Reviewer's report

Title: Integration of Transcript Expression, Copy Number and LOH Analysis of Infiltrating Ductal Carcinoma of the Breast.

Version: 2 Date: 22 January 2010

Reviewer: Ian Campbell

Reviewer's report:

This article describes an integrated analysis of gene copy number/LOH and gene expression in a small cohort of infiltrating breast ductal carcinomas with the aim of identifying genes which drive breast tumorigenesis. The adoption of an integrated approach is becoming well established and should provide a more reliable way of identifying driver genes from background noise. The study is an advance on some previous publication in that the authors have used a higher resolution SNP array but the number of cases analysed (only 12 cases were used for the integrated analysis.) is much smaller than some more recent published studies. The small number of cases severely limits the significants of the manuscript and the relevance of the conclusions.

Major compulsory revision.

1. Throughout the manuscript there is confusion as to the number of samples used for the integrated analysis. In the abstract it is stated that data from 22 cases was integrated but in the methods (page 7) it states that only 16 samples were analysed for gene expression. Furthermore in the results only 12 cases were said to have had matching SNP and expression data (page 11). In addition on page 12 the authors state the copy number data for these 12 cases was somewhat different to the copy number data for the 22 samples. Does this mean that these 12 samples were completely different to the 22 samples or just a subset of the 22 samples? Either way the number of samples actually used n the integrated analysis needs to be clearly articulated and this needs to be reflected in the abstract.

2. Four control breast samples were used to compare gene expression but no details are given as to where these cam from and whether they are appropriated. For examples were these from pre or post menopausal women and was the breast epithelium micro dissected as per the tumour samples.

3. As pointed out by the authors, an advantage of SNP arrays is that it can be used to generated LOH information which can detect regions of copy number neutral LOH. However, since the authors did not used matching normal DNA, much of the potential high resolution is lost though having to use a statistical approach to predict LOH. It is the view of this author that the analysis should be repeated using matching normal DNA.

4. The authors detected a remarkable low number of regions of copy number loss (page 8) and state that this is often reported I the literature (page 16). In fact
my reviewing of the literature suggests that losses are almost as common as gains and this is certainly reflected in the publication by Andre et al (reference 14) where an equal number of regions of gain and loss are reported. The authors need to investigate this discrepancy more thoroughly by reviewing the literature to assess the proportion of gains and losses detected in previous studies. It would also be useful to see some actual examples of the 250k copy number profiles generated in this study. As it stands I am suspicious that there is some systematic bias in the data.

5. The description for the integration of the LOH and CNA data is unclear. The author state that they looked for overlap of CNAs and LOH but surely all regions of copy number loss should correlate with regions of LOH? (subject to the accuracy of the non paired LOH analysis). This section needs much more attention to describe more clearly what was done.

Minor essential revisions
1. The text in Figure 2 and Figure 5 are very pixelated and cannot be read clearly.
2. The text needs to be proof read as there are at least several places where sentences don’t make sense.

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: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests