Reviewer's report

Title: Polymorphisms in the Epidermal Growth Factor Receptor Gene and the Risk of Primary Lung Cancer: a case-control study

Version: 1 Date: 13 July 2007

Reviewer: Hongbing B Shen

Reviewer's report:

General
This study described a hospital-based case-control study on 582 lung cancer cases and 582 controls frequency-matched to the cases on age and gender and investigated the association of polymorphisms in EGFR and primary lung cancer. The authors found that the 181946 TT genotype was associated with a significantly decreased risk of lung cancer (adjusted OR = 0.63, 95% CI = 0.45-0.88, P = 0.007) when compared with the 181946 CC + CT genotypes, especially in ever-smokers. In addition, the CGGCT haplotype with the 181946C allele was related to a significantly increased risk of lung cancer when compared to the most common haplotype CGGTT carrying the 181946T allele (adjusted OR = 1.49, 95% CI = 1.09-2.05). Overall, the study design was sound and the manuscript was well prepared. However, several concerns need to be addressed before publication.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. The authors first examined the frequencies of 39 candidate polymorphisms in the EGFR gene in 27 healthy Korean individuals. They indicated in the discussion (Page 12, Line 8-9) that another 27 cases were also detected and those frequency data could be added in Table 1. In addition, the detail information on the methods of examining the frequencies of 39 candidate polymorphisms needs to be provided in the Materials and Methods section.

2. MAF above 10% in 27 healthy Korean subjects was defined as the criteria of selecting the polymorphisms in this study. However, the polymorphism 162093G>A (rs10251977) was selected with MAF of 0.07 and the other two SNPs 142232 (rs17336800) and 151904T>A (rs17337023) with MAF above 0.10 were excluded from the subsequent analyses (Table 1). Please confirm and explain this point.

3. In discussion (Page 12, Paragraph 1), the authors stressed that the SNP 162093G>A was the true locus associated with modulated the lung cancer risk, which seemed redundant and not significant, because lack of the biological evidence. The authors could not exclude the possibility of LD with other potentially functional polymorphisms (e.g. polymorphisms located in introns or flanking region, polymorphisms of insertion/deletion, and so on).
4. In table 2, similar genotypes distributions were shown for 142285G>A, 181946C>T and 187114T>A, which indicated a possible LD, and the authors may try to recalculate the indicator of D prime and r square with another software (e.g. haploview).

5. Brief description of characteristics (age, sex, smoking status and pack-years smoking) among cases and controls was needed, though detailed information was shown in previous studies.

6. The ORs were just adjusted by pack-years of smoking in Table 2, but age was also adjusted in Table 3 and Table 4. This needs to be addressed.

7. Other haplotypes with frequencies less than 5% were excluded in Table 4, which seemed not appropriate.

8. Page 12 line 1: 181946 C-to-A should be 181946 C-to-T.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Discretionary Revisions (which the author can choose to ignore)

What next?: Accept after minor essential revisions

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.