Reviewer’s report

Title: Genome sequencing analysis of blood cells identifies germline haplotypes strongly associated with drug resistance in osteosarcoma patients

Version: 0 Date: 27 Mar 2018

Reviewer: Edwin Choy

Reviewer's report:

This is a well-written manuscript describing an analysis of two data-sets to try to identify genetic determinants of chemoresponsiveness in patients with osteosarcoma. The authors performed whole genome sequencing of 15 treated patients from the INOVA dataset and performed association studies for genes that associated with relapse outcome in both the INOVA and TARGET datasets. Using a p-value of \( \leq 0.05 \), the authors identified 110,000 haplotypes that associated with relapse in the INOVA dataset, and 2178 haplotypes that associated (to p value of < 0.05) in the TARGET dataset. They then found a total of 231 overlapping haplotypes. The problem with this entire approach is that even if the relapse phenotype was randomly determined with no genomic correlations, by random chance the study would identify similar numbers of associated haplotypes. Furthermore, by random chance, among the 110,000 haplotypes identified in the INOA dataset and 2178 haplotypes identified in the TARGET dataset, one would also find about 200-300 overlapping haplotypes. This merely means that the 110,000 haplotypes seen in the INOVA dataset comprise 10\% of the genome tested in the TARGET dataset. A p value of 0.05 does not adequately account for the sheer number of hypothesis being tested in a whole genome association study. A more appropriate statistical analysis would employ permutation analysis with LOD scores and only results that are significant genome-wide (p value of 10e-7) would be significant. The TARGET dataset should then be re-analyzed to see if the significant INOVA haplotypes remain significant. But merely overlapping the haplotype lists does not confirm the findings of a discovery dataset. I agree with the authors that their findings "need to be validated in a larger cohort of patients". But until they are able to do that, this small, extremely underpowered study does not convince me that their gene list is statistically meaningful.

Likewise, their analysis of DMET genes using 59 patients would also result in approximately 4,543 hits seen in this manuscript even if their phenotypes were randomly generated.

In conclusion, the authors fail to show that their gene lists were not results of random, statistically inevitable lists of genes that could have been generated even if the phenotypes being tested were randomly permuted.

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

No
Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

No

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

No

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.

I recommend additional statistical review

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

Acceptable

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.
no conflict

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