PD-L1 and the Clinical Outcomes of Ovarian Cancer: Meta-Analysis and Bioinformatical Analysis

Xiaoling Shi1, Liu Yun2, Liu Chaoqun Liu3, Hui Ding4, Dan Liang1, Fang Geng5, Haiying Yu1, Jinxiu Ban1, Jiajing Li1, Wentao Gui1, Yi Sun1,5*

Abstract

Objective: A meta-analysis was performed to analyze the association between PD-L1 expression and overall survival (OS) in various tumors and to identify potential targets through biological information analysis. Methods: the data were collected from PubMed and Cochrane library, the all analysis of our study were conducted by STATA software and online website. Results: Ten articles (including 11 studies) that met all inclusion criteria were obtained. The combined HR showed that high PD-L1 expression was significantly associated with poor overall survival (HR = 1.84, 95% CI: 1.15-2.93). Pathway analysis revealed that the upregulated genes were primarily involved in biological processes, including nucleic acid transcription, biosynthesis and negative regulation of cell metabolism. The downregulated genes were primarily involved in the regulation of cell cycle, including chromosome separation and DNA metabolism. The top ten genes that were identified were hub genes (CDK1, CCNB1, CCNA2, KIF11, CDC20, UBE2C, NCAPG, AURKA, AURKB, CHEK1), which had significant function in cell differentiation and virus infection. The Kaplan-Meier survival curve indicated that CCNB1, KIF11, UBE2C, NCAPG, AURKA and CHEK1 were statistically significant (P<0.05). Conclusion: PD-L1 was found to be a latent biomarker for predicting the prognostic value of cancer and also a therapeutic target.

Keywords: PD-L1- cancer- prognosis role- meta-analysis- bioinformatics analysis

PD-L1, also known as CD274 or B7-H1, is a costimulatory glycoprotein belonging to the B7 family. It is expressed by macrophage lineage cells and is considered a potential regulator of anti-tumor immunity in various types of cancers in humans. It can induce apoptosis of T lymphocyte by binding to the PD-1 receptor on activated T cells (Ghebeh et al., 2006; Kim et al., 2015), thereby inhibiting anti-tumor immune response (Song et al., 2014). In the clinic, the PD-L1/PD1 immune checkpoint is blocked by using anti-PD1 or anti-PD-L1 antibodies to relieve tumor-induced immunosuppression (Bregar et al., 2017). By doing so, the immune system is reactivated, and capable of attacking tumor cells (Ali et al., 2015; Katsuya et al., 2015; Masugi et al., 2017). Immunological checkpoint inhibitors for PD-L1/PD1 have been tested in several clinical trials, including Nivolumab, Pembrolizumab, Atezolizumab, Avelumab and Durvalumab. Most efficacy of checkpoint blockade therapy has been seen in solid tumors such as melanoma, small cell lung cancer and kidney cancer. However, only 10% to 30% of the population was found to be sensitive to immunological checkpoint inhibitor treatment. There are some types of tumors, such as prostate cancer, and few individuals were sensitive to treatment with immunological checkpoint inhibitors. Some studies about prostate cancer have shown that the tumor cells play a role in inhibiting anti-tumor immunity systemic release of exosomes containing PD-L1 in the blood stream. As a result, checkpoint inhibitors were not effective in some tumor patients (Delaunay et al., 2018; Zhou et al., 2017).

Ovarian cancer (OV) is a common tumor in women and ranks third in the list of malignant tumors. Due to lack of effective screening methods, patients are diagnosed with advanced disease, with very low 5-years survival rate (Chatterjee et al., 2017; Drakes et al., 2018; Ojalvo et al., 2018). In recent years, bioinformatics analysis based on gene expression array has become an effective new method for discerning new genes and
understanding the underlying molecular mechanisms of cancer. Bioinformatic analyses play an important role in studying the gene expression profile of ovarian cancer. With this analyses, the prognostic factors associated with poor cancer prognosis can then be identified, and used to stratify patients as low-risk and high-risk patients. Thus, a treatment plan can be initiated on time, which can help prolong the survival period. Overall, the discovery of prognostic factors can accurately predict clinical outcomes, identify new predictors and pave way for the discovery of new therapeutic targets. Therefore, we performed meta-analysis and bioinformatics analysis to elucidate the relationship between PD-L1 expression and overall survival, as well as therapeutic targets.

