Reviewer’s report

Title: A gene expression profile for accurate tumour cell percentage scoring of breast tumour tissue: microarray diagnosis eligibility

Version: 1 Date: 9 April 2009

Reviewer: Jen-Tsan Ashley A Chi

Reviewer’s report:

Review for A gene expression profile for accurate tumour cell percentage scoring of breast tumour tissue: microarray diagnosis eligibility by Ropeman et al.

This paper addresses an interesting question of using microarray data to estimate the tumor cell percentage (TCP) using microarrays. Although this question is not urgent right now, but this issue will be important if the use of microarrays for breast cancers become extensive, which is part of the mission of Agenda where these authors work. Even though this issue is not of concern in current medicine, I could see the value of these studies. To achieve this goal, the authors first use the estimated TCP from pathologists as “golden standards” and phenotypes for supervised analysis on the training datasets. From these supervised analysis, a small subset of 13 genes (all highly expressed in tumors with TCP) were put forward due to their ability first in the cross-validation sets and then in an independent cohort of patients to test the ability of these genes in predicting the TCP in more breast cancer patients.

Although I can see the value of this study, I have some some major concerns which need to be addressed.

1. Theoretically, the TCP can be represented by the degree of gene expression of either epithelial cells (for breast carcinoma) or proliferation cluster (for mitotic activities) as surrogate. It may be of value for the authors to compare the performance of 13 genelists with these two possible gene signatures to represent the TCP.

2. The 13 gene listed in Table I is of limited obvious biological information other than ANAPC7 (associated with proliferation). Previous text in the result mentioned several cancer-related processes associated with the 35 gene TCP. What are the genes in the 35 genes which account for these GO term enrichment? I am pretty sure that no such GO enrichment would be found for the 13 gene list in Table I. It is of concern of using 13 genes without no obvious biological linkage to represent a obvious histopathological phenotypes (TCP) under microscopy.

3. Although the authors claims that this 13 gene score system can separate the tumors with low, medium and high TCP reasonable well. But the actual data between these groups had significant overlap. More relevant analysis may be to
use the TCP gene expression score to separate the tumors into 3 groups – high, medium and low TCP score and see whether this classification will correctly predict the rating of pathologists.

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests