Reviewer's report

Title: A computation model to predict bone metastasis in breast cancer by integrating the dysregulated pathways

Version: 2 Date: 28 April 2014

Reviewer: Haiquan Li

Reviewer's report:

The paper describes an approach to predict bone metastasis in breast cancer patients. Unlike classic methods, this one combines predictions from multiple deregulated pathways for an ensemble decision. The potentially distinct deregulation directions among multiple genes in a pathway are also well modeled in this classifier.

Minor Essential revisions:

1) The type of class values is not explicitly described. There are three types of category values and two types of numeric values mentioned in the paper. The former includes bone-metastasis patients, non-bone metastasis patients and non-metastasis patients, while the latter includes bone metastasis risk and survival days. In case of category class values, the problem is a classification problem; otherwise, it is a regression problem.

2) Mann-Whitney U-test is usually applied on two sample/population test and is seldom used in testing a subset versus a superset of samples. Distinct samples are suggested for such test. For example, the test between deregulated genes and candidate genes should be revised to deregulated genes and non-deregulated candidate genes.

3) The title of Y-axis of Figure 2 is confusing. Since all samples are included (e.g. 380 samples in Panel A), the title of “probability of bone-metastasis-free survival” does not make sense. In addition, the K-M curves in Supplementary Figure S1-S3 are not cited in the manuscript.

4) In the literature, the claim of few studies using gene signatures for classification is not correct. There are a lot of such studies and at least two tests are commercially available for breast cancer, which are MammaPrint and Oncotype DX.

5) Some grammar issues are in the manuscript: there is a running on sentence in the last paragraph of the Section “Distinguishing bone metastasis risks by DPBM”; the “train dataset” is a term less frequently used, and is sugusted to revise as “training dataset”.

Discretionary Revisions

1) In case the problem is a classification, did the authors check the precision and recall (or other equivalent measures) in the training, testing, and independent datasets? Were the results good enough and why if not? The discussion of these
issues will enhance the paper.

2) The hazard ratios in the training, testing, and independent datasets are impressing. However, please address the potential overestimation issue arising from unbiased number of controls versus controls. A random sampling of controls with the same amount of cases in each dataset will unveil the robustness of the hazard ratios (p-values are expected to be less significant though due to the reduction of sample size).

3) The method employed a single cutoff (from Gene Expression Grade Index) for each pathway to make a prediction. Has other classification methods been exploited based on all the deregulated gene signatures directly, such as SVM?

Level of interest: An article of importance in its field

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.