Materials and Methods

Meta-analysis

The meta-analysis was based on observational epidemiological studies and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) (Liberati et al., 2009).

Search methods and inclusion criteria

In this paper, PubMed and Cochrane library were used to search for studies, using the following keywords: (PD-L1 OR CD274 OR B7-H1) AND Cancer. The search was completed on Sept 11th, 2018, and the search was not restricted by language.

Following inclusion criteria was used: (I) the articles reported cancer prognosis; (II) they measured PD-L1 expression; (III) they reported the hazard ratio (HR) for overall survival; (IV) they reported the association between PD-L1 expression and overall survival (OS); (V) the sample size of the study was greater than twenty; (VI) the study reported clinical results.

Definition and data extraction

Overall survival was defined as the interval between the medical treatment and the death of the patients or the last observation.

We adopted the same methods to extract data and summarize the data to control bias. The combined hazard ratio (HR) and 95% confidence intervals (95% CI) were used as the effective measures to analyze survival outcomes. At the same time data, including the first author, year of publication, cancer, cancer type (system) and stage, number of patients, age, the method of detecting PD-L1 and the follow-up time, were also extracted from the studies.

Statistical analysis

Firstly, in order to evaluate the impact of the PD-L1 status on tumor prognosis, we extracted the HR and their 95% CI. HR > 1 indicated a poor survival in PD-L1 high expression group. When the 95%CI for the combined HR did not overlap with 1, the negative impact of PD-L1 on survival was deemed statistically significant. Then, we used I² statistics to assess heterogeneity among studies. I² value range was 0-100%, with higher I² value indicating that the impact of inter-heterogeneity was greater in meta-analysis (Higgins et al., 2003). The results will be considered statistically significant if the P<0.05 and I² value were used as the measurement standards. I² <25% indicated no heterogeneity. I² = 25-50% indicated moderate heterogeneity. Similarly I²> 50% indicated strong heterogeneity (Higgins et al., 2003). Random-effects model was adopted for analysis when heterogeneity was found among studies, whereas the fixed-effects model was adopted when data was homogenous. In addition, meta-regression analysis and sub-group analyses were performed to identify potential sources of heterogeneity. Finally, the Begg’s funnel plot was used to analyze any existing publication bias in our research. Finally, P <0.05 was considered statistically significant. All analyses were conducted by using the 12.0 ATATA software.

Bioinformatical analysis

Microarray data and expression analysis of DEGs

GSE39204 gene expression profiles dataset was downloaded from the GEO database, which contains public functional genomics datasets. The platform for GSE39204 was GPL570 (HG-U133_Plus_2) Affymetrix Human Genome U133 Plus 2.0 Array) and it was submitted by Kaoru Abiko. The database contains a total of 66 ovarian cancer samples, including 31 samples with PD-L1 positive expression and 33 samples with PD-L1 negative expression. In this study, the samples were matched according to the stage of the tumor. Four positive samples were selected, including GSM 958092, GSM 958106, GSM 958107 and GSM 958114. And four negative samples were also selected, including GSM958059, GSM 958062, GSM 968064 and GSM 958065. Through the GEO 2R online tool analysis, we found all genes with differential PD-L1 expression. Finally, we found 479 statistically significantly different genes with |log FC|≥1 and P<0.05 as the standards.

Gene ontology and pathway enrichment analysis of DEGs

David (https://david.ncifcrf.gov/) has a set of annotation tools for investigators to use. The tools are usually used to perform GO and KEGG analysis to screen for different genes. The differential genes obtained via David online tool analysis with P-value < 0.05 were considered to be statistically significant.

