

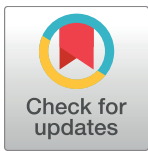
RESEARCH ARTICLE

Transcriptome-based signature predicts the effect of taxol in serous ovarian cancer

Shunyu Hou, Jianrong Dai*

The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, Jiangsu, P.R. China

* daijianrong2018@163.com



Abstract

Taxol is a widely used chemotherapy drug used clinically for ovarian cancer, although the response to Taxol among individuals varies due to the heterogeneity among ovarian cancer patients. In this work, we analyzed differences in the prognostic effect of gene expression and Taxol usage in the Cancer Genome Atlas (TCGA) dataset and identified specific genes associated with the Taxol effect. Using the Cox regression model, a risk model (Taxol score) was developed to assess the outcome of ovarian cancer patients who underwent chemotherapy with Taxol. According to the results, survival was significantly associated with the Taxol score. Moreover, the patients in the high and low Taxol score group had different responses to Taxol. This result was further validated in another two independent datasets. The correlation between clinicopathological indicators was also analyzed, and we determined that the Taxol score is not associated with age, pathological stage, or Taxol treatment, while there was significant correlation with tumor size and grade. Gene Set Enrichment Analysis (GSEA) showed that various signaling pathways including ECM receptor, drug metabolism and ascorbate metabolism pathways were significantly enriched in the high Taxol score group. Collectively, these results indicate that the model is robust for predicting the effectiveness of Taxol by reflecting the various cell statuses of serous ovarian carcinoma.

OPEN ACCESS

Citation: Hou S, Dai J (2018) Transcriptome-based signature predicts the effect of taxol in serous ovarian cancer. PLoS ONE 13(3): e0192812. <https://doi.org/10.1371/journal.pone.0192812>

Editor: Aamir Ahmad, University of South Alabama Mitchell Cancer Institute, UNITED STATES

Received: December 5, 2017

Accepted: January 30, 2018

Published: March 1, 2018

Copyright: © 2018 Hou, Dai. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The third-party data underlying this study are publicly available from the following URLs: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63885> <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE9891> <https://xenabrowser.net/datapages/?dataset=TCGA.OV.sampleMap/HiSeqV2&host=https://tcga.xenahubs.net>.

Funding: This work was supported by the Special Funds for Industrial Technology Innovation of Suzhou SYS201567 to Jianrong Dai, Clinical Medical Centers of Suzhou SZZX201505 to Jianrong Dai, Introduce Project of Clinical Medicine

Introduction

Ovarian carcinoma is one of the most common causes of cancer related death [1]. It was estimated that there were 52,100 new cases and 22,500 deaths in 2015 according to a recent cancer statistic in China [2]. Among the subtypes of ovarian cancer, serous ovarian carcinoma comprises the majority of the cases. Since radiation is not suitable for serous ovarian carcinoma therapy and targeted drugs have not been widely used, chemotherapy is currently the major treatment method for serous ovarian carcinoma after surgery. Among the drugs used for chemotherapy, Taxol is one of the most important drugs for therapy.

Due to the heterogeneity of ovarian serous carcinoma, Taxol resistance was detected in a large proportion of patients. This allowed for biomarkers and genes associated with Taxol resistance to be identified. For example, high expression of forkhead box protein M1 (FOXO1) was shown to contribute to paclitaxel resistance and correlated with survival of

Experts of Suzhou SZYJTD201707 to Jianrong Dai. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

ovarian carcinoma[3]. Inhibition of the IL6-STAT3 axis also had an effect on Taxol resistance in ovarian cancer cells[4]. The clinical significance of Cluster of Differentiation 44 (CD44) up-regulation was evaluated, and it is was also shown to be associated with Taxol resistance and the prognosis of ovarian cancer[5]. In addition to protein-coding genes, miRNAs were also widely identified and reported to contribute to Taxol resistance[6]. MiR-197 was reported to predict survival of ovarian cancer patients via the regulation of Taxol resistant ovarian cancer cells. Additionally, exosomal transfer of stroma-derived miR21 was shown to confer paclitaxel resistance in ovarian cancer cells through targeting of APAF1.

