Reviewer's report

Title: Polymorphisms in Cyclooxygenase-2 gene and breast cancer risk in Brazilians: a case-control study

Version: 1  Date: 5 March 2010

Reviewer: Thilo Dork

Reviewer's report:

The manuscript by Viana-Jorge and coworkers describes a study of PTGS2 gene variants in a series of breast cancer cases and controls from Brazil. The authors focussed on the promoter and 3´-untranslated regions of the gene and confirmed the presence of nine known polymorphisms among 67 healthy subjects. They went on to screen 402 additional healthy volunteers for the four most common polymorphisms to study ethnic differences but did not observe frequency differences among three groups stratified by self-percepted skin colour. Finally, in a subsequent case-control study of 318 breast cancer patients and 273 controls, one of the four polymorphisms (rs 5275) showed up with more heterozygotes among cases than controls, a result that the authors consider as significant after multivariate regression analysis (p=.049).

It is plausible, in principle, to investigate PTGS2 variants in a Brazilian breast cancer case-control series. COX2 has been identified as a risk factor for breast cancer metastasis, although its role for general breast cancer risk is less evident. However, the present study is relatively small and has limited power to detect moderate breast cancer risks. These limitations need to be critically discussed. The authors report a result for one SNP that is, at best, marginally significant and is not corrected for multiple testing. There are several ambiguities in this study that need to be addressed as I will explicate in more detail below.

1. In view of the title, it should be explained why only the promoter and 3´-UTR but not the coding regions have been sequenced. If the analysis of the 10 exons is part of another manuscript, this should be indicated.

2. The authors present as one significant result in the abstract and in the text that risk for breast cancer were associated with age. However, if I understand Table IV correctly, they simply show that their controls were younger than their cases. This is a sampling issue rather than a risk issue.

3. It is outlined in the Material and Methods section that the authors stratified their population by skin colour. However, they do not tell us about the distribution of these subgroups in the breast cancer case-control study.

4. It is stated that three of the four analysed common polymorphisms were in strong linkage disequilibrium (Results, Table III). Table III does not provide LD or r(2) values, but the p values seem to indicate that the alleles at -765, -1290 and
8473 are highly correlated. If this is the case, why has only the SNP at 8473 been found associated with breast cancer? Also, from Table VI it would appear that the allele 8473C is highly correlated with A-1195 (group 2)?

5. The genotypic distribution of the PTGS2*T8473C polymorphism is described as being significantly different between cases and controls. However, as is evident from Table V, this is a borderline result and mainly due to an excess of heterozygotes (or underrepresentation of both homozygous genotypes) in the cases. Accordingly, the association disappears in Table VI which basically gives an allelic OR for T8473C in group 2. I presume that there were also no association if the authors calculated odds ratios and 95% CI under dominant or recessive models, or perform a trend test. This part of data evaluation is missing here.

It is difficult to reconcile such a heterozygote excess with a plausible model or biological hypothesis. In regard that the case series shows a strong deviation from Hardy-Weinberg equilibrium, the genotypes in the cases need to be re-evaluated very carefully. As far as I understand from Table I, the genotyping for T8473C has been done with DHPLC. It would seem to be important to assess the genotypes with a more specific method and to pay attention to samples whose genotypes may not conform to the previously observed linkage disequilibrium with the other polymorphisms.

6. The numbers of individuals in Table VI are confusing. Neither do they sum up to the total given at the bottom, nor are they in line with those given in Table V. Do I understand correctly that only some half of the individuals genotyped for Table V contribute to the haplotype evaluation in Table VI?

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests