Reviewer’s report

Title: A methodology for utilization of predictive genomic signatures in FFPE samples

Version: 3 Date: 12 April 2011

Reviewer: Bin Tean Teh

Reviewer’s report:

The work presented in the manuscript is worthwhile since the ability to extract biologically useful data from FFPE samples will open up many research opportunities. As such, understanding what genome-wide information can be obtained from these samples and the reliability of these data is important.

All mathematical and computational methodology presented is sound and appropriately applied. Methods are well described and the manuscript is well written.

The manuscript lends some support to idea that the activation of oncogenic signatures can be assessed through data generated by the MessageAmp Premier methodology. However, the author’s give little discussion of the limitations of the paper.

Firstly, why did the authors limit themselves to the RAS and MYC signatures? To use so few makes it difficult how broad the application of the FFPE data could be. It would be interesting to explore a broader array of signatures and correlate their activation. This would provide a much more solid basis for generalising and concluding that ‘Reliable and consistent predictions of oncogenic pathway activities can be obtained from FFPE tumor tissue samples’.

Furthermore, perhaps this is beyond the scope of the paper, but another interesting bioinformatic experiment would be to calculate enrichment scores for various defined canonical pathways (e.g. KEGG), not solely oncogenic signatures. The scores for FF and FFPE could then be clustered as was done at a gene level in the manuscript (Figure 2). This would make it possible to assess whether the data are more concordant on a pathway level.

The previous suggestion is born out of the observation that the FF and FFPE data do not show concordance at the gene level but do at the signature level (at least for RAS and MYC). It may be true that biological pathways generally show greater correlation than genes.

It would be nice if the authors could offer some discussion of why there should be such a discordance between gene level data in FF vs FFPE but such a strong concordance at the level of the oncogenic signatures. There is no mention of this in their conclusions.
What was the motivation of switching between MAS 5.0 normalisation and RMA in different analyses?

How long were the FFPE samples stored? In the conclusion of the manuscript, the claim that stored FFPE tissues could be used in studies where the clinical outcome is already known, comes with the implicit assumption that they will have been stored for a reasonably long time. It would be nice to know how the length of time stored affects the RNA degradation and genomic profiling via MAP, at the very least in terms of QC statistics such as those quoted in this paper (GAPDH ratios etc.). Essentially, can reliable data be extracted from tissues stored long-term? The question of broad reliability is an important facet here due to the problems with the gene level data and poor QC metrics for the FFPE samples. The conclusions section could benefit from some discussion of the limitations of the study.

Practically speaking, when analysing data, there would have to be strict guidelines on what conclusions could be drawn and reliability of the data - especially as regards the dichotomy of gene level and pathway level data. As a result, it would be beneficial for the authors to discuss the limitations of gene level data in their conclusions and give some idea of precisely the type of reliable data that can be extracted.

The data presented is encouraging and suggests it may be possible to extract some more broad biological info from FFPE tissues than was previously possible. However, looking into only the RAS and MYC signatures does not permit any conclusions about the wider array of major biological pathways or, indeed, how consistent and reliable the MAP data is over a broad range of oncogenic signatures. Specifically, it does not yet point to a method to extract the 'quality whole-genome expression data' mentioned in the conclusions as a foundation for multiple novel avenues of research.

**Level of interest:** An article of outstanding merit and interest in its field

**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