Reviewer’s report

Title: Single-Cell Sequencing Analysis Characterizes Common and Cell-Lineage-Specific Mutations in a Muscle-Invasive Bladder Tumor

Version: 1 Date: 17 May 2012

Reviewer: Choon Kiat K Ong

Reviewer’s report:

Li et al did single-cell exome sequencing on 55 cells extracted from a muscle-invasive subtype of bladder tumor; standard range of analysis performed on the filtered data: SNV, tumor phylogeny to infer evolution, prevalence screening. Touched only slightly on CNA and pathways analysis. The conclusion being, given the data available, the clonal evolution model is the best description for bladder cancer development.

Major Compulsory Revision:

1) While a fair amount of work and resources were expended, the authors reaching a bit on their conclusions. Their findings are new and interesting but their methods in assessing genetic details and clonal evolution are not new (nature paper on breast cancer single cell sequencing, the two recent Cell papers from BGI and Swanton’s paper in NEJM). Only minor variations in analysis.

2) The authors have mentioned and emphasized the important of “selective pressures” and “cancer driver genes” (Page 10 paragraph 2 and page 14 paragraph 2 are the examples) in bladder cancer development yet did no analysis to back their claims when tools for calculating driver genes probability were already developed and published years ago at Sanger by Greenman et al (http://www.ncbi.nlm.nih.gov/pubmed/16783027). I would suggest the authors to perform some of these analyses to back up their claims. This will elevate the quality of the analysis.

3) The authors’ use statistics to avoid doing independent validations of their predicted mutations; statistics are meant to support the claim of a lesser amount of validations are needed to support an overall FD rate and not to claim no validations are necessary using statistics especially since they are looking at a new type of cancer.

Authors’ claim in p9 that using tissue sequencing, they “… could only identify 134 (30.25% of 443) … mutations … using single-cell analysis,” and that “… single cell analysis had a much higher sensitivity for identifying rare mutations.”

This could be due to the shallow depth exome seq done on the single-cells (see sup fig S2D) that a lot of these “high confidence” mutations predicted as in fact false positives. I would suggest that the authors perform some validation:
I. validated randomly ~15 mutations in the common 134 mutation pool across tissue and single cells to ascertain the accuracy of the overlap

II. Are there predicted high confidence somatic mutations in the tissue but not in any of the single cells? If so, validated those mutations in the tissues and the single cells to ascertain the false negative rate of the single cell sequencing and false positive rate of the tissue.

III. There are 309 mutations that were predicted in single cell seq that were not predicted in tissue seq; to ascertain these mutations are due to the increased sensitivity of the single cell seq technique, the author need to randomly validate ~25-30 of these mutations in both the single cells where the mutations are predicted using sanger seq and also in tissue using mass spec. The authors' has shown the clonal split to be ~1/3-1/3-1/3 and well within the detection range of the mass spec technique they employ to generate fig 2b. If the DNA amount from single cells are not sufficient for sanger validation, I will at least ask the mass spec part of the validation to be done in order to prove increased sensitivity of single cell seq over tissue seq.

4) The CNV-LOH analysis is only performed for the tissue seq; can this also be performed for single-cell seq? This will serve two purposes:

One, to see the common CNV-LOHs between clonal populations and to see if there are any additional changes unique to the sub-populations not detectable via tissue analysis. This will also allow us to see if clonal population split is in concordance between CNV-LOH and SNV analysis.

Two, to see if the normal cells being sequenced are really “normal” ... they should lack any of the anomalies found in the tumor cells. This should support the PCA done using mutant alleles.

Minor Essential Revision:

1) The supplementary tables and figures should labelled so there is no confusion to the readers when searching for the tables referenced.

2) Supplementary table S1a: Specifically sample BC-8 that shows a mean depth of 18.53 but a coverage of 86.20% ... is there a typo in entering the mean depth? Any particular reason why the exome sequencing is inconsistent?

3) Supplementary table S4a,b the x-axis labelling is unhelpful ... what is Variable?

4) Supplementary table S4a shows FDR vs sequencing depth to be zero or close to zero. Are the authors' saying that FDR is invariant as a function of sequencing depth or there is a correlation but the axis needs to be adjusted to show it?

5) The authors' should not use Navin et al 2011 nature publication (http://www.nature.com/nature/journal/v472/n7341/full/nature09807.html) as a comparison as they were employing shallow WHOLE genome sequencing to quantify copy number while the authors' are exome sequencing for SNV
detections. Similar techniques but sequencing depth requirement for accuracy is completely different between copy number detection and SNV detection and it would be disingenuous of the authors' to claim a meaningful comparison. A more appropriate comparison would be to compare against Xu et al 2012 Cell publication (http://www.ncbi.nlm.nih.gov/pubmed/22385958) and/or Hu et al 2012 Cell publication (http://www.ncbi.nlm.nih.gov/pubmed/22385957). Both publications are more recent and more applicable to the authors' study.

6) Several typo errors: Bottom of page 7, 12.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests.