Reviewer's report

Title: Using logistic regression to improve the prognostic value of microarray gene expression data sets: application to early-stage squamous cell carcinoma of the lung

Version: 3  Date: 15 February 2014

Reviewer: Zhifu Sun

Reviewer's report:

The manuscript “Using logistic regression to improve the prognostic value of microarray gene expression data sets: application to early-stage squamous cell carcinoma of the lung” applied logistic regression to select prognostic genes that predict short and long survival of squamous cell carcinoma of the lung. The authors first selected the survival extremes of 20 short and 20 long survivors from a public dataset with 107 stage I and II cases and used the model to find candidates and then refined the list by sliding windows. Twenty four prognostic genes were found by the approach and most of these were immune response related. In spite of the rigorous work and potential value, lack of validation in a different dataset (or datasets) and insufficient evidence to demonstrate its superior performances to commonly used methods make it difficult to assess its practical value and usage.

Major Compulsory Revisions

1) Most commonly used survival analysis for continuous predictors is Cox proportional hazard model, but the paper did not mention or compare with it.

2) The results need to be validated in at least one independent dataset to demonstrate its value. Without the validation, it is not clear about its generalizability or it is better than the published literature.

Discretionary Revisions

1) Lung cancer patient survival is rather complex and contributed by multiple factors. Tumor molecular signature is just part of story. The logistic regression had no other factors considered in the model. Selecting the small subset for such analysis probably made that impossible.

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