Reviewer Report

Title: Transcriptome annotation in the cloud: complexity, best practices and cost.

Version: Original Submission   Date: 9/17/2020

Reviewer name: Saskia Hiltemann

Reviewer Comments to Author:

The authors provide a comparison of two cloud-based solutions for running BLAST-based transcriptomics analysis.
With cloud-based solutions becoming more popular in science, I think this comparison, along with the practical recommendations provided in this manuscript will be interesting to readers.
Some suggestions for enhancements below:
1) The authors mention that there are numerous genomics companies in the space of cloud based biocomputations, citing reference 14 which discusses some of the legal responsibilities of groups doing such cloud based analyses. However, the authors do not connect this discussion back to the use of GCP or AWS, where users would need to obtain similar gaurantees of data security which may not be possible to obtain. Thus a comparison of pricing against the private firms which provide similar services would be interesting to see if they provide specific guarantees that affect researchers with data that requires specific legal requirements.
2) I really like the inclusion of a best practices section with practical recommendations, and would love to see this section expanded:
 a) In the first point the authors state: "We recommend CWL because the resulting product is portable and scalable,and it can be executed across a variety of computational environments as dissimilar as personal laptops or the cloud". However, these features are not exclusive to CWL, and solutions such as NextFlow and SnakeMake (and probably others) would also fit this description (and both of these also offer point 2 Conda and containerization). Please elaborate on your recommendation to include discussion of other workflow management systems, and explain in more detail why you would recommend CWL over these other solutions.
 b) In point 5, you recommend that users "Execute a small test in the cloud to find the best instance type for a workflow". Do you have any further practical recommendations about how users can best go about this? E.g. how does one define a "small test run" from a full datasets, and how can they predict how this will scale up to the full analysis and determine the most suitable machine types?
3) It could be nice to expand the section about the Jupyter notebooks, and how these are being used, perhaps with some screenshots of results, and/or a small schematic showing that (if I understand correctly): the user interacts with the Jupyter notebook on their local machine, which in turn configures the cloud resources and starts the CWL workflow on the cloud, and then fetches the relevant results back and analyses them and displays results to the user. I think a schematic to this effect would be helpful for less technical readers and this in combination with some screenshots of the analysis results in Jupyter will increase the appeal of your work to research scientists.
4) In the conclusion the authors state "In our opinion, the choice of a cloud platform is not dependent

on the workflow but, rather, on the specific details of the cloud provider". However, I don't believe the authors can make this statement having tested only a singe workflow. So please rephrase the conclusion, or compare performance of different workflows covering a range of different characteristics (e.g. one that is memory-intensive, one that is CPU-intensive, and one that requires a lot of data transfer) and showing whether this conclusion holds, or whether some of the "specific details of the cloud provider" may make it more or less suitable for certain types of workflows.

5) AWS, Google, and Azure are probably the "big 3" providers that most readers will l know about, and they might wonder why Azure was not included here and how it would compare. I understand that the authors cannot compare all providers, but it may be useful to at least mention Azure in the introduction where different providers are mentioned, and briefly explain if there were any specific reasons why you chose to compare AWS and GCP, and whether the same methodology could also be applied to Azure.

Level of Interest

Please indicate how interesting you found the manuscript: Choose an item.

Quality of Written English

Please indicate the quality of language in the manuscript: Choose an item.

Declaration of Competing Interests

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

- 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?
- 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?
- Do you hold or are you currently applying for any patents relating to the content of the manuscript?
- 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?
- Do you have any other financial competing interests?
- 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.

Choose an item.

To further support our reviewers, we have joined with Publons, where you can gain additional credit to further highlight your hard work (see: https://publons.com/journal/530/gigascience). On publication of this paper, your review will be automatically added to Publons, you can then choose whether or not to claim your Publons credit. I understand this statement.

Yes Choose an item.