Reviewer's report

Title: Low-penetrance alleles predisposing to sporadic colorectal cancers: a French case-controlled genetic association study

Version: 3 Date: 23 July 2008

Reviewer: Victor Moreno

Reviewer's report:

The authors have successfully addressed the concerns of previous version, but a few issues remain:

1) Conditional logistic regression. I think the authors have just added to the text that "conditional" logistic regression (CLR) was used to account for the matching, but the actual analysis was done with unconditional logistic regression. I think so because if CLR is used, you cannot further adjust for age and sex if these are the matching variables and all cases and controls without matching pair have to be excluded and there is no mention to that and the final number of matched pairs.

2) Multiple testing. Authors have followed one of my suggestions, the estimation of FDR. They have opted for the q-value approach of Storey. I have plugged the p-values from table 1 into the qvalue program and I get a q-value of 0.46 for the most significant p-value (0.003). My interpretation is that even this association (PPARG, recessive model) has about 50% chances of being a false positive result. Nothing to say about less significant results. A better explanation in a supplement of how they reach the cited q-values (0.03 and 0.005) would be useful.

3) Combination of genotypes - Table 2. I think I overlooked a serious problem in previous review. Why the presence of minor alleles for PLA2G2A and PPARG means a mixed pattern and the homozigous common alleles is not informative to differentiate from protective (reference) or predisposing. If minor alleles for these SNPs are strongly protective (individual OR < 1) they should predict protection but not predisposition.

This analysis of the combination of genotypes, that might be the only original contribution of this paper, is not well explained in methods, the results are paradoxical or there are typos in table 2, and we don't know anything about the performance of alternative models to the one proposed. The method used seems to be prone to extreme over fitting and there is no validation. The authors cannot rely in others to validate their proposal. Nobody is going to try it without some prior information that this model has some theoretical external validity. At least an estimate of that using bootstrap or cross-validation should be provided. The consistency analysis is not very useful for that and seems more a failed trial to find subgroups of more significant results.
Level of interest: An article of importance in its field

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