Protein-Protein Interaction (PPI) Network

DEGs of PPI were gathered from STRING, which is a tool for analyzing protein-protein interaction. We used STRING to analyze all differential genes and adopted score >0.4 as the standard. The software Cytoscape was used to form PPI networks, using the plug-in MCODE of cytoscape to select for the two modules of the expression. Finally, we used the plugin CytoHubba to identify the top ten hub genes.

Survival analysis

Kaplan-Meier plotter was used to evaluate the impact of genes or gene combination on survival in breast, ovarian, lung, gastric, colon, prostate, GBM, LGG, melanoma, DLBCL, RCC, AML, and 14 other cancer
with overall survival, we performed subgroup analysis and meta-regression to study the cause of the resulting heterogeneity, including year of publication, country, number of patients, age, cancer category and stage, method (Table 2). Finally, we found that the number of patients (Adj $R^2 = 50.73\%$), the cancer type (Adj $R^2 = 76.39\%$), and the detection method (Adj $R^2 = 77.33\%$) might be the sources of heterogeneity. Finally, the results showed that the combined HR and CI were 1.84(1.15-2.93) and I² values were 72.1%.

**Sensitivity analysis and publication bias**

Through sensitivity analysis, we found that the outcome of sensitivity analysis was more stable when one case was discarded (Inaguma et al., 2017). In order to detect publication bias in the meta-analysis, the Begg’s funnel plot method was applied. The results showed that no bias was found from these studies. The funnel plot result is as follows: $P=0.599$ (Figure 3)

**Bioinformatical analysis**

**Identification of differential genes in ovarian cancer**

We analyzed 4 PD-L1 positive samples and 4 PD-L1 negative samples. All microarray data in the samples were also analyzed using the GEO 2R online software, and differential genes were obtained with $P < 0.05$ and $|\log FC| \geq 1$ as the cut-off criterion. As a result, 479 DEGs were obtained, which contained 202 upregulated genes and 277 downregulated genes in the analysis of GSE39204 (Table 3). The heat map of differential gene expression is presented in Figure 4.

**GO and KEGG analysis in ovarian cancer**

GO analysis showed that the expression of differential genes was mainly associated with the regulation of cell cycle process. The upregulated genes were primarily involved in biological processes, including nucleic acid transcription, biosynthesis, and negative regulation of cellular metabolism. On the other hand, the downregulated genes were primarily involved in cell cycle regulation, including chromosome segregation and DNA metabolism (Table 4 and Figure 5). By KEGG PATHWAY enrichment analysis, the upregulated genes were mainly enriched in cancer pathways and rheumatoid arthritis, whereas the downregulated genes were mainly enriched in cell cycle, viral carcinogenesis, and P53 signaling pathway (Table 5 and Figure 6).

**Network module analysis of protein-protein interaction for DEGs**

We input DEGs into STRING to predict the interaction between proteins, and construct PPI network by Cytoscape. Then, we analyzed it through the plug-in MCODE in the cytoscape software and selected the two most important modules (Figure 7). We also performed a function annotation analysis of genes in these two important modules. The results indicated that module A was mainly related to cell cycle and viral carcinogenesis, whereas module B was related to protein catabolism regulation. The top ten hub genes of the node were identified by using the plug-in CytoHubba, including CDK1, CCNB1, CDK2, CDK4, CDKN1A, CCNB2, CDK6, MAPK1, CDKN1C, and MAPK3.
Survival analysis

Further analysis by the Kaplan-Meier survival curve indicated that CCNB1, KIF11, UBE2C, NCAPG, AURKA, AURKB, CHEK1 were statistically related to the survival time of ovarian cancer (P < 0.05). The result indicated that the high expression of CCNB1, UBE2C and AURKA has a significant impact on the prognosis of ovarian cancer.

Discussion

In this study, we conducted a meta-analysis to study an association between PD-L1 expression and overall survival (OS) in various tumors. The combined HR showed that high PD-L1 expression was significantly associated with poor overall survival (HR = 1.84, 95% CI: 1.15-2.93), and a high heterogeneity exponent (I² = 72.1%, P≤ 0.001). These findings suggest that PD-L1 is a considerable prognostic factor for poor survival in ovarian, renal, thyroid and colorectal cancer patients. Then, we used biological information analysis to analyze ovarian cancer tissue samples, including PD-L1 high expressing and low expressing samples. As a result, 479 different genes were obtained, which contained 202 upregulated genes and 277 downregulated genes. Among them, the high expression of CCNB1, UBE2C and AURKA in the prognosis of ovarian cancer has been shown to be significant. Therefore, these proteins may be important prognostic factors and potential therapeutic targets.

When we analyzed the studies by group analysis and meta-regression, no study had addressed the heterogeneity. Therefore, the highly significant heterogeneity shown in the results could be due to inclusion of patients at different baseline characteristics (year, number of samples, cancer type, the method for detecting PD-L1 and the follow-up time). It is also possible that the reported methods used in the study were different, which could lead to differences between studies. Furthermore, there is no standard for controlling confounding factors across the studies. These problems may have contributed to error in the HR assessment. Therefore, research bias and significant heterogeneity may be resulted from several factors.

Besides, for many studies, the expression of PD-L1 was mostly depended on immunohistochemistry, and there is no unified detection method for quantitative assessment of PD-L1 expression at present (Inaguma et al., 2017; Katsuya et al., 2015; Mansfield et al., 2014). For the multiple samples with positive expression of PD-L1, the associated detection method has been developed. But these tests have not been compared, standardized or prospectively validated. Therefore, it is difficult to determine whether PD-L1 positive expression in tumors is a consistent finding. In addition, there is also no clear standard to use as the cut-off value for PD-L1 expression. Similarly, different detection antibodies have different cut-off values and the researcher often define cut-off values based on experimental conditions.

These technical variations may affect the result and may affect the prognostic value PD-L1 expression, suggesting that cut-off value may also be an important source of heterogeneity in the assessment of PD-L1 expression. Therefore, there is need for establishment of a new research standard to compare the results between different studies, where PD-L1 can be used as a useful predictive biomarker.

The Begg's test showed the P value of 0.599, greater than 0.1, and therefore was not statistically significant, suggesting that no publication bias existed between the included studies. However, funnel plots showed small asymmetries, suggesting that there may have been some epidemiological biases. In order to decrease publication bias, we retrieved the literature by using PubMed and Cochrane library and without limiting language. However, there may be some limitations due to the small number of contained articles and the small sample capacity. In addition, unpublished studies that may contain empty results could not be retrieved.

In bioinformatics analysis, GO analysis showed that the expression of differential genes was mainly related to the regulation of cell cycle process. The signaling pathway and KEGG PATHWAY enrichment analysis resulted showed enrichment of the cell cycle process. PPI analysis indicated that module A was mainly related to cell cycle and viral carcinogenesis, whereas module B was related to regulation of protein catalysis. The two modules were found to be statistically significant, with MCODE score> 10 and FDR <0.05. The top ten hub genes of the node were identified to be CDK1, CCNB1, CCNA2, KIF11, CDC20, UBE2C, NCAPG, AURKA, AURKB and CHEK. CDK1 regulates centrosome circulation and mitotic onset and plays an essential function in controlling the eukaryotic cell cycle. CCNB1 controls the cell cycle in G2/M (mitotic) transition and belongs to the cyclin family. CCNA2 functions with cyclin-dependent protein kinase CDK1 or CDK2 to form a specific serine / threonine protein kinase holoenzyme complex. KIF11 is a kinesin required to build bipolar spindle during mitosis. CDC20 is required for promoting protein ligase activity of the complex/circle (APC/C) at a later stage. UBE2C accepts proteins from the E1 complex and catalyzes its covalent attachment to other proteins. NCAPG is a mitotic serine / threonine kinase that helps regulate cell cycle progression. AURKA has a basic function in centromeres to ensure correct alignment and separation of chromosomes and is necessary for chromatin-induced microtubule stabilization and spindle assembly. AURKB is a serine / threonine-protein kinase component of the chromosomal complex (CPC), a complex that acts as a key regulator of mitosis. CHEK1 is a serine / threonine-protein kinase that is required for checkpoint-mediated cell cycle arrest and DNA damage or for DNA replication to activate DNA repair.

