Reviewer's report

Title: A case-only study of gene-environment interaction between genetic susceptibility variants in NOD2 and cigarette smoking in Crohn's disease aetiology

Version: 1 Date: 4 December 2011

Reviewer: Inga Peter

Reviewer's report:

This is a nicely done study that applied a not commonly used statistical approach, departure from multiplicativity, to estimate a genotype-environment interaction in Crohn’s disease patients. While the reported findings are intriguing, there are several issues that should be clarified and addressed in order to improve the paper:

Major Compulsory Revisions

1. Analyses should be limited to individuals with genotype data for ALL three NOD2 variants; otherwise individuals with missing genotypes can be misclassified with regard to their NOD2 carrier status.

2. There are significant differences (almost 2-fold) in the proportion of individuals with family history of CD between participants with and without data on smoking status at diagnosis. How would the authors explain this bias? Family history was not included in Supplemental Table 2 that compared other traits between included and excluded CD cases.

3. The authors acknowledge that age of disease diagnosis and family history may both confound the observed results, however only age at diagnosis was adjusted for. Family history should be added to the regression model and tested for confounding.

4. On page 7, line 6, the authors state that “……and family history of IBD were not associated with genotype for any of the three risk variants (Supplementary Table 1).” However, they can’t ignore the fact that carriage of R702W was associated with family history (p=0.007) even in the context of Bonferroni correction. Is it possible that the best explanation for the observed findings is that carriers of the risk alleles are more likely to have family history of IBD and therefore more aware of the risk of smoking?

5. Was the year of recruitment, added as a covariate to the regression model, significant? Has the rate of smoking changed between 1995 and 2010 when the study cohort was recruited?

Minor Compulsory Revisions

1. It would be helpful if Table 1 included P-values for comparison between the two groups of smokers and non-smokers.
2. Tables 2 and 3 should add % of individuals in each category to make it easier to follow.

3. The sentence “The observed interaction OR between history of ever having smoked and the 1007fs allele was one-third lower than expected and nearly 50% less than expected when smoking status at the time of diagnosis was analyzed” (Discussion, page 8) should be rewritten to clarify to a general reader what lower than expected interaction means.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.