Survival-associated transcriptome analysis in ovarian cancer

Xiaofeng Xu¹,*, Xuan Zhou¹,*, Yijin Wang¹,², Tao Liu¹,³, Jian Fu⁴, Qian Yang⁵, Jun Wu¹, Huaijun Zhou¹

¹Department of Gynecology, Nanjing Drum Tower Hospital, Affiliated to Nanjing University Medical School, Nanjing, 210008
²Medical College, Southeast University, Nanjing, 210008
³Medical College, Nanjing University, Nanjing, 210008
⁴Department of Gynecology, Suqian People’s Hospital of Nanjing Drum Tower Hospital Group, Suqian, 223800
⁵Department of Gynecology and Obstetrics, The pukou Hospital of Nanjing, The Fourth Affiliated Hospital of Nanjing Medical University, Nanjing, 210031 (China)

Summary

Purpose of investigation: Ovarian Cancer (OC) is one of the most lethal gynecologic cancers worldwide. Despite the standard treatment, including radical resection, systemic chemotherapy, and targeted drugs for patients, survival rates remain low. This study provides new ideas for the diagnosis and treatment of Ovarian Cancer. Material and Methods: We performed Kaplan-Meier analysis on the transcriptome of Ovarian Cancer based on RNA-Seq data from The Cancer Genome Atlas (TCGA). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) enrichment were used for pathway and functional enrichment. Protein-protein interaction (PPI) network was constructed and visualized by STRING and Cytoscape. Results: A total of 1693 genes associated with survival were identified. The Kyoto Encyclopedia of Genes and Genomes pathway and Gene Ontology enrichment analysis revealed that these selected genes were differently enriched in numerous functional pathways. The top ten hub genes (RIPK4, HSPA8, FOS, STAT1, CD40LG, FGF2, RAC1, CXCR4, PRPF19, and CXCL10) were identified in our PPI network. Three highly connected cluster modules were differently enriched in several functional pathways. Conclusion: These key biomarkers in Ovarian Cancer may have diagnostic and therapeutic value in the future.

Key words: Ovarian Neoplasms; Survival Analysis; Transcriptome.

Background

Ovarian Cancer (OC) is one of the most lethal gynecologic cancers worldwide [1]. It was predicted that approximately 22,240 women will be diagnosed with ovarian cancer in 2018, and 14,070 will die from the disease [2]. Although the standard treatment, including radical resection, systemic chemotherapy, and targeted drugs for patients, produces a high response rate of 40-60%, less than half of women diagnosed with OC survive beyond 5 years [1]. However, there are few established molecular prognostic or predictive markers for this cancer type [3], and it is important to develop novel biomarkers for prognosis prediction of OC.

Accompanied by the arrival of the era of ‘big data’, high-throughput experiments such as gene expression microarrays have expanded our understanding of the underlying mechanisms between cancer development and genomic background [4]. The Cancer Genome Atlas (TCGA) is one of the biggest public databases for cancers, which contains transcriptome sequencing data from 131,019 samples of 43 kinds of cancer. With the help of bioinformatic technologies for microarray data, a lot of research on the abnormal regulation mechanisms in cancer has been conducted using the TCGA database. However, previous studies mainly focused on the relationship between a single gene and survival, such as Sprouty2 [5], IL-8 [6], to examine the potential etiology and prognosis of OC.

In the present study, we first identified a series of genes associated with patient survival in OC using the TCGA database. These selected genes were conducted for gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Furthermore, protein-protein interaction (PPI) networks and module analysis were also visualized using Cytoscape software to search for key genes that were related to survival and might be involved in the development of OC.

Material and Methods

Data resources

RNA-seq data for the TCGA-OV project was downloaded from TCGA official website (https://cancergenome.nih.gov). The corresponding clinical information was downloaded from http://www.cbioportal.org. A total of 374 ovarian serous cystadenocarcinoma samples with information of RNA profile and clinical data were included. Data were collated and extracted for further analysis.

Data processing and survival analysis

We applied the R package ‘survival’ to perform Kaplan-Meier analysis and assess survival difference across cases based on gene expression. Grouping depended on the median level of gene expression. These genes whose expres-
sion correlated with overall survival were identified. For the overall survival rates, the log-rank test was used to compare the significant differences in univariate analysis between subgroups. A p value of less than 0.05 was considered statistically significant.

Functional and pathway enrichment analysis of genes associated with survival

The Gene Ontology (GO) (http://www.geneontology.org) database can provide functional classification for genomic data, including categories of biological processes (BP), cellular component (CC), and molecular function (MF) [7]. The Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.ad.jp/kegg/) pathway analysis aims to identify and visualize significantly enriched pathways of molecular interactions, reactions, and relations [8]. To investigate the potential functions of the selected genes associated with survival, GO and KEGG pathway enrichment analyses using Database for Annotation, Visualization and Integrated Discovery (DAVID) were conducted (https://david.ncifcrf.gov/home.jsp) [9].

Protein-protein interaction (PPI) network building and interrelation analysis

The PPI network generated by STRING (Version 10.5, http://string-db.org/) was imported into Cytoscape software (version 3.60, http://www.cytoscape.org) to collect and integrate interactions between proteins [10,11]. Significant modules in the visible PPI network using the Molecular Complex Detection (MCODE) plugin were screened. The default parameters of MCODE were used: degree cutoff > 2, node score cutoff > 0.2, k-score > 2, maximum depth = 100 [12].

Results

Identification of genes associated with patients’ survival

Since survival rate is often used for evaluating cancer prognosis and the value of basic research on molecular mechanisms is ultimately assessed by clinical transformation, the study focused on the genes associated with survival. A total of 374 OC patients with information of RNA profile and clinical data were analyzed. We wrote the R language code for ‘survival’ package to calculate the p values and filter the genes (p < 0.05) associated with OS rate of patient groups divided by the median level of gene expression. A total of 1693 genes associated with survival were identified for further study (Supplementary Table 1). The survival curves for the first eight genes with the minimum p values are displayed in Figure 1.
Table 3. — *KEGG pathway and GO enrichment analysis of genes in module cluster three.*

| Category       | Pathway ID    | Pathway description                               | Count | False discovery rate |
|----------------|---------------|---------------------------------------------------|-------|----------------------|
| GOTERM_BP      | GO.0002696    | positive regulation of leukocyte activation       | 8     | 1.46E-08             |
|                | GO.0051249    | regulation of lymphocyte activation               | 8     | 8.40E-08             |
|                | GO.0032653    | regulation of interleukin-10 production            | 5     | 1.29E-07             |
| GOTERM_CC      | GO.0009897    | external side of plasma membrane                   | 7     | 1.34E-07             |
|                | GO.0009986    | cell surface                                       | 8     | 5.08E-06             |
|                | GO.0098552    | side of membrane                                   | 6     | 9.95E-05             |
| KEGG_PATHWAY   | hsa04514      | Cell adhesion molecules (CAMs)                     | 6     | 1.06E-07             |
|                | hsa04060      | Cytokine-cytokine receptor interaction             | 4     | 0.00298              |
|                | hsa04640      | Hematopoietic cell lineage                         | 3     | 0.00298              |

Figure 1. — A total of 1693 genes associated with survival were identified. The survival curves for the first eight genes with the minimum P values were shown. (A) BATF2, (B) CD38, (C) NLRP12, (D) RXFP1, (E) LAMTOR5-AS, (F) ZNF561, (G) ITGAD, (H) SLC22A2.

**GO and KEGG pathway enrichment analysis of genes associated with survival**

To identify potential functions of genes associated with survival, the selected 1693 genes to DAVID for KEGG pathways and GO enrichment analysis were uploaded. The top 5 mostly enriched KEGG pathways of these genes were pathways in cancer, cytokine receptor interaction, axon guidance, cell adhesion molecules, and osteoclast differentiation (Figure 2A). GO analysis revealed that these selected genes were mostly enriched in DNA-templated transcription, signal transduction, positive regulation of transcription from RNA polymerase II promoter, negative regulation of transcription from RNA polymerase II promoter, DNA-templated positive regulation of transcription for BP term, extracellular space, cell surface, proteinaceous extracellular matrix, extracellular matrix, axon for CC term and transcription factor activity of sequence-specific DNA binding, sequence-specific DNA binding, identical protein binding, RNA polymerase II core promoter proximal region sequence-specific DNA binding, protein kinase binding for MF term. (Figure 2B-D). The most significantly enriched pathways and enrichment terms are displayed in Supplementary Table 2.

**PPI network modeling**

Given a list of known proteins, the STRING database can identify their interactions and predict proteins with assigned confidence scores [10]. The selected 1693 genes were uploaded to the STRING website to perform PPI network. As a result, a total of 1,240 nodes and 3,365 edges were established in our PPI network, with an average node degree of 5.41 and an average local clustering coefficient of 0.35. The top ten hub genes with the
highest degree were receptor interacting serine/threonine kinase 4 (RIPK4), heat shock 70 kDa protein 8 (HSPA8), fos proto-oncogene (FOS), signal transducer and activator of transcription 1 (STAT1), CD40 ligand (CD40LG), basic fibroblast growth factor (FGF2), Ras-related C3 botulinum toxin substrate 1 (RAC1), C-X-C chemokine receptor type 4 (CXCR4), pre-mRNA-processing factor 19 (PRPF19), and C-X-C motif chemokine 10 (CXCL10) (Supplementary Figure 1). Figure 3 displays the survival curves of these hub genes in OC.

Furthermore, we used MCODE algorithm to establish subnets of the PPI network that are likely to represent molecular complexes. The top three highly connected cluster modules are displayed in Figure 4. The functional annotations of the genes involved in the modules were determined using STRING. The most significantly enriched GO terms were SCF ubiquitin ligase complex, cellular metal ion homeostasis, G-protein coupled receptor signaling pathway for the module cluster one (Table 1), mRNA splicing via spliceosome, RNA splicing, mRNA processing for the module cluster two (Table 2), and positive regulation of leukocyte activation, regulation of lymphocyte activation, regulation of interleukin-10 production for the module cluster three (Table 3). KEGG pathway enrichment analyses showed that neuroactive ligand-receptor interaction, spliceosome and cell adhesion molecules (CAMs) were the most enriched pathways for the module cluster one, two and three, respectively.

Discussion

Though there have been advances in current therapeutics, OC remains one of the most deadly cancers. Researchers are exploring significant prognostic biomarkers for OC, in order to provide early intervention. Numerous relevant genes associated with OC survival have been demonstrated previously. Sprouty 2 expression was revealed to significantly impact tumor behavior with predictive value as an independent prognostic factor for survival and recurrence [5]. KIF20A overexpression predicted unfavorable clinical outcome and had potential to serve as a useful prognostic biomarker for epithelial OC patients [13]. The high preoperative serum level of Creatinine was associated with poor survival, which indicated that it might be an additional independent prognostic parameter in patients with epithelial OC [14]. Nucleus and/or cytoplasm of β-catenin expression might be associated with tumor progression and subsequently represent a predictive factor of poor prognosis in OC patients [15]. Wilms’ tumor 1 (WT1), an antigen target, was a biomarker for poor prognosis, particularly when combined with altered p53 in OC [16]. Penzvalto Z. et al. (2014) reported that Mitogen activated protein kinase/extracellular signal regulated kinase 1 (MEK1) represented a promising candidate prognostic biomarker and correlated with response rates to platinum-based chemotherapy in OC [17]. However, most studies primarily focused on a single gene as a potential predictive biomarker.
Survival-associated transcriptome analysis in ovarian cancer

In this study, for the first time, we performed Kaplan-Meier analysis for all the genes in OC cases from TCGA project using bioinformation technology. As a result, a total of 1693 genes associated with survival were identified. KEGG pathway and GO enrichment analysis were then performed. These selected genes were differently enriched in numerous functional pathways. In addition, we carried out the PPI network construction and modeling through Cytoscape. The top three highly connected cluster modules were also enriched in several functional pathways.

As most medical bioinformation studies focused on differently expressed genes between tumor and normal samples, we firstly paid attention to all the genes associated with survival. Thus, for the first time, we identified a series of genes related to survival in OC, which would help us comprehensively understand the molecular mechanisms of OC. During the PPI network construction, the top ten hub genes (RIPK4, HSPA8, FOS, STAT1, CD40LG, FGF2, RAC1, CXCR4, PRPF19, and CXCL10) were demonstrated. RIPK4 has been reported to be aberrantly expressed in several cancer types. It was overexpressed in human ovarian adenocarcinomas compared to noncancerous ovarian tissue samples (Ref). Though interacting with the adaptor protein DVL2, RIPK4 could stimulate Wnt signaling pathway with the co-receptor LR6, which suggests that RIPK4 overexpression may contribute to the growth of OC [18]. Gong et al. (2018) revealed that the expression of RIPK4 was up-regulated in nasopharyngeal carcinoma (NPC) tissues and it could promote the growth and anchorage-independent growth of NPC cells, which demonstrated the oncogenic roles of RIPK4 in NPC and suggested that RIPK4 might be a therapeutic target [19]. High expression of RIPK4 was also observed in bladder urothelial carcinoma tissues and was an independent predictor for poor overall survival [20]. HSPA8 could negatively regulate MLK4 and MLK3, which revealed an important function for MLK4 in modulating MLK3 activity in hot stress responses in OC cells [21]. In pancreatic cancer cells, HSPA8 overexpression was able to enhance cell viability, diminishing the effects of Maslinic acid [22]. c-FOS overexpression could inhibit OC growth by changing adhesion both in vivo and vitro [23]. In metastases, c-Fos and Fos B expression were significantly lower than in the respective primary OC tissues, which also indicated that FOS might be a potential cancer suppressor gene [24]. Trinh B. revealed that STAT1 could interact with DLX4 and promote ovarian tumor angiogenesis in part by stimulating iNOS expression [25]. STAT1 was found overexpressed in patients with high-grade serous OC after the development of clinical platinum resistance. Further knockdown of STAT1 significantly enhanced apoptotic response to platinum treatment in resistant OC cells, which suggests that STAT1 might take part in platinum resistance in OC [26]. Slattery ML. et al. (20xx) demonstrated that CD40LG was one of the genes that could be modified by diet and lifestyle factors, which appeared to be important mediators in breast cancer risk [27], FGF2 down-regulated E-cadherin expression through the activation of PI3K/Akt/mTOR and MAPK/ERK signaling pathways in human OC cells [28]. A prognostic value of serum FGF2 was also determined in OC [29]. RAC1 took part in angiogenesis in OC [30], and its overexpress-
sion was associated with cell epithelial-mesenchymal transition and poor OC prognosis [31]. CXCR4 could promote cell proliferation, migration, invasion, and metastasis of OC [32,33]. Overexpression of CXCR4 was significantly associated with cisplatin-based chemotherapy resistance and it could be a prognostic factor in epithelial OC [34]. PRPF19 was one of the seven genes with different copy number alterations between matched highly and minimally invasive/migratory OC cell subclones, which could be specifically targeted for the treatment and prognosis of advanced OC [35]. High expression of CXCL10 was associated with almost doubled overall survival and it was confirmed as an independent validation set, which supported the notion that CXCL10 exerted a tumor-suppressive function in OC [36].

Conclusion

The present study is the first to identify a series of genes associated with survival in OC. KEGG pathway and GO enrichment analysis revealed that these selected genes were differently enriched in numerous functional pathways. In addition, we carried out the PPI network construction and modeling and the top ten hub genes (RIPK4, HSPA8, FOS, STAT1, CD40LG, FGF2, RAC1, CXCR4, PRPF19, and CXCL10) were identified. These findings may have diagnostic and therapeutic value in the future. However, more research is necessary to clarify the roles of these genes.

Author contributions

X. X. and H. Z. designed the study. X. X. and X. Z. prepared the manuscript. X. X., X. Z. and H.Z. checked the data. Y. W., T. L. and J. F. performed data collection. Q. Y. and J. W. performed statistical analysis. H. Z. conceived and supervised the project.

Acknowledgements

We thank Juan Chen from CCN department of ZTE for her computer technology support. This work was supported by a grant from the Natural Science Foundation of Jiangsu Province (Ref: BK20151096) and the Key Projects of National Health and Family Planning Commission of Nanjing City under Grant (Ref: ZKX17015).

Conflicts of interest

The authors have no conflict of interests to declare.

Submitted: January 11, 2019
Accepted: April 09, 2019
Published: June 15, 2020

References

[1] Smith R.A., Andrews K.S., Brooks D., Fedewa S.A., Manassaram-Baptiste D., Saslow D., et al.: “Cancer screening in the United States, 2018: A review of current American Cancer Society guidelines and current issues in cancer screening”. CA. Cancer J. Clin., 2018, 68, 297.

[2] Siegel R.L., Miller K.D., Jemal A.: “Cancer statistics, 2018”. CA. Cancer J. Clin., 2018, 68, 7.

[3] Oliveira-Ferrer L., Goswami R., Galatenko V., Ding Y., Eylmann K., Legler K., et al.: “Prognostic Impact of CEACAM1 in Node-Negative Ovarian Cancer Patients”. Dis. Markers, 2018, 2018, 1.

[4] Zhao H., Li H.: “Network-based meta-analysis in the identification of biomarkers for papillary thyroid cancer”. Gene., 2018, 661, 160.

[5] Masoumi-Moghaddam S., Amini A., Wei A.Q., Robertson G., Mor-
Hein S., Mahner S., Kanowski C., Löning T., Jänicke F., Milde-Stronach E.A., Alfraidi A., Rana N., Datler C., Studd J.B., Agarwal Gong Y., Luo X., Yang J., Jiang Q., Liu Z.: “RIPK4 promoted the Guo Q., Gao B.L., Zhang X.J., Liu G.C., Xu F., Fan Q.Y., Ogata H., Goto S., Sato K., Fujibuchi W., Bono H., Kanehisa M.: “The homeo- Long H., Li G., Wen X., Lv Y., Zhang M., Wei Y., Huang X., McGann J.C., Liu B.Y., Hannoush R.N., Lill J.R., Pham Bronger H., Singer J., Windmüller C., Reuning U., Zech D., Del-Gonzalez-Villasana V., Fuentes-Mattei E., Ivan C., Dalton Blessing N.A., Kasturirangan S., Zink E.M., Schroyer A.L., Chadee Saito R., Smoot M.E., Ono K., Ruscheinski J., Wang P. L, Lotia S, Pénzváltó Z., Lánczyk A., Lénárt J., Meggyesházi N., Krenács T., Gil M., Komorowski M.P., Seshadri M., Rokita H., McGray A.J., Szklarczyk D., Morris J.H., Cook H., Kuhn M., Wyder S., Si- Huang D.W., Sherman B.T., Tan Q., Kir J., Liu D., Bryant D., et al.: “Prognostic Value of Serum Creati- sthaler L., Reiser E., et al: “Prognostic Value of Serum Creati- liu J.Y., Zeng Q.H., Cao P.G., Xie D., Chen X., Yang F., Lafleur J., Hefler-Frischmuth K., Grimm C., Schwameis R., Gen- liu J., Zhang W., Sun X., Chen J., Li Y., Niu C., et al.: “Overexpression of kinesin family member 20A is associated with unfavorable clinical outcome and tumor progression in epithelial ovarian cancer”. Cancer Manag. Res., 2015, 10, 5327. 177. 1337. 363. 661. 239. 137. 104. 2127. 193. 5327. 2014, 45, 362. 578. 2014, 2498. 2014, 28. 2012, 33. 2015, 5327. 2014, 1536. 2011, 59083. 670. 2013, 273. 2015, 18, 5127. 2018, 1069. 2015, 5327. 2012, 169. 2015, 22, 47. 2015, 1, 2320. 2014, 47, 33. 2014, 28, 32. 2012, 19, 19. 2011, 5, 20. 2010, 10, 3433. 2012, 3, 129. 2010, 8, 5127. 2012, 9, 1069. 2012, 33. 2015, 137. 2018, 119, 462. 2012, 35. 2013, 139, 1441. 2018, 208, 1. 2016, 14, 837. 2018, 5, 1. 2018, 118, 1617. 2017, 257. 2015, 21, 2127. 2018, 2014, 193, 5327. 2015, 22, 247. 2014, 45, 362. 2012, 18, 5127. 2015, 8, 59083. 2017, 770. 2015, 12, 35. 2015, 179. 2018, 110, 753. 2017, 2127. 2015, 21, 2127. 2014, 2127. 2013, 169. 2015, 22, 177. 2015, 5, 1. 2013, 29. 2012, 208, 1. 2015, 193, 5327. 2014, 45, 362. 2017, 21, 2127. 2013, 169. 2015, 193, 5327. 2014, 47, 33. 2014, 47, 33. 2014, 47, 33. 2012, 19, 19. 2012, 19, 19. 2012, 19, 19. 2012, 19, 19.