Reviewer’s report

Title: Complex structural rearrangements are present in high-grade dysplastic Barrett's oesophagus samples

Version: 0 Date: 22 Jun 2018

Reviewer: Carlo Maley

Reviewer's report:

This paper adds genome sequencing from 16 Barrett's esophagus (BE) samples and compares them to 22 previously reported esophageal adenocarcinomas (EAC). However, 6 of those samples (including 5 of the 7 non-dysplastic samples) were from patients who had synchronous EAC and so are not good models of pre-cursor lesions (those patients may have had worse exposures and/or genomic instability, over longer periods of time than the non-progressors). This may (or may not) explain why there were few statistically significant differences between the non-dysplastic and the dysplastic BE samples. The previous study, by Stachler et al. compared dysplastic to non-dysplastic samples, all from patients with synchronous EAC. 4 of the samples in the present study came from two patients that did not progress to EAC. Those samples have very low mutational/structural burdens, consistent with previous studies. With so few samples, from such different conditions, it is hard to be confident in any trends in the data.

The main value I see in this study is the contribution of whole genome sequences from BE without concomitant EAC. Each of those samples tells a story, but together, there aren't enough of them to determine trends. The authors focus on the fact that 2 of the 5 dysplastic samples had evidence of chromosomal catastrophes (2 others had "complex events"). This provides preliminary evidence for a hypothesis that those catastrophes occur before EAC and drive rapid progression to EAC.

The conclusion "The presence of complex localised rearrangements in dysplastic 86 samples may be a trigger for the progression from BE to EAC." While technically true that it may be a trigger, the data is not very strong - there are many differences between non-progressor tissue and progressor tissue - that does not imply causation.

The authors report comparisons of 2 biopsies separated by 3-4 years from each of two non-progressors. It is difficult to interpret that data with N=2 and lacking any data on the spatial heterogeneity in BE. We know from SNP array studies that there is extensive spatial heterogeneity in BE, so it is hard to know if the differences between the two biopsies is due to taking samples in different locations or to changes over time in the BE segment. The fact that the number of mutations went down over time in one patient suggests that spatial heterogeneity may be playing the dominant role.
Furthermore, the authors report only 13% and 42% overlap in the mutations present in the two biopsies from the same patient. However, this often depends on the mutation calling pipeline. If there is poor coverage or quality of sequencing in a locus in one of the two biopsies, a mutation that is actually present there can easily be missed. Since they authors use the consensus between two mutational callers, they are being relative conservative about calling mutations, and so are vulnerable to this kind of false negative result. A further analysis should look for week support for a mutation in one sample if there is strong support for it in another sample. Furthermore, GATK is designed for germline mutations, not somatic mutations. That means, it is not designed to deal with the clonal heterogeneity within the sample. A caller like MuTect, EBCall or Strelka is preferable (See https://doi.org/10.1371/journal.pone.0186175).

The differences between EAC and BE with respect to mutational signature 1 should be controlled for age of the patient.

Determining the baseline ploidy in order to do copy number analyses is difficult, and has dramatic effects on the estimations of copy number changes. The authors did not describe how this was done in order to be sure the baseline is diploid.

The telomere length data is also hard to interpret, as it is a function of both proliferation and telomerase/ALT activity. More importantly, the minimum length among the chromosomes of any one cell, not mean length is more important for generating chromosomal alterations.

**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.

Yes

**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.

Not relevant to this manuscript
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.

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