Reviewer's report

Title: Low expression of a few genes indicates good prognosis in estrogen receptor positive breast cancer

Version: 2 Date: 15 April 2009

Reviewer: Md Rafiul Hassan

Reviewer's report:

1. Is the question posed by the authors well defined?

The prognosis of diseases through analysis of gene expression data is a challenging task and has appeal in many fields that include cancer biology and disease pathway. Author tried to solve the prognosis in estrogen receptor positive breast cancer which enables a better treatment for early stage breast cancer patients.

2. Are the methods appropriate and well described?

This paper proposes an accelerated progression relapse test for breast cancer prognosis that uses few genes and a statistical modelling. As far as I can tell, the idea in the paper is to use the existing mixture model and finds out the threshold point at which the expression level of a multi-state probe differs relating to the metastatic cases and non-metastatic cases. The experimental findings and post-analysis of the result proves the efficiency of the modelling. However, it is not clear how the optimal threshold point is identified. As I understood, author has chosen the threshold point/ cut-off point through the inspection of the mixture model, which is subjective. Unless a mathematical expression is provided to find out the optimal cut-point c from the mixture Gaussian distribution I don’t find any usefulness of the proposed approach. Hence, detailed description of the approach with required mathematical proves are required to accept the paper for publication.

It would have been great, if author can compare the experimental findings with the existing unsupervised clustering techniques (e.g. clustering using single hidden markov model, Unsupervised Learning of Finite Mixture Models, BIRCH clustering etc.)

3. Are the data sound?

Author has used dataset collected from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) which seems sound and ok.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?

Author has followed the exact standard of BMC publications to prepare the manuscript and it adhere to the relevant standards for reporting and data
5. Are the discussion and conclusions well balanced and adequately supported by the data?

The discussion and conclusions have been fairly written to analyse the findings of the paper.

6. Are limitations of the work clearly stated?

What are the limitations of the approach is missing in the paper.

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?

A brief introduction about the parameters of mixture model would make the paper more understandable.

8. Do the title and abstract accurately convey what has been found?

The title and abstract of the paper reflects the summary of the findings described in the paper.

9. Is the writing acceptable?

The paper has been well written and the English is acceptable.

- Major Compulsory Revisions

This paper proposes an accelerated progression relapse test for breast cancer prognosis that uses few genes and a statistical modelling. As far as I can tell, the idea in the paper is to use the existing mixture model and finds out the threshold point at which the expression level of a multi-state probe differs relating to the metastatic cases and non-metastatic cases. The experimental findings and post-analysis of the result proves the efficiency of the modelling. However, it is not clear how the optimal threshold point is identified. As I understood, author has chosen the threshold point/ cut-off point through the inspection of the mixture model, which is subjective. Unless a mathematical expression is provided to find out the optimal cut-point c from the mixture Gaussian distribution I don’t find any usefulness of the proposed approach. Hence, detailed description of the approach with required mathematical proves are required to accept the paper for publication.

It would have been great, if author can compare the experimental findings with the existing unsupervised clustering techniques (e.g. clustering using single hidden markov model, Unsupervised Learning of Finite Mixture Models, BIRCH clustering etc.)
Quality of written English: Acceptable