Reviewer 1

1. When using synthetic DNA to evaluate the sensitivity and sensibility of RVD2, much is dedicated to (currently) unrealistic coverages. For example, in the Figure1 and Table1 (Figure 3) the authors presents a sequence coverage of 5.000x and even 55.000x. Despite being very useful for the own authors as high hypothetical control, I think ~30x - 100x coverage is more informative, and presenting these extremely high coverages can be misleading.

The reviewer correctly points out that much next-generation DNA sequencing data has a median depth in the 30x - 100x range. **We have added a paragraph to our discussion of the sensitivity and specificity results (second paragraph in Section 5.2).** In that paragraph, we discuss three primary median depth ranges and typical applications for those areas. Briefly, whole genome sequencing typically has depths in the range of 10x - 100x as noted. In addition, targeted cancer gene sequencing typically has median read depths in the range 100x - 1,000x because the region of interest is smaller and there is interest in measuring tumor heterogeneity. Microbial and viral sequencing can have median depths from 1,000x - 100,000x because one is interested in identifying rare subpopulations that may prove resistant to antibiotics or antivirals.

We have chosen a broad range of median depths for our results so that researchers in whole genome sequencing, targeted sequencing and microbial sequencing can see how the methodology would perform in their application area.

2. Despite having the whole genome sequencing for HCC1187 and HCC1187BL cell lines, the authors focus all presented analysis one a single gene. It's not clear why the authors used the coordinates chr7:154738059154782774 (~45Kb), excluding the three first exons (four introns) and the last exon and most of the last intron, when, in fact, PAXIP1 has ~59Kb. No further explanation is given regarding why this specific region was chosen.

Thank you for identifying this important omission. The region of interest for the PAXIP1 gene should have covered the entire 59Kb coding region. We ran RVD2 on the current NCBI start/stop codons (http://www.ncbi.nlm.nih.gov/gene/22976) to give the expanded region of chr7:154,735,400 - 154,794,682. This expanded region yields additional mutations. In particular, using the expanded region, we identified 5 additional germline variants and 2 additional somatic variants. In expanding the region of interest the previously identified germline variants at chr7:154,766,700 and chr7:154,781,769 are not called. A close inspection shows that these loci have very similar point estimates of the allele frequency in the case and control in the shortened region of interest. In the analysis with the expanded region, RVD2 estimates a lower across-locus precision parameter (M0) and thus does not call these loci. The control M0 parameter is 5.7 in the previous analysis and 4.2 in the 59kbp region.

3. Instead of using DNA sequencing of primary tumors genomes, which are highly heterogeneous and present many rare (single nucleotide) variants, the authors used the DNA from a commercially available cell line, which is colonially expanded (genetically homogenous). Interestingly, chromosome 7, where the investigated gene is located, is tetraploid on HCC1187[1], therefore, one would expect to find only allele frequencies 0%, 25%, 50%, 75 and 100% on variants located in this chromosome. Figure 5 seems to be in agreement to this hypothesis, except for variants 49704 and 81769, which has marginally lower estimations of allele frequency.Therefore, the variants detected on HCC1187 are not actually rare variants (MAF=25%) and, as shown on Figure 3 (Table1), may be detected by most of currently available methods. Despite the limitations, the analysis is complete, but it would be more informative to apply RVD2 to publicly available (and validated) DNAseq from primary tumors.

Thank you for pointing out the evidence that HCC1887 is tetraploid for chromosome 7. **We have incorporated this evidence into our discussion of the data with the reference your provided.** We agree that it would be informaive to apply RVD2 to publicly available and validated DNAseq from primary tumors. We have searched extensively for such a data set and we have experienced three limitations. First, RVD2 requires sequencing data from a control sample. The control sample does not need to be matched, but should be of the same sequence as the case. This control sample allows RVD2 to measure the locus-specific error rate and call significant deviations from that error rate in the case. This requirement limits the pool of possible data sets. Second, we require a reasonable read depth (as shown by the synthetic DNA experiments) to identify rare variants. Many DNAseq data sets have a read depth of 10-15x because the investigators are seeking to identify clear variants. Few data sets have a read depth higher than 15x. Finally, we wanted a data set for which the originators provide a set of validated variants. Again, this requirement limits the number of available data sets. Taken together, these requirements are met by the HCC1187 data set. We expect that with the increased interest in tumor heterogeneity, investigators will be generating more data sets that meet these criteria. It will be interesting to explore those data sets with RVD2.

4. The authors cite a paper from 2011 to refer for the Synthetic DNA sequencing. Explicit details on the technology, read length and number of reads generated would greatly ease readers understanding of this section.

Section 4.1 describing the details of the experimental protocol for the synthetic sequencing data is indeed brief. **We have added the details the reviewer suggested and as expected, the narrative is much clearer.** Briefly, Supplementary Table S1 in the Nucleic Acid Research publication shows that each sample has roughly 1,000,000 35bp paired-end reads with no errors in the multiplexing barcodes to give an average read depth of ~100,000 across the synthetic amplicon. We describe this and more detail as suggested in the protocol section.

5. The authors greatly expose the comparison of sensibility, specificity, true positive and true negative rates. It seems that the caveat of the method is on memory consumption and processing time. It would be very informative to explicitly have a table with execution time and memory consumption of all utilized methods, preferably in a more realistic scenario, more specifically, a whole genome analysis.

The memory consumption and computational time for the algorithm is an important consideration. We intentionally focused on the accuracy of the methodology here and so chose Metropolis-Hastings sampling as our inference algorithm. Metropolis-Hastings is known to converge to sampling from the true posterior distribution in the limit. As such, presently the algorithm is not as computationally fast as it could be. To improve the speed of the algorithm, we have implemented a multi-threaded version to take advantage of a multi-core computational configuration. **We provide computational speed results in Section 5.1 and Supplementary Section 8 as suggested and discuss avenues for improvement that we are currently pursuing.**

Reviewer 2:

1. page 4 line 22; check the signs of the threshold boundary conditions in your hypothesis tests.. looks like one should be less than minus Tau.

Thank you for this observant correction. Indeed, the equation on page 4 line 22 should be minus tau. **We have corrected the equation in the main text as well as the same equation in the supplementary information in the caption for Figure S6.**

2. On preprocessing methods (4.1); please check what you mean by the C50 option of BWA. That should be something else.

Thank you for catching this important detail in the description of our methodology. We used the aligned bam files described previously, and ran samtools mmpileup with the -C50 flag to filter for high quality reads. The -C50 options is recommended in the samtools manual: http://samtools.sourceforge.net/mpileup.shtml. **We have corrected the manuscript to reflect this.**

3. On figure 6 panel A; at position index 35, 36, 37 and position 50; please check that the symbols shown for miscall versus mutant calls are as you intend them.. looks like two are reversed.

We are unable to identify what symbols are reversed. The RVD2 germline and somatic calls in Figure 6 do correspond to the same calls in Figure 5. Figure 5 and Figure 6 do use different symbols to denote their data, but the data is consistent.