measured in reads per kilobase of transcript per million mapped reads (RPKM) (17). Only markers with an average RPKM value of three experimental duplicates ≥1 were defined as expressed markers. Thirdly, the DEMs of the four samples were detected referring to the method developed by Audic and Claverie (18). The P-values were adjusted using the Benjamin and Hochberg correction (19) in the R programming language. According to the criteria, expressed markers with fold change ≥1.5 and false discovery rate-adjusted P-value ≤0.01 were defined as differentially-expressed. Spearman's coefficient was calculated to demonstrate the correlation between experimental duplicates (20).

Results

Expressed marker analysis. In the present study, a total of 563 markers were selected, all of which were documented to be associated with ovarian cancer. Therefore, it was expected that they would be expressed in the samples. The expression levels of the markers were validated by RPKM, which is a quantitative value with widespread use in RNA-seq analysis (21). As a result, 484 markers (~85.97%) were detected to be expressed in one or more samples. It was found that the RPKM values of the markers were similar in the four samples (Fig. 1A). The majority of the markers were expressed in all four samples (411 of 484; ~84.92%), with each sample also exhibiting specifically-expressed genes (Fig. 1B).

Certain genes in the same gene family have similar roles in metabolic processes, therefore it was inferred that they may be expressed synchronously in the process of transcription. To verify this hypothesis, all the known gene families found in UCSC were selected, and the 484 markers were categorized into 94 gene families. By comparing the RPKM values of these markers using cluster analysis (Fig. 2), it was identified that the three experimental replicates of each cell line can be clustered together. Samples of the two patients could be distinguished. However, samples of PEO14 and PEO23 could not be separated by the cluster analysis. This inferred that the differences between the two individuals had a greater effect on the expression of the marker than the differences between the tumor stages.

Pathway interaction network. A KEGG pathway interaction network was developed (Fig. 3). In the 484 expressed markers, 257 markers were detected to be enriched in 44 pathways, and 72 markers appeared in pathways associated with cancer (Fig. 3, pink). AKT2 (22,23), a putative oncogene, was enriched in the majority of the cancer-associated pathways in the network, including small cell lung cancer, non-small cell lung cancer, thyroid cancer and bladder cancer. Similarly, E2F2 and BRCA2 also appeared in several cancer-associated pathways. With the exception of cancer, pathways associated

<table>
<thead>
<tr>
<th>Sample</th>
<th>Read length (bp)</th>
<th>Patient ID</th>
<th>Stage</th>
<th>Raw bases</th>
<th>Bases after filtering</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEO1</td>
<td>42*2</td>
<td>2</td>
<td>First recurrence</td>
<td>1.3 G</td>
<td>1.2 G</td>
</tr>
<tr>
<td>PEO1</td>
<td>42*2</td>
<td>2</td>
<td>First recurrence</td>
<td>1.4 G</td>
<td>1.2 G</td>
</tr>
<tr>
<td>PEO1</td>
<td>42*2</td>
<td>2</td>
<td>First recurrence</td>
<td>1.7 G</td>
<td>1.5 G</td>
</tr>
<tr>
<td>PEO4</td>
<td>42*2</td>
<td>2</td>
<td>Second recurrence</td>
<td>1.6 G</td>
<td>1.3 G</td>
</tr>
<tr>
<td>PEO4</td>
<td>42*2</td>
<td>2</td>
<td>Second recurrence</td>
<td>1.5 G</td>
<td>1.3 G</td>
</tr>
<tr>
<td>PEO4</td>
<td>42*2</td>
<td>2</td>
<td>Second recurrence</td>
<td>1.4 G</td>
<td>1.3 G</td>
</tr>
<tr>
<td>PEO14</td>
<td>43*2</td>
<td>3</td>
<td>Presentation</td>
<td>1.2 G</td>
<td>1.1 G</td>
</tr>
<tr>
<td>PEO14</td>
<td>43*2</td>
<td>3</td>
<td>Presentation</td>
<td>933.6 M</td>
<td>860 M</td>
</tr>
<tr>
<td>PEO14</td>
<td>43*2</td>
<td>3</td>
<td>Presentation</td>
<td>1.3 G</td>
<td>1.2 G</td>
</tr>
<tr>
<td>PEO23</td>
<td>42*2</td>
<td>3</td>
<td>Recurrence</td>
<td>1.8 G</td>
<td>1.6 G</td>
</tr>
<tr>
<td>PEO23</td>
<td>42*2</td>
<td>3</td>
<td>Recurrence</td>
<td>1.7 G</td>
<td>1.5 G</td>
</tr>
<tr>
<td>PEO23</td>
<td>43*2</td>
<td>3</td>
<td>Recurrence</td>
<td>701.8 M</td>
<td>670 M</td>
</tr>
</tbody>
</table>
Markers associated with ovarian cancer were analyzed in the present study. To verify which markers were expressed in tumor cells, RNA-seq data analysis was applied. Of a total of 563 markers that were selected, 484 markers (~85.97%) were expressed in human ovarian tumor cells, which gave clear evidence that these are markers of ovarian cancer from the respect of the whole transcriptome. Differences between individuals are often omitted, but the present result demonstrated that expression levels of markers were more likely to be different between individuals than between tumors of different stages. In addition, more DEMs were detected between two different tumor stages of the same patient compared with between patients. Thus, in a future study, more samples from various ovarian cancer patients should be sequenced to prevent interindividual differences affecting the data.

In the expressed markers, the majority of the markers were enriched in pathways associated with cancer (72 markers), signaling (129 markers) and infection (35 markers). AKT2, a putative oncogene (24), was enriched in the majority of the cancer-associated pathways in the network and was first found to be aberrantly expressed in human ovarian cancer (25). Similarly, E2F2 and BRCA2 were also enriched in several cancer-associated pathways. The existence of pathways associated with infection in the pathway network once again with infection also appeared in the network with 35 markers. In the network, the p53 signaling pathway was linked with oocyte meiosis and progesterone-mediated oocyte maturation. The mTOR signaling pathway was also found to be associated with the progesterone-mediated oocyte maturation.

**DEMs analysis.** By comparing the changes of expression levels of markers from the presentation stage to the recurrence stage, it was found that certain markers had roles in recurrence in ovarian cancer. In the present study, the expression levels of markers between different patients with different tumor stages were compared. More DEMs were detected between two different tumor stages of the same patient, compared with between patients (Fig. 4). DEMs were detected in four pairs of comparisons, PEO14 vs. PEO23, PEO14 vs. PEO1, PEO14 vs. PEO4, and PEO4 vs. PEO4 vs. PEO1. As a result, 85 markers were found to be differentially-expressed.

**Discussion**

Markers associated with ovarian cancer were analyzed in the present study. To verify which markers were expressed in tumor cells, RNA-seq data analysis was applied. Of a total of 563 markers that were selected, 484 markers (~85.97%) were expressed in human ovarian tumor cells, which gave clear evidence that these are markers of ovarian cancer from the respect of the whole transcriptome. Differences between individuals are often omitted, but the present result demonstrated that expression levels of markers were more likely to be different between individuals than between tumors of different stages. In addition, more DEMs were detected between two different tumor stages of the same patient compared with between patients. Thus, in a future study, more samples from various ovarian cancer patients should be sequenced to prevent interindividual differences affecting the data.

In the expressed markers, the majority of the markers were enriched in pathways associated with cancer (72 markers), signaling (129 markers) and infection (35 markers). AKT2, a putative oncogene (24), was enriched in the majority of the cancer-associated pathways in the network and was first found to be aberrantly expressed in human ovarian cancer (25). Similarly, E2F2 and BRCA2 were also enriched in several cancer-associated pathways. The existence of pathways associated with infection in the pathway network once again
Figure 3. Pathways interaction network of expressed markers. Circle nodes stand for the pathways; lines between them show their associations. Pink words are pathways associated with cancer, blue indicates signaling pathways, yellow indicates pathways associated with infection, and the two red pathways are involved with oocytes.

Figure 4. DEMs of four groups of comparatives. Blue points indicate markers upregulated in samples shown on the x-axis, while the red points indicate the downregulated genes and the green points indicate those without significant changes. (A) DEMs between PEO1 and PEO4 samples; (B) DEMs between PEO14 and PEO23 samples; (C) DEMs between PEO1 and PEO14 samples; (D) DEMs between PEO4 and PEO14 samples. DEMs, differentially-expressed markers.
supported the view that external infections, such as bacteria and viruses may be associated with oncogenes in the induction of the occurrence of carcinomas (26-28). Signaling pathways also have important roles in cancer progression (29), and there were 129 detected markers in this type of pathway. Inactivation of p53 is implicated in tumor progression and numerous activated oncogenes included in the p53 signaling pathway have been elaborately investigated (30,31). In the network determined in the present study, the p53 signaling pathway was linked with oocyte meiosis and progesterone-mediated oocyte maturation (http://www.genome.jp/kegg/pathway/hsa/hsa04115.html). As has been established, the oocyte is the area where ovarian carcinoma formation is initiated, and unbalanced estrogen metabolism is hypothesized to be one of the causes of ovarian carcinoma (32). Thus, markers enriched in this pathway may have a role in the oncogenesis of ovarian cancer. The mTOR signaling pathway was also found to be associated with progesterone-mediated oocyte maturation. The deregulation of mTOR, a downstream effector of the AKT pathway, has been reported to have effects on tumor progression (33).

All the markers that were selected are associated with the presentation, development and diagnosis of ovarian cancer, according to the findings from associated studies. In total, 85 markers were abnormally expressed in tumor cells from patients who had recurrences. Of the 33 upregulated markers, **TMP** (34), **MAL2** (35), **ERCC** (36), **CD** (37), **KLK** (38) and **SCARA3** (39) were also observed to be over-expressed in previous studies. The **CD9** gene encodes a member of transmembrane 4 superfamily-tetraspanin family (40), which regulates cell surface glycoprotein function in differentiation and signal transduction. Notably, gene expression is involved in the suppression of cancer cell motility and metastasis (41). Although a study found that downregulation of **CD9** may be associated with the process of ovarian tumor dissemination (42), the present study demonstrated that **CD9** had a high expression level (RPKM) in all four samples (1290.7 in PEO14, 2141.98 in PEO23 with a significant increase at the recurrence stage). **AKTIP** may have a role in apoptosis; however, at present no studies have ruled out its potential target for taxol resistance in ovarian cancer. Clin Cancer Res 14: 3149-3155, 2008.

In conclusion, DEMs of ovarian cancer samples at different stages and the associations between them can be clearly investigated by examining the RNA-seq data. These markers may provide novel prospects for further studies on ovarian cancer.

**References**


