Reviewer’s report

Title: Serum diagnosis of diffuse large B-cell lymphomas and further identification of response to therapy using SELDI-TOF-MS and Tree Analysis Patterning

Version: Date: 4 May 2007

Reviewer: Antonia Vlahou

Reviewer’s report:

General
In the manuscript by Zhang et al, SELDI-TOF-MS profiling is being performed on serum samples of patients with diffuse large B cell lymphoma (DLBCL). Decision trees are being generated using the Biomarker Patterns Software which can discriminate with high accuracy, in respective blinded test sets, DLBCL from controls, poor prognosis from good prognosis and relapse versus non relapse DLBCLs. In general the study is carefully conducted and although the authors do not report the identity of the discriminatory peaks, there is interest due to the clinical problem (DLBCL diagnosis and prognosis) it addresses.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)
1) SELDI-TOF-MS analysis has been employed in several cases by now for the diagnosis/prognosis of various diseases in serum and urine. In all cases, high accuracy rates in disease detection were received. Nevertheless, the published studies also revealed that the technique faces various reproducibility problems and is prone to artifacts. There is lack of discussion on these issues in the manuscript. It is recommended that the authors discuss the reproducibility aspects of the SELDI profiling assay, its employment as a “biomarker discovery” versus “pattern analysis” tool, how reproducibility problems are being addressed in recent studies and also how the authors addressed some of these issues in the presented study.
2) The authors report the statistically “top-scored” peaks and also the peaks utilized for the generation of the decision tree. There is very limited overlap between the two ie, only 2 peaks found to differ significantly were employed by the algorithm for the generation of the decision tree. The authors should try to explain/discuss on this issue.
3) Were the samples positioned randomly on the chips? Randomization is very important when performing this type of analysis, as shown by multiple studies. This information is missing from “materials and methods”.
4) How many peaks (or peak clusters following the second pass peak selection) in total were detected?
5) It is not very clear how the reproducibility study was conducted. The authors report that they “compared 10 selected M/Z peaks from an unaffected case study”. Do they mean that they run one control sample multiple (how many?) times? Also were all samples analyzed in duplicate?

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
1) In “methods” the authors mention that they applied the sample on H4 chips and they describe the respective technique whereas in the “results” they report that they employed WCX. This discrepancy should be clarified.
2) In the abstract the authors should mention that the accuracy rates they provide refer to the classification of the blinded test sets.
3) “A proteomic data set reported by Adam et al was used to classify the samples in this study” The meaning of this sentence is not clear. Do the authors want to say that they employed the experimental protocol described in the study by Adam et al?

Discretionary Revisions (which the author can choose to ignore)

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major
compulsory revisions

**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