Reviewer’s report

Title: Allele-specific disparity in breast cancer

Version: 2 Date: 29 October 2011

Reviewer: Kun Huang

Reviewer’s report:

This paper presented the authors’ search for allele-specific disparity in breast cancer using paired blood tumor samples from a large cohort of breast cancer patients in different stages. Using a workflow/algorithm developed in this paper using B-allele frequency and log R ratio of Illumina SNP data to estimate Euclidean distances, the authors were able to detect SNPs with substantial difference between blood and tumor with high occurrence frequencies among the patients. The results are consistent with known genes for breast cancer and the frequency increases with stages of breast cancer development.

There are a few comments:

1. in Suppl. Figure 1, some part of the flowchart is not complete. In one cell it says "Label & find Tumor' samples have shifted from heterozygous towards". But not clear on "towards" what?
2. The description of the same algorithm in the top part of page 7 is not clear. The "First" step is described but it is unclear what are the following steps (there is no "second" or "next" steps).
3. Equation (1) described how S is computed. But how is S used? It is not clear in the following description.
4. Figure 6 does not seem to be the network functions. It is the IPA output for Functional Analysis and bar heights are minus log of the p-values (from Fisher's exact test) instead of FDR. In IPA, networks were presented using network scores, not p-values.
5. Some texts are not clear, for instance, the last sentence of the "Results and Discussion" section is "Comparing the two gene-sets to see if the with regards to amplification and deletion the stem cell genes are more often altered by deletion than by amplification (Table 5B)." Besides the grammatical error, it is not clear on what the authors are trying to present. This is the results section, what are results derived from this comparison?
6. I would suggest the author separate Results and Discussion into two sections. It is not always clear on what are the results and what are the implications.
7. Several reference literatures (e.g., 19, 34, 35) are not complete.

Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Needs some language corrections before being published

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

Declaration of competing interests:

I declare that I have no competing interests.