Though efforts have been devoted to discovering novel biomarkers for guiding chemotherapy using Taxol, none of these biomarkers were used clinically due to their low performance. Taxol resistance is related to multiple signaling pathways; therefore, a single biomarker is not sufficient to predict the effectiveness of Taxol. Transcriptome-based models utilize the expression level of multiple genes in various pathways and develop a model to guide prognosis and therapy[7–10]. Since this modeling incorporates a variety of cell statuses, including cell adhesion and immune cell infiltration, it has been emphasized and has become prevalent in many cancer types.

In this article, genes related to survival and Taxol resistance were identified from the TCGA dataset. Using the Cox multivariate regression model, a scoring system was developed to evaluate the effectiveness of the model. The Taxol score was significantly associated with survival; patients with a low Taxol score had showed better survival when they underwent Taxol chemotherapy, while those with a high Taxol score did not show better survival when they underwent Taxol chemotherapy. These observations were confirmed in other validation datasets. The correlation between the Taxol score and clinicopathological indicators was analyzed, and KEGG pathways associated with Taxol score were identified.

Materials and methods

Sample enrollment and pre-processing of data

The samples in TCGA, GSE63885[11] and GSE9891[12] with records of Taxol treatment were enrolled in this study. Recurrent and metastatic samples were excluded from this study. The expression matrix of the TCGA dataset was downloaded from UCSC Xena (xena.ucsc.edu), and the expression was transformed to log₂ RFPM values using the original data. The Z-score was calculated for each gene across the samples. The genes expressed in less than 50% of the samples were excluded. The raw expression data of GSE63885 and GSE9891 were downloaded from GEO (www.ncbi.nih.gov/geo). After background correction and normalization, the expression values were z-score transformed to remove the platform bias. The information of cohorts was listed in [S1 Table](#).

Gene selection and model development

To exclude genes that contribute to the survival and were independent from Taxol response, correlation analyses between gene expression and survival of ovarian serous carcinoma patients were implemented in the TCGA-Taxol receiving group and the TCGA-Taxol depleted group. Genes significantly associated with survival in the TCGA-Taxol receiving group ($p < 0.05$) while not significantly associated with survival in the TCGA-Taxol depleted group ($p > 0.05$) were selected as candidate genes. All the combinations of the candidate genes were identified, that means all combinations from $\bigcup_1^n C_n^i$ was tested (where n refers to the candidate genes and C indicates the combination, and i is from 1-n), and we developed a risk model using Cox multivariate regression with each combination using the following formula and

each combination,

$$\text{Taxol score} = \sum_i^n c_i x_i$$

where c_i is the coefficient resolved and x_i is the relative gene expression level. The survival differences between those with a low-score and those with a high-score were compared using Kaplan-Meier survival analysis. The combination model with the smallest p value was selected as the optimized combination.

Statistical analysis

Survival difference analysis, Cox multivariate regression, and univariate regression was carried out using the “survival” package on the R platform (www.r-project.org). The heatmap of candidate gene expression was plotted using the “pheatmap” package. Gene Set Enrichment Analysis was implemented using java software[13], with TCGA gene expression matrix and high-score/low-score separation using the median Taxol score as a cutoff.

Results

Candidate gene selection and model development

Taxol-resistant genes were selected using the following criteria in the TCGA dataset. First, the expression levels of candidate genes that were significantly associated with survival in the Taxol-receiving subgroup (patients who received Taxol therapy) according to the Cox univariate regression ($p < 0.05$) were determined. Second, the expression levels of candidate genes that were not significantly associated with survival ($p > 0.05$) in the Taxol-depleted group (patients who did not received Taxol therapy) were determined using Cox univariate regression to exclude to probability that the gene is a prognostic gene for ovarian cancer but not involved in Taxol resistance. Thirteen genes were then selected as candidate genes for model development. To eliminate redundancy and optimize the panel, all combinations of these thirteen genes were enumerated. The Cox multivariate model was developed using these combinations to calculate the Taxol score in the Taxol-receiving subgroup. Using the median Taxol score value as a cutoff, the survival difference between those with a high Taxol score and a low Taxol score was compared, and the p values were calculated using Kaplan-Meier survival analysis. Finally, a combination of six genes (CLIP2, LAX1, PCK2, THOC1, LEPR, GSTZ1) was used for model development, and the Taxol score was calculated as follows:

$$\begin{aligned} \text{Taxol score} = & (0.114148691 * \text{CLIP2}) + (-0.190271239 * \text{LAX1}) + (-0.157333524 * \text{PCK2}) \\ & + (-0.151729207 * \text{THOC1}) + (0.146653652 * \text{LEPR}) \\ & + (-0.092779415 * \text{GSTZ1}) \end{aligned}$$

where the gene symbol represents the relative expression level. Genes with positive values suggest that these genes contribute to the Taxol resistance, and vice versa.

Model performance in the TCGA dataset and validation dataset

The performance of the model was first evaluated in the training dataset, which was the TCGA dataset. The survival difference of the high and low Taxol score group was compared using Kaplan-Meier survival analysis. As expected, the patients with a high Taxol score had no significant survival difference between the Taxol-receiving and the Taxol-depleted group (Fig 1A and 1B). The good performance of the model may result from the over-fitness of the training dataset; therefore, validation cohorts were employed to assay the robustness of the model across datasets. GSE63885 and GSE9891 were used for validation due to the explicated Taxol

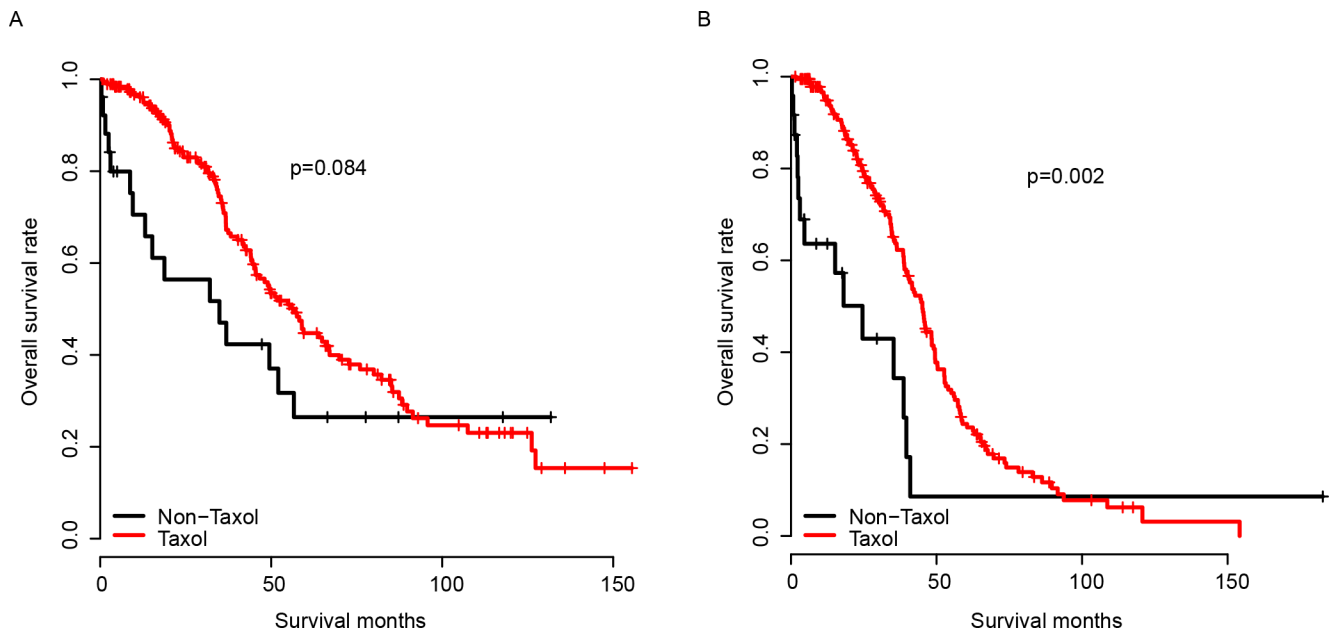


Fig 1. The performance of Taxol score in the training dataset. The survival of Taxol-receiving and Taxol-depleted groups was comparable in the high Taxol score group (A), while survival of the Taxol-receiving group was significantly better than the Taxol-depleted group in the low Taxol score group (B). The Taxol and Non-Taxol group refer to patients received and did not received Taxol (identical to Taxol-receiving and Taxol-depleted group).

<https://doi.org/10.1371/journal.pone.0192812.g001>

chemotherapy records. The Taxol score of each sample in the validation datasets was calculated using the same coefficients as the training dataset. The Taxol score was divided into a high Taxol score group and a low Taxol score group using the median Taxol score value in each group, as was done for the TCGA cohort. As expected, the patients who underwent Taxol chemotherapy had a better survival than those that did not undergo Taxol treatment in the low Taxol group. However, no significant difference in the survival was observed in the high Taxol group between patients who underwent Taxol chemotherapy or those who did not. This observation was confirmed in both the GSE63885 and GSE9891 groups (Fig 2A and 2D). In summary, the Taxol score is robust in predicting serous ovarian cancer’s response to Taxol in both the training cohort (TCGA) and the validation datasets (GSE63885 and GSE9891).

Clinical indicators, KEGG pathways and Taxol score

The correlation between the clinical indicators and the Taxol score was evaluated. As shown in Fig 3A, the Taxol score was independent from age, stage and Taxol treatment, while it was significantly associated with primary tumor size and grade ($p = 0.02$ and 0.0046 , respectively).

To assess the biological status that the Taxol score may reflect, significantly enriched KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways were evaluated using the Gene Set Enrichment Analysis (GSEA) by comparing the expression difference between the high Taxol group and the low Taxol group using the median value. As expected, metabolic signaling pathways including drug metabolism, cytochrome P450, and the ascorbate aldarate metabolism pathway were significantly enriched (Fig 3B and 3E). The ECM receptor signaling pathway was also enriched. In summary, the Taxol score reflects various biological states of serous ovarian cancer and predicts patient response to Taxol.

Discussion

Taxol is one of the most important chemotherapy drugs for ovarian serous carcinoma. Since the response to Taxol varies across individuals, single molecular biomarkers for drug usage

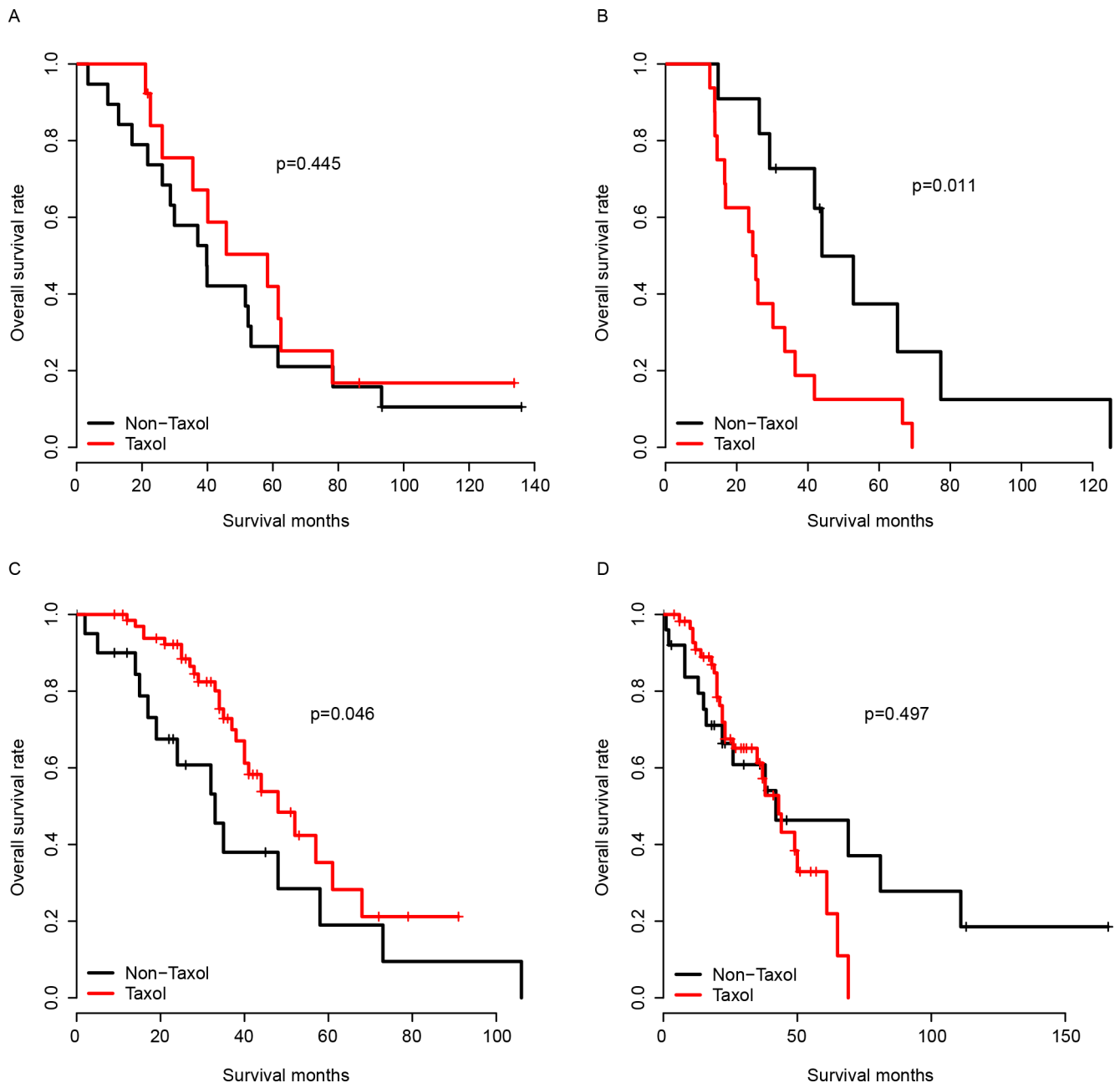


Fig 2. Performance of Taxol score in validation datasets. The performance of Taxol score was validated in the high Taxol score group (A) and low Taxol score group (B) in the GSE63885 cohort and GSE9891 cohort (C-D, high and low Taxol score group).

<https://doi.org/10.1371/journal.pone.0192812.g002>

have been developed in the past decades. However, none of these single molecular biomarkers have been used clinically due to their lack of robustness. The heterogeneity of serous ovarian cancer impacts the response to Taxol, which relies on many cellular processes including the cancer subtype and drug metabolism. Therefore, it is reasonable that the response to Taxol cannot be simply predicted using a single gene biomarker. To overcome this weakness, we introduced a Taxol response model and a Taxol score based on gene expression to predict the response of ovarian serous carcinoma to Taxol. The Taxol score successfully predicted the response of Taxol in three independent datasets.

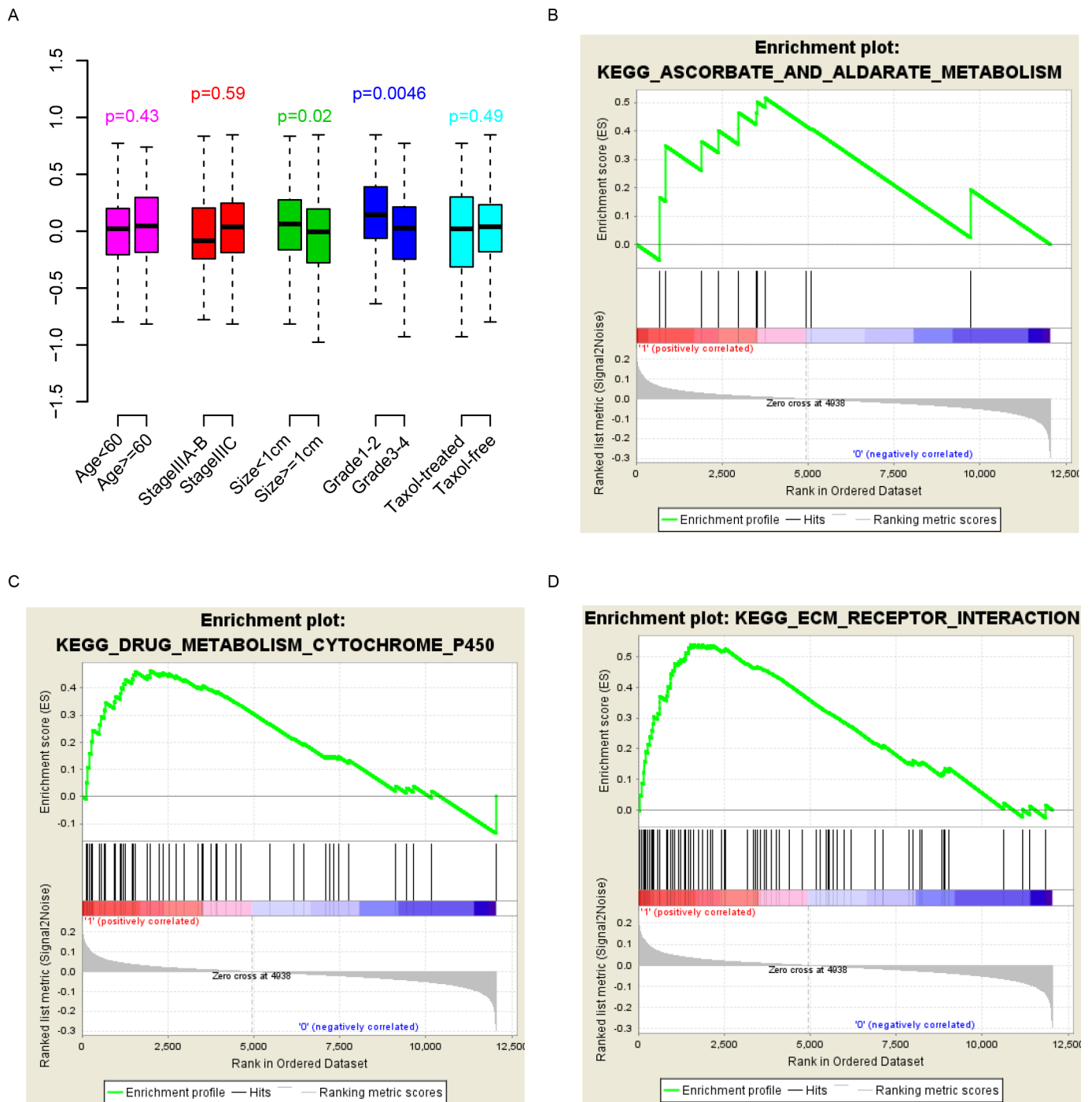


Fig 3. Pathways, clinical indicators and Taxol score. The Taxol score is independent of most indicators (A). Pathways associated with Taxol score were identified.

<https://doi.org/10.1371/journal.pone.0192812.g003>

Limitations exist in this work. First, this is a retrospective study. The patients who received Taxol chemotherapy also received other treatment methods, including other chemotherapy drugs; therefore, a bias may have arisen. Additionally, the dosage used was not always clear. Second, the tumor response to Taxol should have been measured by tumor size as it is better than the survival information; however, these data are unavailable.

Conclusion

The developed Taxol score is robust in predicting the sensitivity to Taxol across serous ovarian samples and cohorts via reflecting various status of serous ovarian carcinoma.

Supporting information

S1 Table. The raw data accessions and URLs. All data is uploaded by third party. (XLSX)

Author Contributions

Conceptualization: Shunyu Hou, Jianrong Dai.

Data curation: Shunyu Hou, Jianrong Dai.

Formal analysis: Shunyu Hou, Jianrong Dai.

Funding acquisition: Jianrong Dai.

Investigation: Shunyu Hou, Jianrong Dai.

Methodology: Jianrong Dai.

Resources: Shunyu Hou.

Supervision: Shunyu Hou, Jianrong Dai.

Validation: Shunyu Hou, Jianrong Dai.

Visualization: Shunyu Hou, Jianrong Dai.

Writing – original draft: Shunyu Hou, Jianrong Dai.

Writing – review & editing: Shunyu Hou, Jianrong Dai.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA: a cancer journal for clinicians*. 2016; 66(1):7–30. Epub 2016/01/09. <https://doi.org/10.3322/caac.21332> PMID: 26742998.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA: a cancer journal for clinicians*. 2016; 66(2):115–32. Epub 2016/01/26. <https://doi.org/10.3322/caac.21338> PMID: 26808342.
3. Zhao F, Siu MK, Jiang L, Tam KF, Ngan HY, Le XF, et al. Overexpression of forkhead box protein M1 (FOXO1) in ovarian cancer correlates with poor patient survival and contributes to paclitaxel resistance. *PloS one*. 2014; 9(11):e113478. Epub 2014/11/21. <https://doi.org/10.1371/journal.pone.0113478> PMID: 25411964; PubMed Central PMCID: PMC4239070.
4. Suh YA, Jo SY, Lee HY, Lee C. Inhibition of IL-6/STAT3 axis and targeting Axl and Tyro3 receptor tyrosine kinases by apigenin circumvent taxol resistance in ovarian cancer cells. *International journal of oncology*. 2015; 46(3):1405–11. Epub 2014/12/30. <https://doi.org/10.3892/ijo.2014.2808> PMID: 25544427.
5. Gao Y, Foster R, Yang X, Feng Y, Shen JK, Mankin HJ, et al. Up-regulation of CD44 in the development of metastasis, recurrence and drug resistance of ovarian cancer. *Oncotarget*. 2015; 6(11):9313–26. Epub 2015/04/01. <https://doi.org/10.18632/oncotarget.3220> PMID: 25823654; PubMed Central PMCID: PMC4496219.
6. Kim YW, Kim EY, Jeon D, Liu JL, Kim HS, Choi JW, et al. Differential microRNA expression signatures and cell type-specific association with Taxol resistance in ovarian cancer cells. *Drug design, development and therapy*. 2014; 8:293–314. Epub 2014/03/05. <https://doi.org/10.2147/DDDT.S51969> PMID: 24591819; PubMed Central PMCID: PMC43938445.
7. Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *The New England journal of medicine*. 2016; 375(8):717–29. Epub 2016/08/25. <https://doi.org/10.1056/NEJMoa1602253> PMID: 27557300.

8. Salazar R, Roepman P, Capella G, Moreno V, Simon I, Dreezen C, et al. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2011; 29(1):17–24. Epub 2010/11/26. <https://doi.org/10.1200/jco.2010.30.1077> PMID: 21098318.
9. Bou Samra E, Klein B, Commes T, Moreaux J. Identification of a 20-gene expression-based risk score as a predictor of clinical outcome in chronic lymphocytic leukemia patients. *BioMed research international*. 2014; 2014:423174. Epub 2014/06/03. <https://doi.org/10.1155/2014/423174> PMID: 24883311; PubMed Central PMCID: PMC4026849.
10. Kim SK, Kim SY, Kim JH, Roh SA, Cho DH, Kim YS, et al. A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. *Molecular oncology*. 2014; 8(8):1653–66. Epub 2014/07/23. <https://doi.org/10.1016/j.molonc.2014.06.016> PMID: 25049118.
11. Lisowska KM, Olbryt M, Dudaladava V, Pamula-Pilat J, Kujawa K, Grzybowska E, et al. Gene expression analysis in ovarian cancer—faults and hints from DNA microarray study. *Frontiers in oncology*. 2014; 4:6. Epub 2014/01/31. <https://doi.org/10.3389/fonc.2014.00006> PMID: 24478986; PubMed Central PMCID: PMC3904181.
12. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2008; 14(16):5198–208. Epub 2008/08/14. <https://doi.org/10.1158/1078-0432.ccr-08-0196> PMID: 18698038.
13. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102(43):15545–50. Epub 2005/10/04. <https://doi.org/10.1073/pnas.0506580102> PMID: 16199517; PubMed Central PMCID: PMC1239896.