

Reviewer Report

Title: Bioinformatics applications on Apache Spark

Version: Revision 1 **Date:** 6/16/2018

Reviewer name: Joshua W. K. Ho

Reviewer Comments to Author:

The quality of the manuscript has been improved after revision. I believe this review represents a timely contribution on the use of big data frameworks (such as Spark) in bioinformatics. Nonetheless, I still feel the sections on 'Spark in motif analysis' and 'Spark in population genomic inference' are still very confusing and potentially misleading. For example, calling 'motif' a type of noise in NGS data, and naively equating motifs with TFBS is incorrect. This indicates the authors may not fully understand these terms. Also, I still do not understand what do they mean by 'genomic inference' - estimating ancestry? population admixture? Simply re-iterating words in my questions is not sufficient. The authors should either significantly rewrite both sections and have these checked by experts in these areas, or should simply remove them. I am happy with the rest of the manuscript.

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.

I agree to the open peer review policy of the journal

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