Reviewer’s report

Title: ClickGene: an open cloud-based platform for big pan-cancer data genome-wide association study, visualization and exploration

Version: 0 Date: 05 Feb 2019

Reviewer: Pablo Cámara

Reviewer's report:

Bi et al. present an online application (ClickGene) for plotting processed genomic data of large cancer cohorts. There are multiple available tools of this type, the most well-known perhaps being cBioPortal. These tools are useful for performing quick exploratory analyses of processed cancer data and generating hypothesis. The specific tool introduced in this paper allows the user to create several types of plots that are not included in cBioPortal. It also performs simple statistics. The online application comes pre-loaded with the TCGA dataset, but also allows the user to upload their own dataset.

Although I find the tool is worth publishing, there are multiple aspects with the way it is presented in the paper that I think need to be revised:

1. I find the discussion section in the paper to be very misleading. The authors focus the discussion on results that they claim are new or inconsistent with the current literature. However, as far as I see, none of those results represent an inconsistency, but rather a misinterpretation of the results by the authors. For instance, the authors look at copy number alterations of chromosome arms 1p and 19q in glioblastoma. Co-deletion of 1p and 19q is a well-characterized event in low-grade IDH-mutant oligodendrogliaoma, to the extent that it is now one of the defining characteristics of IDH-mutant oligodendrogliaoma according to the WHO Classification of Brain Tumors. In high-grade gliomas, 19q deletions are rare, usually only observed in secondary IDH-mutant glioblastomas. Most glioblastomas therefore do not carry a 1p/19q co-deletion. In the paper, the authors look at the average copy number of 19q in the TCGA glioblastoma cohort and find that it is above 2 (figure 6) and conclude that "19q should not be considered as a deletion". I do not see any inconsistency here with the previous literature. Moreover, looking at the mean is misleading, as it is particularly sensitive to outliers. If instead, we consider the median, we observe that 19q indeed is very often deleted in low-grade gliomas, and occasionally also in high-grade glioblastomas, as evidenced by a similar plot produced with ClickGene but using the median instead of the mean. In my opinion, the paper would improve substantially if the authors focus the discussion section in presenting some examples of analyses of the TCGA data using ClickGene, highlighting the consistency with previous studies.
2. In the discussion, the authors present dynamic time wrapping (DTW) scores for the copy number alterations of adenocarcinomas and squamous cell carcinomas (table 1), and highlight differences among different cancer types. However, the differences in score are small (<5%) and the authors do not perform any statistical test to assess the significance of these differences. Are they consistent with random fluctuations? The authors could use some approach, like bootstrapping, to better assess the significance of these differences.

3. The application includes simple statistical tests. However, significances in the plots produced by ClickGene are not adjusted for multiple hypothesis testing. I find this is an important aspect in this context, as many of the tests are performed thousands of times to produce a plot. In my opinion, the authors should implement at least one of the standard procedures for controlling the false discovery rate (e.g. Benjamini-Hochberg).

4. Although in general the online application works reasonably OK, it often gets stuck without giving an error. For instance, if I try to generate a Beeswarm plot of the copy number variations of gene ADAMTS12 in lung adenocarcinoma (LUAD), the application goes into the "Loading..." screen indefinitely, without giving an error.

5. Overall, I find the writing could be much improved. There are many grammatical errors, as well as colloquial expressions that do not seem suited for a scientific publication.

**Level of interest**
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

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

Not suitable for publication unless extensively edited

**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.

I declare that I have no competing interests

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

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world’s journals. Signing up is quick, easy and free!

No