Gynecologic Oncology 141 (2016) 49–56

Contents lists available at ScienceDirect

Gynecologic Oncology

journal homepage: www.elsevier.com/locate/ygyno

Prediction of taxane and platinum sensitivity in ovarian cancer based on gene expression profiles

Ryusuke Murakami [a], Noriomi Matsumura [a,*,1], J.B. Brown [b], Zhipeng Wang [c], Ken Yamaguchi [a], Kaoru Abiko [a], Yumiko Yoshioka [a], Junzo Hamanishi [a], Tsukasa Baba [a], Masafumi Koshiyama [a], Masaki Mandai [d], Ryo Yamada [e], Ikuo Konishi [a]

[a] Department of Gynecology and Obstetrics, Graduate School of Medicine, Kyoto University, Japan
[b] Kyoto University Graduate School of Medicine, Center for Medical Education, Life Science Informatics Research Unit, Japan
[c] Laboratory for Malignancy Control Research, Medical Innovation Center, Kyoto University Graduate School of Medicine, Japan
[d] Department of Obstetrics and Gynecology, Kinki University Faculty of Medicine, Japan
[e] The Center for Genomic Medicine, Kyoto University, Japan

HIGHLIGHTS

- Gene expression-driven scores predict platinum/taxane response
- Association of proposed scores with HGSOC gene expression subtypes
- Complimentary roles of platinum and taxane for personalized therapy

ARTICLE INFO

Article history:
Received 16 September 2015
Received in revised form 25 January 2016
Accepted 21 February 2016

Keywords:
Ovarian cancer
Gene expression profile
Prediction
Taxane
Platinum
Sensitivity

ABSTRACT

Objective. Prognoses of ovarian cancer (OC) have improved with the paclitaxel-carboplatin regimen. However, it remains unclear which cases exhibit a genuine benefit from taxane or from platinum. We aimed to predict taxane and platinum sensitivity in OC via gene expression.

Methods. We identified differentially expressed genes in responsive and resistant cases from advanced OC biopsy expression dataset GSE15622, containing responses to paclitaxel or carboplatin monotherapy. These genes generated a scoring system for prediction of drug response by applying single-sample gene set enrichment analysis. Discriminative metrics termed the T-score and C-score were derived.

Results. High C-score levels were significant in responders compared to non-responders in a separate cisplatin treatment dataset (GSE18864, $p = 0.043$). High C-score groups also had significantly better progression-free survival in three OC datasets (The Cancer Genome Atlas, TCGA: $p = 0.02$; GSE9891: $p = 0.03$; GSE30161: $p = 0.001$).

In two additional datasets of advanced OC, high T-scores could associate taxane and platinum regimens with better survival than non-taxane and platinum regimens (GSE9891: $p < 0.0001$; GSE3149: $p = 0.045$), whereas in cases with low T-scores, different chemotherapeutic regimens did not result in a significant difference. Assessing TCGA and GSE9891, T-scores were elevated in the C1/Mesenchymal subtype, whereas C-scores were elevated in the C5/Proliferative subtype and were lower in the C1/Mesenchymal subtype ($p < 0.0001$, respectively). C- and T-scores negatively correlated with each other, suggesting complementary roles of taxane and platinum.

Conclusions. Our proposal and finding of a scoring system that could predict platinum or taxane response could be useful to develop individualized treatments to ovarian cancer.

1. Introduction

The overall survival rate of women with epithelial ovarian cancer has improved since combination therapy of platinum- and (non-platinum) taxane-based drugs was introduced [1]. However, despite numerous phase-III randomized clinical trials since the combination's introduction, the standard of care remains the same for the primary

* Corresponding author at: 54 Shogoin Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan.
E-mail address: noriomi@kuhp.kyoto-u.ac.jp (N. Matsumura).
[1] Present address: 54 Shogoin Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan.

http://dx.doi.org/10.1016/j.ygyno.2016.02.027
0090-8258/

treatment of advanced epithelial ovarian cancer [2]. In both ovarian cancer and more general cancer research, specific gene mutations and amplifications are increasingly being used as biomarkers to predict sensitivity or resistance to molecularly targeted drugs [3]. Particularly for ovarian cancer, BRCA1/2 mutations as well as defects or aberrations in DNA repair-related genes are reported to be good predictive biomarkers for sensitivity to poly ADP-ribose polymerase (PARP) inhibitors [4]. These highly efficient biomarkers can help patients suffering from diverse cancers with a high response rate and can help avoid unnecessary treatment. In contrast to targeted therapies, there are fewer effective biomarkers for response to cytotoxic agents. In addition, many predictive models for response to chemotherapy in ovarian cancer have been based on combination chemotherapy rather than single therapy regimens [5]. Moreover, there is no clinical trial matching patients with cytotoxic drugs based on specific biomarkers to realize personalized cancer treatment. Therefore, we consider that predictive models for a combination chemotherapeutic regimen of cytotoxic agents should be assessed by evaluating monotherapy-based predictive biomarkers.

