Reviewer’s report

Title: Structural Variation of Centromeric Endogenous Retroviruses in Human Populations and their Impact on Cutaneous T-cell Lymphoma, Sézary Syndrome, and HIV infection

Version: 0 Date: 05 Sep 2018

Reviewer: Jacklyn Hellwege

Reviewer's report:

This manuscript by Kaplan et al describes a thorough investigation into population and various disease frequencies of K111 and K222 proviruses. They show that the K111 deletion is uncommon in African-origin populations, and also that the deletion associates with worse phenotypic outcomes in cutaneous T-cell lymphomas. The paper is largely descriptive and provides adequate context for its findings. There are some areas which could benefit from improvement however.

1. The K222 findings seem secondary and rather distracting from the bulk of the manuscript which seems to revolve around K111. Although the two proviruses are clearly related to one another, including K222 complicates an already nuanced description. Specifically, the finding that total K111 and K222 combined is different between CTCL patients and controls does not add much value to the manuscript when the technical interpretation is that K111 is likely driving the difference and has already been presented much earlier in the results.

2. The term pericentric instability is first mentioned in the results on page 12 as having already been referred to earlier in the paper. A more complete description of what the author's interpretation of this term would be beneficial.

3. The closing paragraph of the introduction mentions briefly that a goal is to characterize K111 frequencies in diverse populations, however, the discussion is dominated by these observations rather than the disease specific comparisons. A bit of balance in expectation vs delivery would help the reader to orient themselves.

4. It might be helpful to restrict comparisons from the CTCL group to those without documented bone marrow transplant, specifically, excluding SZ4, from sensitivity analyses would assist in interpretation of those results. Secondary comment: in the results (page 13, line 209) it states that the K111 status of the donor is unknown, while in the discussion (page 28 line 484) seems to be stating overtly that the donor marrow is definite.
5. The description of copy number analysis results on page 20 beginning on line 327 documents very large numbers of copies (up to 58000/100pg DNA), however the accompanying figure axis is labeled K111 copies per 100pg/DNA and a range from 0 to 6. This is confusing and the scale seems off.

6. There are multiple instances of numbers in tables not matching those in text. While some may be explained by the combining of groups, I don't believe all are. For instance, line 279 on page 17 reads "...while in African and African American women it was 2.16% (n=142) (Table 4)". In table 4, however, only 139 women from Nigeria are listed, and the prevalence is 1.44%. Please double check table and text values for concordance.

Minor note: on page 14 line 235, the K111 prevalence in healthy African American individuals was ostensibly compared to Caucasian HIV patients, which have not previously been discussed, and for reasons which are not entirely clear.

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

Quality of written English
Please indicate the quality of language in the manuscript:
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.