Reviewer Report

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

Version: Original Submission **Date:** 11/30/2020

Reviewer name: Joanna Achinger-Kawecka

Reviewer Comments to Author:

Dozmorov, Tyc et al. present guidelines for analysis of Hi-C data generated from PDX models with respect to mouse DNA contamination in the sample. They use in silico spike-in of mouse Hi-C reads and actual Hi-C data from PDX samples to show that different approaches to mapping mouse and human reads and read processing do not affect the final Hi-C maps.
This is an important work and will be of high value for the 3D cancer genome field, however I do not think that the results presented by the authors justify the conclusions and therefore more analysis needs to be done before this manuscript can be published.
The key analyses that need to be performed are to look at the effect of mouse spike-in reads or mouse cell contamination on chromatin interactions. Presented results focus only on high-level domain structures (TADs) and are limited to look at the total number/size of TADs called. TAD boundaries called from Hi-C data have been previously shown to be highly overlapping between mouse and human genomes as well as in some other species (as recently discussed in Eres and Gilad, Trends in Genetics, 2020). However, the main correlation between TAD calls in different datasets can be explained by the use of the same calling algorithm. Therefore, TADs and TAD boundaries are not a good measure of the effect of mouse cell contamination in Hi-C data. Instead, the analyses should be focused on chromatin interactions (or enhancer-promoter interactions), which are more cell-type specific. Authors need to show how many mouse-specific interactions are present in the final Hi-C data from PDX samples as well as look at the enrichment of all valid interactions for mouse vs human enhancers, promoters and CTCF binding (using public histone mark data or chromHMM and CTCF ChIP-seq).
Minor comments:
1. The difference between two Hi-C kits used (Library 1 vs Library 2) including names of the kits and restriction enzymes used should be included somewhere at the front of the results section.
2. Can the 40% duplication in Library 1 (Phase Genomics kit) be explained by over-sequencing of the library that is not complex enough due to only one RE used in the kit?
3. Fig. 5 - TAD number and sizes are not a good quality metric for this question as they are mainly driven by the type of the algorithm used to call TADs. The authors should instead include analysis of the actual insulation score/directionality index that underlines the TAD calls and show correlation between the scores, PCA/MDS plot and look at overlap between called boundaries to see if there are any mouse-specific TAD boundaries that are present in the in silico Hi-C data and in vivo PDX data.
4. Authors should look at interactions that are associated with mouse-specific genes - can these be observed in the in the in silico Hi-C data and in vivo PDX data? Some visual examples are needed as well.
5. It is expected that PDX Hi-C data will show more intra-population heterogeneity as compared to cell

line Hi-C data. This will affect "background" noise interactions, which may be present only in small sub-populations of cells and therefore affect the signal to noise ratio. Can this be clarified from the different analysis pipelines used and therefore be a key consideration for researchers when deciding on the best pipeline to use for PDX samples?
6. In PDX tumour samples, mouse fibroblasts have been shown to infiltrate tumours and introduce mouse signal in the analyses data. Can the authors look at fibroblast-specific interactions (e.g. based on fibroblast genes) in the PDX data to see if these can be detected?

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 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.