Thus far, cell line-based signatures derived from single chemotherapeutic agent response have been used for prediction models of response to combination chemotherapy [6]. However, commonly used cancer cell lines have been reported to have distinct patterns in mutations or copy number alterations that differ from those of clinical ovarian tumors [7]. In addition, the tumor micro-environment, which is different from cell culture, can confer innate resistance to chemotherapy. Heterogeneous cell types within tumors can actively influence therapeutic response and shape resistance [8]. Recently, two large-scale cancer cell line databases, the Cancer Genome Project (CGP) and the Cancer Cell Line Encyclopedia (CCLE), assayed a panel of several hundred cancer cell lines for gene expression and pharmacological drug response to a variety of anticancer drugs [9,10]. It was found that while gene expression is well correlated between the two databases, the measured drug responses were highly discordant [11]. Considering the inconsistency of drug response between these two big cell line projects, we can expect that patient-derived tumors will exhibit a similar diversity of responses to even the same therapy, and hence patient-derived tumors, rather than cell lines, should be used to generate predictive models for anti-cancer drug response in clinical settings.

Recently, the Australian Ovarian Cancer Study (AOCS) and The Cancer Genome Atlas (TCGA) studied the transcriptome of high-grade serous ovarian cancer (HGSOC) and identified four gene expression subtypes with distinct prognoses: C1/Mesenchymal, C2/Immunoreactive, C4/Differentiated and C5/Proliferative subtypes [12–15]. Gene expression profiles reflect distinct biological states with associated features, and therefore one hypothesis is that gene expression subtypes might also associate with drug response. A stratification of ovarian cancer using gene expression profiles could therefore be utilized for prediction of an individual's response to taxane or platinum treatment. In short, such stratification could help us to predict those patients who might exhibit a genuine benefit from taxane or platinum, and conversely, those who are likely to exhibit resistance. The aim of this study is to elucidate transcriptome profiles that are correlated with response to taxane or platinum chemotherapy.

We determined taxane- and platinum-predictive response/resistance signatures using a publically available gene expression microarray dataset of isolated paclitaxel or carboplatin treatments in an ovarian cancer clinical trial [16]. Leveraging the single sample gene set enrichment analysis (ssGSEA) scoring system, we explored a prediction model that could predict the response to taxane or platinum in independent gene expression datasets of clinical samples and mouse xenograft models [15]. The work is unique in its capacity to decompose the individual drug contributions in an effort to improve the expectation of response using clinical ovarian cancer samples. In future clinical settings where transcriptome measurement is routinely available and affordable, we should be able to determine the optimum combination treatment for ovarian cancer using separate taxane and platinum sensitivity prediction scores in each patient's diagnosis.

2. Materials and methods

2.1. Statistical analysis (see Supplementary Fig. S1 for a workflow chart and dataset summary)

2.1.1. The 1st step: gene expression microarray datasets

We obtained publically available gene expression microarray datasets from the NCBI's Gene Expression Omnibus (GEO) website (http://www.ncbi.nlm.nih.gov/gds/) and The Cancer Genome Atlas (TCGA) Data Portal (http://cancergenome.nih.gov/); ovarian cancer datasets (GSE15622: Affymetrix HG-U133A2.0, GSE9891: HG-U133plus2.0, GSE3149: HG-U133A, GSE30161: HG-U133plus2.0 and TCGA: HG-U133A), a breast cancer dataset (GSE18864: HG-U133plus2.0), and an epithelial ovarian cancer mouse xenograft dataset (GSE56920: Agilent SurePrint G3 Human GE v2 8x60K Microarray) were retrieved for the following analyses. The Robust Multi-array Average (RMA) method was used for normalization of Affymetrix microarray datasets using the R package “affy” (http://www.R-project.org/). The probe numbers of all Affymetrix datasets were mapped to conform to the HG-U133A gene probe names. Of the 22,215 gene probes, we retained the top half of probes with the highest average expression in the dataset used for gene set selection (see The 2nd step: identification of genes predictive of chemotherapy response) and then filtered this subset to the top half of probes with the highest standard deviation (SD), resulting in 5553 probes retained.

2.1.2. The 2nd step: identification of genes predictive of chemotherapy response

We identified differentially expressed genes in sensitive and resistant cases from the dataset of advanced ovarian cancer GSE15622, which contains data of laparoscopic biopsy specimens and clinical response to paclitaxel or carboplatin monotherapy. Samroc [17] was used for detecting significant differentially expressed genes between responders and non-responders for carboplatin and paclitaxel monotherapies. FDR q-values were calculated from Samroc p-values using the R library “p-adjust”.

Volcano plotting was performed using gene scores from Samroc statistics on the x axis and the negative base-ten logarithm of FDR q-values on the y axis, with points colored by response/resistance (see below). Samroc statistics indicate the difference in gene expression between two groups (responders and non-responders) for a given gene. Samroc statistics that further deviate from 0 indicate a larger difference in gene expression. Signature genes could be either up- or down-regulated, represented by positive values or negative values. Sensitive gene probes and resistant gene probes satisfied a requirement of FDR q-value < 0.0025. Thus, only the most extremely deviated genes remained as signatures of platinum/taxane sensitivity-predictive genes (UP genes) and platinum/taxane resistance-predictive genes (DN genes).

2.1.3. The 3rd step: generation of discriminative C-scores and T-scores

The therapy response-predicting genes were used to generate a scoring system, for evaluating external prediction of taxane or platinum response, by applying single-sample gene set enrichment analysis (ssGSEA). ssGSEA was used to generate scores for predefined signature gene sets in the Molecular Signatures Database. For a given sample, gene expression values were normalized and rank ordered. The empirical cumulative distribution functions (ECDFs) of the genes in the signature (gene set) and the remaining (non-set) genes were calculated. Integration of the difference between the ECDFs yields a statistic which is similar to the one used in GSEA, but based on absolute expression rather than differential expression [18].