Reviewer's report

Title: Genomic investigation of etiologic heterogeneity: methodologic challenges

Version: 1 Date: 6 October 2014

Reviewer: Debashis Ghosh

Reviewer's report:

This paper extends previous work by the authors in terms of attempting to identify subtypes using a combination of multiple "-omics" platforms. In the previous work, multiple tumors from the same individual were available. Now, this is not the case, and as a result, the authors propose a modification of their D statistic, which they call D-. It is illustrated with application to clear cell kidney cancer data from TCGA.

Major compulsory revisions:

1. The lede is buried very much towards the end, but the authors themselves admit on line 403 that "we cannot assert with confidence that 4 is the appropriate number of subtypes." This suggests that the method is not very numerically stable. Some better justification for the number of clusters is needed.

2. lines 203 and 210: the authors should specify exactly how many terms are being added up in the equation. The current presentation is too casual.

3. line 241. The authors used what appeared to be a self-competitive test in the jargon of Goeman and Buhlmann. Can they explain why they used this rather than a self-contained test?

4. lines 252 - 254: The authors should show a plot to verify this phenomenon.

5. line 285: A useful rule of thumb for how many starts are needed for the k-means clustering should be given.

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

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests.