Reviewer's report

Title: Genetically protected young smokers? - A case-control study on lung cancer before the age of 50 years

Version: 1 Date: 12 September 2007

Reviewer: Sara Raimondi

Reviewer's report:

Tobacco smoke is the major risk factor for lung cancer, however genetic factors may play an important role at low dose carcinogen exposure and at younger ages. The aim of the present study was to investigate the role of candidate genes in lung cancer susceptibility in patients with age 50 years or younger. This is an important issue, since it is not known whether gene polymorphisms associated with lung cancer are the same for all age groups, and identifying genetic factors leading to early onset lung cancer could potentially yield strategies to postpone the onset of this tumor, or help formulate new therapeutic approaches.

The study design and methods used for the analysis were appropriate and well described, however methods for SNPs selection should be clarified. I found the presentation of results somewhere confusing. I suggest that authors concentrate the description and discussion only on the most relevant results.

Specific comments follow.

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

MAJOR COMPULSORY REVISIONS

1) METHODS, Selection of Candidate Genes and SNP section: A clearer explanation of the criteria used to chose the study genes should be provided by the authors. They wrote that it was based on “two criteria: published significant association together with plausible biological relevance for lung cancer”. In fact are they performed a Medline search on gene polymorphisms and lung cancer? How many genes did they find? How many genes were excluded because of no evidence of an association with cancer and/or no biological plausibility? How do they choose the final list of 11 SNPs (i.e. the most studied, the most biological relevant, the ones with the strongest association with lung cancer,...)?

2) METHODS, Statistical Methods section: the definition of never smokers seems not in agreement with standard (see World Health Organization. Guidelines for controlling and monitoring the tobacco epidemic. Geneva, Switzerland, 1998.) Please clarify why you chose this classification. Anyway I suggest that the reference group is called “never and light smokers”, since it includes subjects who smoked up to 1 PY. Moreover, clarify in the text and in the Tables whether subjects smoking 1 PY were included in the group of never or moderate smokers.

3) RESULTS: Authors seem to report different ORs for smoke, calculated in the
different models. This is confusing: only one model should be chosen and only
the most likely OR for smoke should be reported. I suggest using a model
included only sex and age as covariates;

4) RESULTS, GPX1 and EPHX1: Presenting different ORs for the same gene
within the same strata (i.e. heavy smokers, current smokers, …) is very
confusing. As for smoking, only one model should be chosen and the
correspondent ORs reported. I would use the one-gene models to provide ORs
for each single gene and the two-genes model to describe the combined effect of
the polymorphisms (genetic Prediction Score);

5) RESULTS, GPX1 and EPHX1 section, lines 12-15: I cannot understand what
is the reference group for this ORs and what is the exposure (at least one
protective allele???). Please clarify.

6) RESULTS, GPX1 and EPHX1 section, line 15: A p-value for the test of
gene-smoking interaction should be given and the test should be described in the
Methods section;

7) DISCUSSION section, last line: The sentence on gene-smoking interaction is
not supported by the results (see previous point);

8) TABLE 1: Please clarify whether there is a different classification of smoke
exposure for cases and controls. If it is the case, a differential bias could occur in
cases and controls, therefore I suggest that authors carry out a sensitivity
analysis classifying cases and controls in the same way;

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

MINOR ESSENTIAL REVISIONS

1) TITLE: The question “genetically protected young smokers?” seems to refer to
a study group included only young smokers, which is not the case of the present
study. I suggest a different title, i.e. “do genetic factors protect from early onset
lung cancer?- A case-control study before the age of 50 years”;

2) METHODS, Statistical Methods section, page 8, lines 4-5: Please, add a
reference for Hardy-Weinberg equilibrium test;

3) METHODS, Statistical Methods section, last line: Please add a line for the
assessment of multiple comparison (Bonferroni correction);

4) RESULTS, Smoking Habits section: delete the last 2 lines on page 9 and the
first 2 lines on page 10: this result is obvious since pack-years calculation based
on the years of exposure, that are therefore likely to increase with increasing
age;

5) RESULTS, both GPX1 section and EPHX1 section: It seems that authors
calculated ORs both for smoking habits (former vs never smokers/ current vs
never smokers) and for “current smoking behaviour” (current vs never+former
smokers). I suggest that ORs for current smoking behaviour will be deleted, since
it does not add any significant information and it is not appropriate to group never
and former smokers-is well known that lifetime smoking increase the risk of
cancer. Anyway, reporting all the results is confusing (lines 12-13 and lines 21-23
on page 11 appeared the same);

6) RESULTS, GPX1 and EPHX1 section, page 12, last line: from table 5 it seems that never smoking cases with gPS>2 are only two. Please correct or clarify.

7) REFERENCE LIST: The references are full of mistakes - i.e. refs 2 and 4 are equals, refs 28 and 39 are equals, refs 56 and 58 are equals, refs 65, 66, 67 and 68 are equals, ....

8) FIGURE 1: It is not easy to distinguish the different lines in the figure and the filled symbols;

9) TABLE 1: There is a typo for male controls, 41-45 years: percentage could not be 372%;

10) TABLE 3: Report the crude ORs beyond the p-values. Add “p-values for” Baumgartner-Weiss-Schindler test in the title;

11) TABLE 5: Use the same groups for SEL and all smokers: “2” could be always grouped with “3 and 4”, or always considered as a separate category.

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

DISCRETIONARY REVISIONS

1) You do not need to report both OR with 95% CI and p-values. I suggest deleting p-values throughout the text when 95% CI were already reported;

2) DISCUSSION: I would move the section “some remarks to the study design” at the end of the discussion, before the Conclusions section;

3) TABLE 1: I suggest using “male/female” instead of “men/women”.

4) TABLE 2: Add a note with P= patients, C= controls. I suggest presenting data as count (%) instead reporting 6 different columns.

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