Reviewer’s report

Title: Use Case Driven Evaluation of Open Databases for Pediatric Cancer Research

Version: 0 Date: 07 Sep 2018

Reviewer: Pablo Cámara

Reviewer's report:

The authors provide a comprehensive overview of the currently-available main databases and online exploratory resources for pediatric cancer research. As highlighted in their manuscript, pediatric cancer research databases are much less developed than analogous databases for adult cancers. This is something that presumably will change in the coming years, as major initiatives, such as the Pediatric Cancer Genome Project, the Children's Brain Tumor Tissue Consortium, or the Treehouse Children Cancer Initiative, continue to generate and deposit their data. Meanwhile, it is important that these initiatives take into account what has been learned over the past years by major consortia in adult cancers (e.g. TCGA and ICGC) with regard to data sharing, formatting, exploratory tools, etc.

Hence, this paper could serve in my opinion two purposes: 1) to provide a good introduction to the main available online tools and resources for pediatric cancer research, and 2) to identify and bring into discussion the main needs and limitations of the current online resources for pediatric cancer research. However, I believe to that end the following points should be taken into consideration:

1. The research content of the manuscript is limited as compared to research articles usually published in BioData Mining. I think it is hard to perform meaningful quantitative comparisons between such a heterogeneous set of databases and tools, and qualitative comparisons can be rather subjective. Furthermore, because online resources for pediatric cancer research are evolving rapidly, the content and conclusions of the manuscript will likely become outdated in the near future. Based on this, I strongly believe the manuscript should be restructured and resubmitted as a "Review" rather than as a "Research Article".

2. The use cases presented in the manuscript in my opinion fail to capture some of the main limitations of current resources. For instance, the availability of processed data (e.g. accessing and downloading expression matrices, somatic variant calls, allele frequencies, etc.) is limited and often restricted to cancer-related genes. This is usually insufficient for robust exploratory/hypothesis-generation studies and one of the main differences between resources like TCGA and current pediatric cancer databases. I believe the quality of the
manuscript would improve by adding a discussion of these aspects which computational biologists working in cancer might find useful.

3. In Figure 2, I find the differences in mutation frequency among various tools/databases striking. Are they due to differences in the tumor subtype composition of the databases on which these tools build, or to differences in the parameters used for variant calling? It could help to have the authors comment on this. As it stands, it seems unclear what information the figure is conveying.

4. I find it might be helpful to further disaggregate the databases into genomic resources (PCGP, ICGC, TARGET, Treehouse, etc.) and clinical/epidemiological resources (ACCIS, POGO-NIS, etc.).

**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:

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.
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