Reviewer's report

Title: STAT3 polymorphism and Helicobacter pylori CagA strains with higher number of EPIYA-C segments independently increase the risk of gastric cancer

Version: 3 Date: 20 April 2015

Reviewer: Douglas Scott Merrell

Reviewer's report:

In the revised manuscript, “STAT3 polymorphism and Helicobacter pylori CagA strains with higher number of EPIYA-C segments independently increase the risk of gastric cancer”, the authors did an excellent job of addressing the prior concerns of the reviewers. Overall, the manuscript is interesting and appears to provide novel information about a different population than previously investigated. Furthermore, the authors looked at both host and bacterial genetics as potential indicators of H. pylori disease progression. Overall, the paper is significantly improved. However, there are still a few concerns to be addressed.

Major Compulsory Revision:

1. The authors state that distilled water was used as a negative control for qRT-PCR. It is the understanding of this reviewer that the distilled water was used in place of cDNA in the qRT-PCR reactions. While this is a negative control, it merely rules out gDNA contamination of the reagents used in the qRT-PCR reaction. In order to rule out gDNA contamination of the cDNA the authors should conduct a cDNA synthesis reaction including all components except the reverse transcriptase (RNA, primers, buffers). This “no RT control” reaction is then used as a template for qRT-PCR. Since the initial reverse transcription reaction lacked the enzyme there should be no cDNA template in this reaction, therefore there CT values from this qRT-PCR should be very large. Without running a “no RT control” the authors cannot rule out gDNA contamination from the RNA/cDNA templates. This is important because gDNA contamination would lead to artificially low CT values and subsequently inflated expression levels for the gene of interest.

2. On lines 145-148 the authors state, “endoscopic biopsy samples... were obtained for histological and microbiologic studies”. Included in this list is the 13C urea breath test, which cannot be performed on an endoscopic sample. Was this performed on the patients prior to endoscopy? Please clarify.

3. Please define in the figure 1 legend what inflammation and activity refer to. In the text line 392 please clarify how the data regarding PMNs and mononuclear cells relates to figure 1. Additionally figure 1 shows activity however, it is not referenced in the text. What do the numbers above the bars indicate (see below)? Overall, I find this figure incredibly confusing and really couldn’t understand what it was meant to show. If possible, it would be better to find another way to show this data in a more user-friendly fashion for the reader.
4. Figure 1: The values above the columns do not correspond to the values listed on the y-axis. For example, under inflammation score 3, there are two columns both with 32 indicated above; however the columns are not the same height. Furthermore, under activity score 3 there is a 22 above a column that is higher than the adjacent column labeled 42. Please label what percent refers to in the y-axis and what the number above the columns represents.

5. Figure 2. Please define the numbers on the y-axis.

6. Line 402: the phrase “rate of pSTAT3 expression was significantly higher” is misleading. As described, the staining evaluates the percent of cells with pSTAT3 staining, which does not evaluate rate. Please add the definition of the scoring system to the figure 3 legend. Amend the figure 3 y-axis to indicate pSTAT3 as opposed to STAT3.

Minor comments:

1. Please be consistent in the spelling of tumor throughout the document (for example: lines 258 and 259). Also, please correct the spelling of lipopolysaccharide (Line 49) and define SNP on line 92.

2. Line 106: Should be designated not designed.

3. Line 164: Please indicate Table 1 after the cagA gene primers are listed.

4. Line 228: it is unclear if both LPS and Pam3Cys are both diluted in RPMI.

5. Line 263-266. Please separate into two sentences: To retrieve antigenicity, the sections were placed in 10 mmol/L citrate buffer solution, pH 6.0, and heated in a microwave for 12 min. Sample were then treated with 3% hydrogen peroxide-metanol for 12 min to block endogenous peroxidase and, rinsed with distilled water.

6. Lines 369 and 378: please change to logistic regression

7. Lines 385-408. Consider combining the text referring to figures 1-4 under a heading similar to evaluation of STAT3 rs744166 polymorphism

8. Line 469-471. “Remarkably, we found that in patients, with gastric cancer predominates the rs744166 GG genotype and infection with CagA strains with higher number EPIYA-C segments.” Please remove comma and move predominates to the end of the sentence otherwise it creates a sentence fragment.

9. Consider reordering figure 4 before 3 showing the staining followed by the quantification of that staining.

10. Please describe figure 4 in the text.

11. Tables 2, 3, and 5 all have superscripts that appear as “-ve” or “+ve” next the “Hp” and “CagA” designations. I’m unsure of what these stand for. What is the “ve”?

Level of interest: An article of importance in its field
Quality of written English: Needs some language corrections before being published

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

Declaration of competing interests: I declare that I have no competing interests