Reviewer's report

Title: Whole exome sequencing of microdissected splenic marginal zone lymphoma: a study to discover novel tumor-specific mutations.

Version: 2
Date: 29 May 2015
Reviewer: Jonathan Strefford

Reviewer's report:

This manuscript reports a limited genomic analysis of splenic marginal zone lymphoma, using whole exome sequencing of two patients and some targeted validation and extension. There is a lot of work here, and the manuscript is nicely prepared and concise.

There is limited novel data reported though, and the study is small and probably under powered, particularly compared to three recent studies identifying KLF2 mutations, that are not mentioned at all. Clearly, it is nice that they have successfully performed WES from micro-dissected material. The exome sequencing is difficult. The authors mention a minimum coverage of 20 and a novel VAF of 10%, so 2 reads for calling a variant is very low and likely to include high levels of false positives. Furthermore, only 66% of the exome was captured at this depth, so the authors are missing a lot here. This is particularly important as the authors emphasize sub-clonal dynamics, which seems even more uncertain in this context. How can the authors be certain of their observations? Also, I see no evidence of a comprehensive comparison with published data? Gene lists are available from all the published studies, so a comparison would be helpful. I wonder if they over play this card, and should make it less of a focus?

Whilst this study is limited, the authors do report a putative novel mutation target SMYD1, with exonic mutations present in two cases, and an additional 3' variant. This is a potentially interesting observation but needs further work. Are these validation variants somatically-acquired? Is this frequency of mutations highly then you would expect based on background rates? Have these mutations been reported in the extensive exome sequencing data published already? Can they screen additional cases? This would strengthen the study hugely.

In conclusion, the reviewer feels that this work falls below a publishable standard in its current form and suggests significant alterations before further consideration.

In addition to the above major compulsory revisions, please consider the below as minor essential revisions.

1) the introduction is missing key references concerning recent papers in this area, focusing on KLF2 from and Italian and UK group, but also a large clinical study published in Clinical Cancer Research (which also performs bioinformatics
analysis of deep-sequencing data to provide some preliminary insights into sub-clonal diversity).

2) The discussion states that their approach to micro-dissecting samples provided better insight into clonal evolution, but there technical design of the experiment does not support this, i.e. such low sequencing depth.

3) We know certain genes are false positives in this type of analysis using MutSig approaches, but the authors do not mention this.

4) I cannot see Figure 1, so cannot assess it. Is the title of this figure correct, as I am not sure it shows ‘coding complexity’?

5) Is the MYD88 mutation in the dup(3q) case duplicated? You would expect this, and it would be interesting to know.

6) What are the coverage stats for certain genes, as we know NOTCH2 is poorly captured in WES studies?

7) The authors should consider providing KLF2 data on their cohort, as this is the most important gene in this disease.

8) I think the discussion is ‘over sold’ and needs to be toned back to be a better reflection of actually what they add to the field. The sub-clonal section particularly is not discussed in the context of the quality and quantity of the sequencing data they have.

Level of interest: An article of limited interest

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests