Reviewer's report

Title: Association between Helicobacter pylori cagA-related genes and clinical outcomes in Colombia and Japan

Version: 2 Date: 1 September 2011

Reviewer: Alex Boussioutas

Reviewer's report:

This is a correlation study that reports an association between certain H. pylori genes and Gastric Cancer in a subpopulation of Columbian patients. The authors selected genes from previous microarray experiments that were correlated with cagA status and used a subgroup of these genes to assess expression of those genes in different patient pathologies in Columbian and Japanese cohorts.

I have a number of questions regarding the manuscript.

Major compulsory revisions

1. This addresses virulence factors in the bacteria but the complex interplay between host and pathogen are not really addressed in this analysis. Were the populations studied tested for established host polymorphisms that increase the risk of GC (eg IL-1R)?

2. I have an issue with the methodology used for selection of genes. Microarray analyses involving correlations require correction for multiple testing. The p values reported here do not appear to be corrected for multiple testing. This would reduce some of the significance of the genes reported as correlated to cagA status. If you perform appropriate multiple testing correction do these genes still reach significance?

3. I think it is insufficient to perform a multivariate analysis of age and gender for risk of GC in these populations. There are other risk factors that could be incorporated that are more relevant for risk of GC (see question 1).

4. It would be relevant to ascertain whether other established H. pylori genetic markers of risk as mentioned in the manuscript are also relevant to risk in these cohorts. The results would serve as a positive control if they indeed correlated as the authors suggest they would.

5. This study seems to investigate presence of genes, however gene expression would be more relevant. Does the level of expression of the genes being studied have biologically significant expression differences? There is no data presented about fold change of gene expression only that there was negative or positive association with cagA status. Whereas a correlation may be statistically significant between genes the expression difference may be marginally different and hence have little biological meaning.

Minor essential revisions
6. There are a number of grammatical errors throughout the manuscript that would need to be addressed

**Level of interest:** An article of limited interest

**Quality of written English:** Not suitable for publication unless extensively edited

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests