Reviewer's report

Title: Do genetic factors protected for early onset lung cancer? - A case-control study before the age of 50 years

Version: 3 Date: 5 November 2007

Reviewer: Sara Raimondi

Reviewer's report:

The authors have addressed many highlighted issues and have solved part of my major criticism on the work. However, I feel that the result section is still confusing due to the large amount of ORs presented for the same variables, especially for smoking.

-------------------------------------------------------------------------

MAJOR COMPULSORY REVISIONS

1) The authors answered to my comment on redundancy stating that “the genes investigated are related to substance metabolizing, tumor suppression or DNA-repair, one might expect any association to be modified by the strength of tobacco exposure. Hence we avoid to estimate a single OR across all groups of smokers”. I agree to include ONE OR for EACH smoking level (moderate, heavy smokers), but I disagree to provide MORE than one ORs for EACH smoking level. On page 11, lines 3-4 from the bottom, OR for moderate smokers is 15.9, for heavy smokers 65.9. Therefore, moderate smokers have a 16-fold higher probability to develop lung cancer comparing with never smokers, and heavy smokers have a 66-fold higher probability. These ORs were obtained by logistic regression model including sex, age, smoke and GPX1. On page 12, lines 13-14, the OR for moderate smokers is 36.4, for heavy smokers 49.79. Therefore moderate smoker have now a 36-fold higher probability to develop lung cancer comparing with never smokers, and heavy smokers have 50-fold higher probability. These ORs were obtained by a logistic regression model including only sex and age as covariates. Do the authors really think that it is important to know that OR for moderate smokers is 16 adjusting by GPX1 and 36 adjusting by EPHX1? The authors should discuss the implications of these different results if they think they are important. Again, my suggestion is to provide only ONE OR for moderate smokers and ONE for heavy smokers, obtained by a logistic regression model including only sex and age as covariates.

2) Also the OR for GPX1 is 0.3 (0.1-0.8) for heavy smokers and 0.6 (0.2-1.6) for moderate smokers in the model II with only one gene. The OR for GPX1 is 0.3 (0.1-0.8) for heavy smokers and 0.4 (0.2-1.3) for moderate smokers in the model II with two genes. I found these ORs pretty equals within the same SEL, therefore I think it is not useful to present all of them. From the results, it seems to me that EPHX1 is not a confounding factor for GPX1-lung cancer association,
therefore I think the authors could present the main effect of GPX1 using ORs obtained by the model including only one gene. The two-genes model is useful to discuss the combined effect of the two genes.

The OR for EPHX1 is 0.2 (0.1-0.7) for moderate smokers and 0.8 (0.3-1.9) for heavy smokers in the model II with only one gene. The ORs for the same gene are 0.3 (0.1-1.1) for moderate smokers and 0.7 (0.3-3.5) for heavy smokers. Again, what is the reason to present both the results from the single-gene and the two-genes models?

I suggest that the authors delete lines 1-9 from page 13 and start the paragraph “GPX1 and EPHX1” with line 10 “As a combined effect of these two polymorphisms”.

3) Please, also delete lines 15-16 on page 13 “This indicates a strong….”. As from the result of the test, the interaction was not significant.

MINOR ESSENTIAL REVISIONS

Table 1: Sex (m/f) instead than (m/w)
Adenocarcinoma instead than adenoma carcinoma
Table 2: The table is not cited in the text
Table 3: The results for CYP1A1 are missing
Table 4: Note: SEL: never and light smokers

Discretionary Revisions (which the author can choose to ignore)

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

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

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