Reviewer's report

Title: Metabolomic Profiles of Hepatocellular Carcinoma in a European Prospective Cohort

Version: 1  Date: 10 June 2015

Reviewer: Daniel Raftery

Reviewer's report:

The authors perform an impressive NMR based metabolomics study in a moderately large, nested case-control set of samples that were collected from multiple sites in a prospective manner. There are a number of positive elements to this study:

There is a strong need for diagnostics related to liver cancer, which is a global problem.

The prospective sample set in this study is very well characterized and well matched with respect to cases and controls. The multiple collection site nature of the study avoids many of the problems that other studies suffer from during validation.

The study is focused on early HCC patients, which are the ones that can be more successfully treated.

NMR is a stable platform, and the authors try multiple statistical approaches which seem to give similar results in that the performance is roughly in line with AFP, the current blood marker used to monitor liver function and HCC treatment.

It’s great to see that the statistical analysis is quite thorough.

However, there are several negatives as well:

I am concerned that the CPMG experiment may be inappropriate for measuring lipids. CPMG is sensitive to the 1H relaxation time and since it is very possible that samples from different patients or acquired at different times of the day will vary in viscosity (due to the water content, or protein content of that particular serum sample), then this will affect the apparent lipid signal size. These confounding effects will cause problems for validation. I would suggest that the authors add/analyze 1D NOESY data if they did acquire it. Otherwise, I would strongly suggest that they delete the lipid peaks in the CPMG spectra. Maybe the authors have some convincing data that shows the linearity of the CPMG data over different lipid relaxation times, or its insensitivity to viscosity changes, but in my opinion this situation is potentially problematic. We have encountered lipid intensity variations that are exaggerated in the CPMG spectrum compared to the 1D NOESY and typically eliminate the lipid region as a result. These days, we actually prefer protein precipitation (See Gowda, Anal Chem, 2015).
On a somewhat related point, some of the loading plots (OPLS coefficients) show the lipid signals to have dispersion lineshapes, which is often indicative of peak shifts. In this case, the lipid lineshape is inhomogeneous and thus the resulting changes may be related to the bias of the CPMG experiment in the lipid region. For example, the OPLS coefficient plot in Figure 3a and 3c and 4a show this situation.

The authors should make sure that these are not due to peak shifting or related to CPMG bias as per the previous point.

I’m also concerned about the presence of propylene glycol primarily in the HCC population. If this is indicative of medication use, then doesn’t this confound the results? Is the model the same by taking these patients out? The authors mention removing the peak in the supplemental, but PG appears in Table 4.

The authors use a statistical reduction method based on correlation that they developed some years ago. As a result, only 16 metabolites pass a test of significant after the statistical reduction. How do the markers compare if one just uses metabolite peaks they can identify in the spectra?

Minor: It would be appreciated if the authors could change the symbol shape in either their case or control data. The use of the same symbol makes it impossible to differentiate them when printed in black and white. Similarly for the ROC curves. I know we’re all never supposed to use paper anymore, but still.

If the authors can satisfactorily address these issues (especially the CPMG related issue) then I would be in favor of publication.

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**

I hold equity interest and an executive role in Matrix-Bio, Inc., which is an early stage company focused on cancer biomarker development. In addition, we have performed and continue to perform liver cancer biomarker discovery in my academic laboratory, and have published 3 papers on this topic, mostly using different methodologies. We also have applied for IP related to HCC detection.