Reviewer’s report

Title: Case-control Indian Buffet Process identifies biomarkers of response to Codrituzumab

Version: 1 Date: 31 Dec 2018

Reviewer: Jason Chia-Hsun Hsieh

Reviewer's report:

The manuscript presented a novel statistical method based on the Indian Buffet Process (IBP) to identify biomarkers predictive of response to treatment with Codrituzumab. It is a good try but I still have some questions to ask before it is considered to be published in the journal of BMC Cancer.

Major Concerns:

1. In the RESULTS, the PK data showing a highly varied range of drug exposure in the treatment arm and only half of the patients receiving appropriate drug exposure could explain why the primary efficacy endpoint was not met. Based on the condition, I believe that the authors should only analyze those patients who received appropriate drug exposure versus who did not. That means the results in Table 5 and Appendix 5-7 (F1: GPC3 IHC3+, F3: NK, CD8, CD45) are truly meaningful. If not, please explain.

2. The authors used PFS as a clinical endpoint and the C-IBP model identified twelve subpopulations from the set of 180 patients and three latent features (F1, F2, and F3). Why the authors did not use responder or OS as endpoints?

3. The method of clustering for getting a feature or signature(s) from numerous and complex factors is very useful in many situations. But the problem for this approach is how to persuade readers that the result from current dataset is applicable for the future group without a validation set?
Minor issues:

1. The methodology of C-IBP is not clear enough, please describe more in details. The software or algorithm the authors used are also required to be well described in the manuscript.

2. The expression of tested items in table 1, 2, 3 is confusing to readers (even to researchers). I suggest the authors express them more clinically readable. For example, 'membranous GPC3 expression level' might be more easy to read and understand rather than 'H_score_mem'.

3. In Table 2B, I am not sure that 'CD3/CD16_ necrotic/stroma' means the percentage of dual CD3+CD16+ cells in necrotic tissue against those in the stroma or other. These way of terminology here seems to be very confusing.

4. In table 2B, I can not find 'P_Necrotic/stroma' and 'AAT, alanine aminotransferase' in the table. Please clarify.

5. In table 2C, is there any clear definition of viable/necrotic tissue and stroma? Please clearly define in the manuscript.

6. In the table 2C, the p-value of CD56dimCD16- was 1.04e-05, what are its hazard ratio and confidence interval, may I ask? Without that, I cannot confirm that the description in line 45-48 on page 10 in Discussion is correct.

7. In Table 2A-2C, are these factors found by C-IBP independently significant? If yes, CD8+NK are a subgroup of NK cells, are they affecting each other (confounding)?

I believe it is a good try that the authors proposed a special method (C-IBP) to explain the results from a failed phase II study. We always want to learn something from clinical data even it was failed. But the natural thing is that the drug exposure is inappropriate, the methodology advances cannot change the fact and data source. The drug group is not really a drug group, which weakens the strength and rationale of Table 2, Table 3 and Table 4.
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?
If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

none

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal