August 1, 2008

Timothy R. Rebbeck, Ph.D.
Editor-in-Chief
Cancer Epidemiology Biomarkers and Prevention

Re: EPI-08-0523. “Association between variants of the 8q24 chromosome and nine smoking-related cancer sites”

Dear Dr Rebbeck,

Thank you for accepting our manuscript for publication in Cancer Epidemiology Biomarkers and Prevention. We greatly appreciate the time you and the reviewers have spent in reviewing this manuscript.

The manuscript has been revised in response to comments and suggestions of reviewers. We plan to submit two copies (1) clean copy of the paper and (2) a copy with all changes highlighted by yellow color.

Following are point-by-point responses to the reviewers’ comments and the corresponding changes the manuscript:

Reviewer #1

Q1. It would be helpful to provide an indication of the power that the study had to detect effects for each of the SNPs, in each of the ethnic groups, given the small to moderate sample sizes for several tumors.

A1 Many statistical authors have pointed out that after as study is completed, power calculations are misleading and all that matters is the confidence intervals. See for example page 160 of Modern Epidemiology 3rd edition (2008) for discussion and citations on this point. In essence power calculations concern only what a study yet to be conducted might find and so is vital for planning; but the intervals show what it did or did not find. Low power translates after the study into wide confidence intervals that include the null, summarized by saying the study results have low precision. Table 4 still does provide power calculations for those who still wish to see them. We simply do not want to encourage this usage.
Both low power and low precision stem from small sample size. In the Discussion session, the small sample size in the Taixing and MSKCC studies is mentioned as a limitation as follows (page 18, lines 14-18):

“The sample sizes in the Taixing study and in the stratified analyses of the LA study may affect the precision of our measurements. As a result, the interval estimates from both the Chinese study (over 200 cases for each site and over 400 controls) and the MSKCC study (172 cases/157 controls) are imprecise.

Q2. It would be useful to test if the MAFs in the different ethnic groups are significantly different from each other.

A2. According to reviewer’s suggestion, we tested the differences of MAFs in the different ethnic groups.

In the Method section, we added the following (Page 10, lines 19-20):

“… differences in minor allele frequencies (MAFs) were evaluated for all three SNPs using the chi-squared test.”

In the Results section, we added the following paragraph (page 12, lines 20-23):

“There were notable variations in the distribution of MAFs in both African-American (the LA study) and Chinese (the Taixing study) when compared to Whites (the LA study) for all SNPs (rs1447295: p<0.0001 and =0.014; rs16901979: p<0.0001 and <0.0001; rs6983267: p<0.0001 and =0.0016, respectively).”

Q3. The deviations from HWE are slight and are not of concern, as long as the QC data are rechecked to ensure that the deviations are not due to genotyping error. Discussion on this could be reduced substantially in the text.

According to reviewer’s advice, we shortened the discussion of HWE as follows (Page 17, lines 13-23):

“We observed two minor deviations in HWE (rs16901979 genotype distributions in African-Americans and rs6983267 genotype distribution in the Chinese population); however, the allelic proportions remained consistent with the previous published literature(19, 23). The chance finding, selection bias, or laboratory genotyping error may potentially lead to the HWE deviations. Since we observed a high QC concordance rate for all 3 SNPs (>99%) in our lab, the possibility of genotyping error is unlikely. Controls in both the Los Angeles and Taixing City studies were randomly selected from the population at risk using algorithms to capture an accurate representation of their respective cities (33, 35). After removing the African-American population in our analysis of rs16901979, we observed similar associations. The association between rs6983267 and liver cancer needs investigation by other larger studies.”
Q4. P-values and FPRP values should be provided, in the text and tables, to only two significant digits to the right of the decimal point.

A4. The changes have been made according to the reviewer’s suggestion. The p-values and FPRP values are now listed to two significant digits and can be found on Table 4 and in the text. These changes of tables can be found on pages 30.

