Reviewer's report

Title: Carboxypeptidase 4 gene variants and early-onset intermediate-to-high risk prostate cancer

Version: 1 Date: 29 September 2008

Reviewer: Elaine Ostrander

Reviewer's report:

In the manuscript "Carboxypeptidase 4 gene variants and early-onset intermediate-to-high risk prostate cancer" by Ross et al., six CPA4 SNPs were analyzed in 1012 men (506 cases and 506 controls) from a hospital-based case-control study from Cleveland, Ohio. The authors reported that one coding variant in CPA4 confers an increase risk (OR 1.83 CI 1.02-3.36, p=0.04) of intermediate-to-high risk prostate cancer among younger patients (< 66 years). In the overall study population, no association between CPA4 variants and prostate cancer risk was found.

There are several strengths to this manuscript, including the decision to study intermediate-to-high risk prostate cancer, and the use of tagging SNPs across the CPA4 gene to assess the association between risk of prostate cancer and the CPA4 locus. A few issues remain to be addressed as outlined below.

While it is really important to discover biomarkers which can distinguish prostate cancer cases who will die of their disease versus those who die with their disease. The authors address this issue by analyzing cases with Gleason >= 7, clinical stage >= T2c, or PSA >= 10ng/mL at diagnosis. While this criteria separates out intermediate-to-high risk prostate cancer, it is important for the authors to investigate if there is any association between high-risk prostate cancer risk and CPA4 variants. Since multiple lines of evidence indicate that for Gleason 7 4+3 is different from 3+4, it might be good to separate out the Gleason 7 (3+4) from Gleason 7 (4+3), 8-10 in the analysis to make sure that the result observed is not driven only by the intermediate-risk prostate cancers.

Additional details to be filled in include:

When choosing the haplotype tagging SNPs, how many LD blocks covered the CPA4 coding region - just one or more than one? How much of the promoter and downstream region are included in the blocks/genomic area utilized? One SNP was excluded due to genotyping assay difficulties. Is there another SNP which could capture the same genetic information as that SNP? Otherwise the full haplotype information of that region is lost.
For the haplotype analysis, what frequency defined a common haplotype in the analysis? Which program did you use from the case-control analysis? Haplovew?

The authors indicate that carrying two copies of the haplotype with the rs2171492 T allele was weakly associated with prostate cancer risk. What was the comparison group? All other haplotypes or just individuals without a copy of the common rs2171492 T allele haplotype. This issue is important as the authors claim that because only one common haplotype contains the T allele (at what freq?) and that the haplotype association was weaker than the SNP alone that “the SNP itself may directly impact development of the more aggressive forms of prostate cancer.” This would be quite a strong statement if the comparison groups are not the same as in the SNP alone analysis. It would be good to see both rs2171492 TT vs TG/GG, and to get a clearer explanation of what the comparison group was for the haplotype analysis.

Finally, were there any blind duplicates in the genotyping and if so please state rate of duplication?

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 I have no competing interests.