Reviewer's report

Title: Discriminating lymphomas and reactive lymphadenopathy in lymph node biopsies by gene expression profiling

Version: 2 **Date:** 11 October 2010

Reviewer: Patricia JTA Groenen

Reviewer's report:

This manuscript describes the application of gene expression profiling in de diagnostics of lymphoma. The group used fresh lymph node specimens from different subtypes of B-cell lymphomas (Classical Hodgkin, FL and DLBCL) and reactive lesions and evaluated the accuracy for classification using gene expression data. A training set and test set was used, and the obtained data were analyzed using 3 different statistical approaches to classify the lymphomas. The classification accuracy rates, in this study of 112 tissue specimens, were between 68.6 – 88.5% (dependent on the comparison), which will not be acceptable for clinical application right now, but larger studies will be needed to improve the analysis. In general, it is a well-written and analyzed study . There are however, some concerns that have not been addressed by the authors and should be included in the revised version.

Minor Essential Revisions:

-The authors should provide information on the size/amount of the tissue specimens that is needed for the analysis. In addition, how much RNA is used in the analysis? Did the authors check for RNA degradation in the samples? Or was there another quality control step to exclude samples from the analysis?

-More information on the correction for batch effect of arrays is needed.

-I would like to see more details on the samples used in the study, to rule out misclassification effects due to: 1) the hospital were the samples came from 2) grading of the FL-cases and the GC/ABC type of DLBCL cases 3) tumor load. Was there a minimal tumor load of the lymphoma cases in this study?

I assume that well-defined and classified cases are included in the present study. Would this technology be able to successful discrimination of cases that are difficult to classify by a pathologist? Did the 5 GC- cases of GCB cell type have the BCL2 translocation and /or the BCL6 translocation?

-The accuracy of RL versus lymphoma cases is limited (80% in the test set). As such, there is a false positivity of 20% of patients receiving the diagnosis of lymphoma instead of having a reactive lesion. And 20% of patients who have a reactive lesion which would be molecularly diagnosed as lymphoma. Can the authors provide more details on the reactive lesions, which may partly explain their findings. How can this accuracy (bening-malignant) be improved according

to the authors?

-More information on the 31 (of the 38 annotatable) genes that are used for classification

Discretionary revisions

- What is the future perspective for this approach? Would this approach be valid also for consultation cases, coming from another hospital, also fresh frozen samples.

- The lower expression of particularly the light chain immunoglobulin genes in reactive lymph node is intriguing. Do the authors have information on the light chain use in their reactive lesions and the lymphoma samples, that may explain this phenomenon.