Q5. The single line of data shown for the "Recessive or Dominant" models, in Table 2, is not clear, as these models test different combinations of genotypes.

A5. Instead of “Recessive or Dominant” we have now changed the presentation to genotypes in Table 2. Changes can be observed on Table 2, pages 26-28.

Q6. A p-value for the test for trend from the additive model would be helpful to show in Table 2 as well.

A6. As suggested, p-values for additive model or trend test have been added in Table 2 (pages 26-28). A sentence was added in the Methods section (page 11, lines 14-16):

“We first analyzed SNP genotypes (CC, CA, AA) as a continuous variable (additive models) and as dummy variables for each cancer site.”

Q7. A comment should be provided on whether or not tests for interactions were significant for effects in never vs. ever smokers, shown in Table 3.

A7. We added the two paragraphs on the changes in odds ratios across smoking (interaction) in the Methods and Results sections as follows:

“For each site, changes in the odds ratios for these three SNPs across levels of tobacco smoking were evaluated using unconditional logistic regression adjusting for previous mentioned confounding factors and by including smoking (never or ever), SNP genotypes (0 and 1) according to the dominant or recessive model, and product terms of each SNP.” (page 11, lines 17-22)

“For rs1447295 (region 1), when stratified by smoking, the only noteworthy change in odds ratio was found in liver cancer (p=0.025), OR\text{adj}=1.96 (95\% \text{ CI}=1.07-3.59) among smokers and 0.90 (95\% \text{ CI}=0.49-1.65) among never-smokers.” (page 14 lines 8-11)
Q8. The authors might want to consider removing data on kidney cancer from the paper, or perhaps mentioning it only in the text, given the very low power to detect effects given the small number of cases.

A8. As suggested the results of kidney cancer have been removed from Tables and a paragraph was added in the text as follows: (Page 13, lines 11-13)

“Lastly, in a pilot study, using the dominant model, we observed an inverse association of rs16901979 with kidney cancer (OR=0.48, 95% CI=0.23, 1.00, data not shown).”

Reviewer #2 (Reviewer Comments to the Author):

Q9. The authors use data from different populations to investigate different cancers. They stated that the population selection details have been described previously. The manuscript would be stronger if the case and control characteristics were tabulated separately within the manuscript or as supplementary material.

A9. As suggested, this supplementary information has now been provided and can be found on page 29, in “Supplementary Table 1”. A sentence was added at the beginning of Results section as follows (page 12, lines 10-11):

“..a short summary for study populations can be found on Supplementary Table 1.”

Q10. Ideally, the same type of cancer should be compared among different populations whose samples and data were collected using identical criteria.

A10. We completely agree with the reviewer. However, the only site we could compare among different populations is esophageal cancer. The major difficulty lies in the fact that the proportions of adenocarcinoma of esophagus in the US population are different from Chinese population. As demonstrated in our results, the associations between three SNPs and esophageal cancer were null in both Chinese and Los Angeles populations, which may in part be attributable to small sample sizes in both the Los Angeles and Chinese population. Nevertheless, we have combined both Los Angeles and Taixing studies to increase precision and observed no obvious associations between three SNPs and esophageal cancer after adjusting for potential confounding variables.

The following paragraph is added. (Page 18, lines 18-20)

“Among esophageal cancer analyses, we combined combining Los Angeles and Taixing study sites to increase precision and observed no obvious associations after adjusting for potential confounding variables.”
Q11. Authors should discuss further the importance of the 8q24 region and what the specific SNPs mean in terms of disease causation.

A11. The discussion of the importance of the 8q24 region has been expanded as follows (page 16, lines 16 – Page 17 line 12):

“The clear association of 8q24 with prostate cancer suggests a potential hormone-related carcinogenic pathway which may be associated with expression of microRNAs in the 8q24 region(48). Our results and those of previous studies shows SNPs of “region 3” are more often observed to be associated in cancer sites other than prostate, indicating this specific “region” may be involved in other carcinogenic pathways, such as a tobacco-related carcinogenic pathway, or a combination of different pathways. Recent studies have observed SNPs between 128.47 to 129.54 Mb, i.e. “region 3.” associated colorectal and ovarian cancers (24, 27-29, 49). Ghoussaini and colleagues reported this cancer associated “region” may be narrower than previously believed, spanning only 128.47 to 128.50 Mb(49). Further research will be required to determine whether 8q24 loci, specifically “region 3,” are associated with smoking-related carcinogenesis. Studies of SNPs in LD with rs6983267, SNPs within “region 3,” and those between 128.47 to 128.50 Mb with smoking-related cancers may also be useful to detect new markers and reveal possible underlying biological mechanisms. Lastly, we cannot exclude the possibility that SNPs beyond “region 3” may also be associated with tobacco-related carcinogenesis and our and our rs6982267 results due to its high MAF providing us more power to detect the observed associations. Thus, functional studies and studies with larger sample sizes should be conducted to further investigate the association of these SNPs with smoking-related cancers.”

Q12. Minor points: Abstract line 15: rs6983267 should be rs1447295. Abstract line 17: "among ever-smokers we observed associations between rs1447295 and liver cancer and rs6983267 and lung cancer" should be deleted as it is a repetition. References: several references are missing journal titles and/or page numbers.

A12 Thank you for pointing out these “minor points”. We apologize for the confusion and we wish to clarify that the points we made in abstract lines 15 and 17 were not errors. We re-organized the presentation of the results as follows (page 3, lines 15-19):

“We also observed a suggestive association between rs6983267 and liver cancer (OR_adj=1.51, 95% CI=0.99, 2.31). When we stratified our analysis by smoking status, rs6983267 was positively associated with lung cancer among ever-smokers (OR_adj=1.45, 95% CI=1.05, 2.00).”

Q13. There are several recent publications that have not been cited: Jacobs et al., Multiple loci identified in a genome-wide association study of prostate cancer, Nat Genet;40(3):310-5 (2008) and Salinas et al. Multiple Independent Genetic Variants in the 8q24 Region Are Associated with Prostate Cancer Risk, CEBP 17 (5): 1203. (2008)

A13. Thanks for suggestion. Additional references have been added (citation 26 and 27) and errors in the references have been corrected (pages 20-24).
Q14. Tables: In Table 1 variables or genotypes are listed as C/C, C/A etc. In Table 2 the genotypes are listed as wt/wt, wt/var etc. and again as CC, CA etc. in Table 3. These inconsistencies are distracting.

A14. Presentation of genotypes in Tables 1 and 2 have been changed to reflect consistent genotypes (page 25-28).

Q15. Authors could discuss if other SNPs in the 8q24 region would be used to further validate the data they present here.

A15. In response to the reviewer’s suggestion, the following has been added in the discussion section (Page 17 line 4-12):

“Studies of SNPs in LD with rs6983267, SNPs within “region 3,” and those between 128.47 to 128.50 Mb with smoking-related cancers may also be useful to detect new markers and reveal possible underlying biological mechanisms. Lastly, we cannot exclude the possibility that SNPs beyond “region 3” may also be associated with tobacco-related carcinogenesis and our rs6982267 results were due to its high MAF providing us with more precision to detect the observed associations. Thus, functional studies and studies with larger sample sizes should be conducted to further investigate the association of these SNPs with smoking-related cancers.”

We hope these changes and responses are sufficient. Again, we appreciate the reviewers’ time and efforts on improving this manuscript. We feel honored to be able to publish our results in your journal.

Yours Sincerely,

Zuo-Feng Zhang, M.D., Ph.D.
Professor of Epidemiology
UCLA School of Public Health
71-225 CHS, Box 951772
650 Charles E. Young Drive, South
Los Angeles, CA 90095-1772
U.S.A.
zfzhang@ucla.edu