Our research also has some shortcomings: Firstly, the number of included studies were inadequate in our analyses, with only 10 articles including 11 studies. As observed above, a smaller sample size is more likely to generate heterogeneity. Secondly, there is a retrospective
accumulation of the baseline characteristics of patients, which may be subject to recall bias. Similarly, data can be lost in the process of follow-up. Compared with randomized controlled trials, retrospective study reduced the accuracy and may also affect the true effect of PD-L1 expression on tumor prognosis. Besides, all data we analyze are from published studies, not the entire data for each patient. Published data tend to have positive effects and the negative data are rarely published. The possibility of confusion between demographics and clinical factors in the study was limited. Finally, all the studies were only tumor patients without healthy population controls, no comparison, so it was impossible to analyze whether PD-L1 expression existed in healthy population.

Immune checkpoint is a kind of immunosuppressive signal molecule in normal body, which plays an immunomodulatory role by regulating the balance of co-stimulatory and co-inhibitory signals (Inman et al., 2007; Kim et al., 2015; Nakanishi et al., 2006). Co-stimulatory and co-inhibitory signals played a significant value in maintaining tolerance, modulating the amplitude and duration of T cell responses. This has potential value for developing new treatment regimens (Inaguma et al., 2017). In addition, the inhibitory effect of PD-L1 exceeds the stimulatory effect that of costimulatory molecules, causing dampening of T cell activation, and disease status progression. Some studies have shown that PD-L1 is correlated with patient survival, therefore is a potential prognostic biomarker for human malignancies (Ghebeh et al., 2006; Katsuya et al., 2015; Loos et al., 2011).

In conclusion, the meta-analysis results show that compared with no or low expression of PD-L1, high expression of PD-L1 could reduce disease-free survival time and had poor long-term prognosis after tumor surgery. Therefore, PD-L1 could be a latent biomarker for predicting the prognostic value of cancer and could also be a therapeutic target (Ritprajak and Azuma, 2015; Zhu et al., 2017). Overall, PD-L1 as a target for tumor treatment may be a promising therapeutic strategy (Loos et al., 2011; Page et al., 2014). It is crucial to adopt standard methods to quantify PD-L1 expression to provide a prognostic valuable threshold for treatment, which in turn will help make better decisions for the treatment of patients with PD-L1 overexpression (Yang et al., 2013).

Author Contribution Statement

Methodology, Yun Liu; Validation, Chaoqun Liu; Formal Analysis, Hui Ding; Investigation, Fang Geng; Resources, Haian Yu; Data Curation, Dan Liang; Writing – Original Draft Preparation, Xiaoling Shi and Yi Sun; Writing – Review & Editing, Xiaoling Shi and Yi Sun; Visualization, Jinxiu Ban; Supervision, Jiajing Li; Project Administration, Wentao Gui; Funding Acquisition, Yi Sun.

Acknowledgments

This work was in part supported by Guangxi Science and Technology Base and Talent Special Project (AD19245003), Natural Science Foundation of Guangdong Province (No. 2018030310412) and National Guangxi Collège Students Innovation and Entrepreneurship Training Program (No. S202110601105, 202110601031).

Conflicts of Interest

The authors declare that they have no competing interests for this research.

References


