Reviewer’s report

Title: Targeted High Throughput Sequencing in Clinical Cancer Settings: Formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity.

Version: 1 Date: 23 May 2011

Reviewer: Ivo G Gut

Reviewer’s report:

In their manuscript “Targeted High Throughput Sequencing in Clinical Cancer Settings:.....” the authors Kerick et al. describe progress they have in combining high resolution analysis of tumour DNA extracted from formalin-fixed sections. They analyse extracted DNA with solution-phase targeted enrichment. The manuscript focuses on two aspects, the extraction of DNA, quality and quantity required for the processing with the enrichment system and the quality in sequence that can be achieved. This part of the work is very interesting and of great value. The second aspect pertains to the possibility to detect somatic differences between DNA taken from different parts of the same tumor section and conclusions that can be drawn from this.

Major issue:

Due to differential coverage copy number variants are detectable according to the authors, which is interesting, however differences in somatic single nucleotide variants are not found (P13 L383. Here there is an issue, the percentage of the entire genome sequence covered in this study is very small (0.1 % and 2%), a published whole-genome somatic variation scan in AML published by Wilson detected only a handful of somatic single nucleotide variants. One would expect a solid tumor to have more somatic mutations than AML, however, even if the number were in the 10s thousands the number that will fall into the enriched target is small. The assumption that there might be a detectable differential is stretched, because the same variant could be present by chance. This result is not sufficiently corroborated. The right experiment would be to carry out a whole genome somatic mutation scan, develop an enrichment bait set for the detected variants and then test DNA samples taken from different parts of the tumor section for these detected variants.

Another important point is that a condition for identifying a variant should also be that the variant is observed both in sense and antisense reads. There is no reference to such a procedure.

The presentation of the manuscript is clear and adheres to the standards laid out by the journal. The data seems sound and the descriptions are clear. As pointed out above, one of the main conclusions seems flawed and should either be reworked or omitted. The content of the work, even in absence of this is of substantial general interest and very timely.
Minor issues:
P2 L65 here “in the long term” should be used rather than “in the long range”
P5 L122 5-10 cm² seems a very large area – is this correct
P6 L151 sentence makes no sense
P7 L185 qPCR does not have hyphen
P7 L201 3-fold coverage seems too low as a threshold
P9 L240 correct individually
P10 L252 the introductory paragraph belongs to the discussion and not the results section
P11 L298 the type of modifications that might occur are speculative relative to this manuscript – they either need to be corroborated with a reference of a measurement
P11 L301 20x coverage – does this refer to greater than or smaller than?
P15 L417 same as L298

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

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