Reviewer's report

Title: Gene-expression of metastasized versus non-metastasized primary head and neck squamous cell carcinomas: a pathway-based analysis

Version: 1 Date: 21 November 2007

Reviewer: Thomas Belbin

Reviewer's report:

General

The manuscript by Hensen et al. describes experiments to elucidate differences in gene expression between metastasized and non-metastasized HNSCC primary tumors. As stated in the manuscript, regional lymph node metastasis is an important prognostic factor in HNSCC, and plays a decisive role in the choice of treatment. The authors use a publically available dataset previously published by Roepman et al as a reference dataset, and their own independent gene expression dataset. The results of their pathway-based analysis are interesting; however, several concerns remain.

--------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

- More information should be included on the reference dataset in the “Methods” section, specifically a brief breakdown of clinical information (tumor site, stage, follow-up, etc.) and how this compares with the validation set.

- Within the results, the authors describe a “tendency to validation” for 7 of 21 genes. What does this phrase mean, and how is this determination carried out?

- In the Discussion section, the authors state that the outcome of array studies is heavily dependent on technical aspects such as the choice of microarray platform. This is a strong statement, particularly in light of recently published papers by a large (MAQC) Consortium on the reliability of microarrays in which the authors actually report a high level of interplatform concordance (Patterson et al. 2006; Shi et al. 2006). Moreover, the second of the two papers in the edition refers to a previous study that reported “little overlap among lists of differentially expressed genes derived from three commercial microarray platforms when the same set of RNA samples was analyzed” (Tan et al. 2003), which would contradict the above statement. In light of this, the authors should include references to studies that support their conclusions.

- In the Discussion section, the 7 genes identified through their gene-based analysis are discussed in light of the current literature, but no attempt was made to validate the observed differential gene expression in their validation dataset using an alternative technology such as real-time PCR or tissue microarray.
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

- Please remove the last sentence of the section on “Reference Dataset” in Methods.

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:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests