Reviewer Report

Title: Chromatin conformation capture (Hi-C) sequencing of patient-derived xenografts: analysis guidelines

Version: Original Submission **Date:** 12/11/2020

Reviewer name: Noam Kaplan

Reviewer Comments to Author:

In recent years Hi-C has been applied to cancer genomes, with the aim of both characterizing cancer-specific alterations of 3D genome organization as well as changes in the 1D genome sequence such as structural variations. Patient-derived xeongrafts (PDX) provide an important system for studying cancer, and is associated with unique technical problems as the human tumor cells are contaminated with mouse DNA. In the current manuscript, the authors test a number of different techniques, both computational and experimental, and try to evaluate which combination provides better quality data. Such a study can be quite useful for other groups pursuing Hi-C in PDX, and while currently this work will be of interest to a relatively small group of specialists, the potential applications of Hi-C in cancer may widen the interest in this type of work in the future.

The authors test 3 different techniques for differentiating mouse from human reads, 3 different Hi-C computational pipelines, and 2 different Hi-C protocols (commercial kits). The authors look at a number of different quality statistics, and conclude that the best combination of approaches is probably "Direct" mapping of both human and mouse reads, Juicer Hi-C pipeline, and the Arima Hi-C kit.

In general the paper is well written, clear and technically sound. My main issue is with the interpretation of the quality statistics.

Main issues:

1) The authors propose a number of different Hi-C quality statistics, but there is often a tradeoff between these statistics. Thus, in order to show that one method is better than others one needs to show an improvement in all parameters. For example, the authors state that Juicer is better than the other pipelines, and this is supported by more valid reads mapped and better cis/trans ratio. However, the Long/short read ratio worse than the other methods. Cis/trans is a good measure for evaluating the amount of random ligation (which yields more cis than trans). However, another type of common bias in Hi-C data is short range cis interactions that may result from a number of causes such as insufficient digestion or contamination by unligated fragments. It is entirely possible that Juicer maps more of these incorrect short range read pairs, and this is reflected by a higher cis/trans ratio and more "valid reads". Thus, it is not possible to determine based on the current metrics that Juicer is better than the other pipelines. Specifically for this case, A possible control for this bias would be to calculate "valid mapped reads" and "cis/trans" using only reads>20kb. More generally, I would be careful with drawing strong conclusions about quality unless all statistics point in the same direction (this is not to say the results are not useful, just that the conclusions might need to more careful).

2) It is unclear why the authors can conclude that a "smaller" power-law exponent is better (note that the way the authors use the term "smaller" is confusing here because these are actually negative

numbers, -1.83 is not smaller than -1.99). Artifacts like background ligation can cause a shallower decay.
3) The same goes for TADs. TAD calling pipelines can be affected by data biases in different ways, especially since these are often hierarchical overlapping structures, and it is certainly not clear whether finding more TADs is better or worse in terms of data quality. For example, with a higher level of background ligation/mismapped reads, it could be more difficult to identify larger TADs, so only the nested TADs are found resulting, in more TADs.
Minor issues:
1) It might be useful in Figure 2 to add a vertical horizontal at the 10% and 30% threshold, where relevant.
2) Is there any mismapping of reads mouse->human in the in silico data?

Methods

Are the methods appropriate to the aims of the study, are they well described, and are necessary controls included? Choose an item.

Conclusions

Are the conclusions adequately supported by the data shown? Choose an item.

Reporting Standards

Does the manuscript adhere to the journal's guidelines on minimum standards of reporting? Choose an item.

Choose an item.

Statistics

Are you able to assess all statistics in the manuscript, including the appropriateness of statistical tests used? 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 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.