Author’s response to reviews

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

Authors:

Melanie Pradier (melanie.fpradier@gmail.com)
Bernhard Reis (bernhard.reis@roche.com)
Lori Jukofsky (lori.jukofsky@roche.com)
Francesca Milletti (fmilletti@kitepharma.com)
Toshihiko Ohtomo (ohtomotsh@chugai-pharm.co.jp)
Fernando Perez-Cruz (fernando.perezcruz@sdsc.ethz)
Oscar Puig (oscar@phosphorus.com)

Version: 2 Date: 31 Jan 2019

Author’s response to reviews:

Dear Editor in Chief and Reviewers,

Thank you very much for your time and valuable comments, which have helped us improve our manuscript. We are pleased that the reviewers considered our work interesting and original. Please find below our answer to the specific comments of reviewers 1 and 3.

To facilitate the reviewing process, we have highlighted in red all improvements to the manuscript. In addition to the main manuscript, we have also updated the Additional Files: 1, 2, 4, 5, and 6. Please let us know if any of our answers is unclear, and we will do our best to clarify it further.

Sincerely,

The authors
## Reviewer 1 comments ##

R1: I think it is an innovative approach for predictive and prognostic biomarkers exploration for biologically complicated tumor types using novel mechanism-based biotherapy. The ability to differentiate the impacts between prognostic & predictive factors and quantify effect size of each biomarker is also brilliant. The comparison of this novel approach with traditional multivariate regression model only deserves more investigation and literature reviews. If more examples of other tumor types with different kinds of biotherapies, analyzed in the same method, could be raised in the recent publications, the convincing power may be quite bigger!

The authors: We thank the reviewer for his positive feedback, and take note of his comment for a future extended work on a wide range of applications and datasets.

## Reviewer 3 comments ##

R3: 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.

The authors: We agree with the reviewer. Indeed, patients who received high drug exposure are those that bring us most information about the therapy, but the low drug exposure group is also informative. Therefore, our approach deliberately includes all patients in order to learn a joint
meaningful patient representation via latent features that manifest differently in different patients. Including patients that received low drug exposure allowed us to discover prognostic factors (in latent feature F1 and F2), as well as drug-specific effects (in latent feature F3) by sharing information (e.g., occurrence of correlation signatures) across patients with varying levels of drug exposure.

Action: We have clarified this point in the main manuscript: “C-IBP deliberately includes all patients in order to learn a joint meaningful patient representation via latent features that manifest differently for different patients. (...) Table 5 and Additional Files 5-7 confirm the statistical significance of the discovered biomarkers by performing two-sample tests only on patients who received high drug exposure (but after having found a latent projection based on information from all patients).”

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?

The authors: PFS was the clinical endpoint of the trial. The study was not powered to detect statistical differences in OS. OS is informative, and can be included in additional exploratory analysis, but the main conclusions must be derived from PFS as endpoint.

Action: We added a footnote to clarify this point: “PFS was the clinical endpoint of the clinical trial. The study was not powered to detect statistical differences in overall survival (OS).”

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?

The authors: We agree, and this is to some extent a current limitation of our study. The clinical trial is large (Phase 2) and allows the method to be deployed accurately, but we agree that an independent dataset would add additional confidence. That said, having a validation set is critical for supervised methods, where the objective is to predict some data Y given some other input
data X. The algorithm is then trained using data from a train set, hyperparameters tested in a validation set (via cross-validation for example), and final performance assessed in a completely independent test set. However, the proposed approach is unsupervised, and it is useful for data exploration purposes. The method finds a projection of the data (analogous to what Principal Component Analysis would deliver), for which a validation set is not absolutely necessary.

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.

The authors: Thank you for your comment, we agree. Action: We have added a more detailed description in the main text, under the Statistical Analysis subsection. We have also improved the Additional File 1; in particular, we have added an illustration of the whole pipeline. We have additionally released the software used in the paper together with extensive API descriptions and examples; this is publicly available at: https://ivaleram.github.io/GLFM/

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'.

The authors: Thank you for the suggestion. Action: We have updated the name of the tested biomarkers to be more clinically readable (see Additional File 2 for a full updated list). We also corrected a couple of inconsistencies in the name of variables between the Additional File 2 and Tables in the main text (see comment 4).

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.
The authors: Thank you for the comment. It is the count of CD3 CD16 double positive cells in the necrotic tissue AND stroma tissue. Necrotic tissue and stroma tissue are counted together. Action: We have clarified this point in the manuscript, namely, “CD3/CD16 necrotic is the count of CD3 CD16 double positive cells in the necrotic tissue and stroma tissue counted together.”

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

The authors: AAT was statistically significant in the case of feature F1, and thus, it can be found in table 2A. Regarding “P_Necrotic/stroma” biomarkers, those were listed in Table 2B as “%_Necrotic/stroma”. Action: We have corrected and standardized the nomenclature in the text, figures, and tables of both the main manuscript and additional files.

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

The authors: This is the amount of viable cells in necrotic tissue and stroma, counted together. Thanks for the comment. Action: we have clarified this point in the manuscript (see also comment 3 above).

6. In the table 2C, the p-value of CD56dimCD16 was 1.04e-5, 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.

The authors: We believe the reviewer refers to the subset CD56dimCD16bright, which was found the most predictive in our study (Table 2C and 4) and also in the literature as related to poor prognosis in HCC (reference 17). Action: We have added the hazard ratio and confidence interval to the table caption, which were 0.75 [0.57; 1.00].

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)?
The authors: Thank you for the comments. All statistical significance tests in this manuscript were performed independently. We cannot rule out confounding effects in these two subgroups of cells (CD8+ NK cells are a subgroup of total NK cells). However, levels of expression of CD8 and CD56 vary, and flow cytometry does not always allow to effectively separate count from surface marker expression levels. For example, there could be constant levels of CD56, which would allow counting NK cells, but variable levels of CD8 across NK cells in different samples, which could be interpreted as more or less count of CD8+ NK cells, or alternative same count of NK cells with higher or lower levels of CD8 surface marker. Sometimes, absolute cell count vs relative level of expression of the surface markers are hard to differentiate.

Action: We have addressed this point in the manuscript by acknowledging that there are possible confounding effects in the Discussion section: “...we remark that, although all statistical tests were independent, we cannot rule out all confounding effects across subgroups.”

R3: 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.

The authors: As the reviewer points out, a varying drug exposure across patients is an important challenge of the data source that cannot be reversed. But this difficulty is what makes C-IBP particularly attractive and suitable for this kind of data. While variable drug exposure confounded previous statistical approaches such as regression models [1], C-IBP was able to unravel statistically significant biomarkers despite of the confounder. Although patients received different levels of drug exposure, C-IBP learned a meaningful latent representation of patients by sharing information from all patients. The learned latent features were used to identify homogeneous subgroups of patients for whom classical statistical analyses were performed.

Action: We have added an additional paragraph in the Discussion section to emphasize this point: “While variable drug exposure (…) confounding effects across subgroups.”