Reviewer’s report

Title: The MLH1 2101C>A (Q701K) variation is related to risk of gastric cancer in Chinese males

Version: 2 Date: 12 August 2011

Reviewer: Eugenia Flores

Reviewer’s report:

The authors perform an assessment of the association between mutations in the DNA repair gene MLH1, with the risk of gastric cancer in Chinese population. The study was conducted in 236 patients with gastric cancer and 240 controls matched for age and sex of the provinces of Jiangsu and Anhui. The question is clear the authors try to determine the significance of the association between mutations in the MLH1 gene with gastric cancer.

Major Compulsory Revisions

RESULTS

1. Table 1, is important to have information on family history of cancer in the controls, and if possible, data from other exposure factors for both groups, which allows for evaluation of potential differences between cases and controls, useful information for confusion assessment by statistical models.

2. Authors should indicate the distribution of cases and controls by province, and if possible by ethnicity, for statistical evaluation of confounding by population stratification, and to avoid bias.

3. In the literature (Kleinbaum and Klein, 2002) suggests three forms of statistical analysis to the case-control studies matched: 1) Stratified analysis, 2) McNemar analysis, 3) Logistic modeling.

One way to perform the stratified analysis is by using a chi-square Mantel-Haenszel (M-H) when studying pairs (case-control) with similar characteristics, where each pair forms a stratum. In this study indicate having studied 236 cases and 240 controls matched for age and sex data shown in Table 1 (p = 0.997 p = 0.915, respectively), however, if 1 case, matched with 1 control, have 236 or 240 pairs, I guess that the information in table 1, where we have 178 cases for men and 58 women and 180 control men and 60 women, failed the 1:1 ratio, making it difficult statistical analysis from the perspective of paired data. The odds ratio (OR) of M-H is an estimate of summary adjusted for matching variables. The McNemar test is exactly equal to the M-H chi square; both are based on couples and not in individuals for the case-control studies matched. Multiple logistic modeling is a suitable method when you need to control for variables not included in the matching, as could be: family history, ethnicity or place of origin, alcohol consumption, smoking, etc, variables that are potential confounding. Conditional logistic regression model is appropriate for
paired data (when you have partners like). When this is not possible, an alternative is to use non-conditional logistic regression, modeling appropriate to be used to present the ORs, CI and p value in Table 2, statistical models should be adjusted for matching variables (age and sex), and other variables not considered in the matching, such as family history, smoking, alcohol consumption and place of origin, or other factors. In the discussion referred information of drinking and smoking in men, is striking that these data have not been incorporated into the results.

4. If possible, using the above logistic model, to evaluate the interaction between the sex variable and the genotypes of MLH1 2101C>A, which is the association that seeks to demonstrate in Table 3.

5. Using the statistical program STATA v.11.1, I performed the calculations of ORs controlled by the gender with the information in Table 3, and the M-H test for homogeneity of the results, indicate that not there is sufficient evidence that the ORs are different, so we cannot conclude that gender is a real effect modifier. Similar results are obtained by unconditional logistic regression. On the other hand, the amplitude of the confidence interval (1.04 - 68.06) of the M-H OR in men (Table 3) indicates uncertainty of the estimator and therefore can be thought of an estimator biased or confused. However, one cannot exclude that the association is real. The intervals may be more accurate with a larger sample size

6. I believe that Table 3 should be eliminated or replaced by the results of logistic regression. One of the limitations of stratified analysis is that there is loss of statistical power, increasing the type 1 error.

7. It is recommended that the results of Table 4, are made by haplotype analysis, and also perform the calculation of linkage disequilibrium. Analysis of current and relevant in the field of genetic association studies. There are several programs in which you can perform this analysis: Haploview, Haplo Stats, SNPStats, etc.

8. Authors should evaluate the Hardy-Weinberg equilibrium in controls for SNPs

DISCUSSION

9. Authors should restructure the discussion.

Minor Essential Revisions:

1. In methods: instead of normal use to refer to patients without cancer, using the word control in the first paragraph of clinical samples .... A total of 240 normal age-matched and sex individuals ... (similar situation in the first paragraph of results).

2. In the third paragraph of methods (Mutation screening), I consider it important to indicate the Primer sequences, and a summary of the conditions of PCR, so that results can be replicated anywhere.

3. The discussion in the penultimate paragraph indicate:......However, it was only marginally associated with the riesk of gastric cancer (p=0.069), this value cannot indicate that there is a marginal association, the limit value set by the authors was 0.05. It is clear that age is a confounding variable. The risk of developing
any type of cancer increases with age.

4. Strengthen the discussion on the limitations of the study, related to the low frequencies of SNPs and mutations found, as well, the sample size.

5. At the end of the conclusions instead of…..it has been demonstrated….. that the MLH1 2101 C>A mutation may contribute to an increased risk of gastric cancer in males. Changed to: it was found that the MLH1 2101 C>A mutation ..........

Referencia


**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**

I declare that I have no competing interests