Reviewer's report

Title: Novel polymorphisms in caspase-8 are associated with breast cancer risk in the California Teachers Study: a nested case control study

Version: 2 Date: 2 September 2015

Reviewer: Jeffrey Smith

Reviewer's report:

This manuscript presents straightforward observations of association between a candidate gene, CASP8, and breast cancer risk, using a nested case control design within the CTS cohort. Rationale is provided for investigation of the candidate, one derived from prior work in the field. Association between CASP8 and breast cancer risk has been investigated by the BCAC consortium since 2007. At present collective data suggest a weak association with risk of low effect size, requiring large sample sizes for its investigation. The NHGRI’s GWAS catalog currently lists no hits at the gene. For the investigation described within this manuscript, stratification of subjects by receptor status was evaluated. This can potentially increase homogeneity and power to detect genetic association within subtypes, and is increasingly being tested. I have followed the journal’s review designations in categories below.

Major Compulsory Revisions:

For the investigation, rationale for selection of SNPs in the region is not described. Detection of association between breast cancer and CASP8 genetic variation would be highly dependent upon how systematic and comprehensive the SNP selection process may have been. The only related detail provided is that “twelve haplotype tagging SNPs” were investigated. Typical tagging strategies are based upon r2, and are less commonly selected based upon imputed haplotypes. What was the reference population and data set for their selection? What criteria and methods were used? Without the detail, the reader can have little confidence in the ability of the study to systematically detect association within the study population. No haplotype-based analyses are conducted. Also pertinent are patterns of LD between the tested tagging SNPs and those with published associations with breast cancer.

Table 2 presents the results of five separate analyses for each of 12 SNPs. Current molecular profiling of tumors into subtypes of Lum A (ER+/HER2-), Lum B (ER+/HER2+), HER2 (ER-/HER2+), and TN/basal (ER-/HER2-) has current clinical relevance. IHC rather than expression profile is instead available for many of the genotyped subjects. Corresponding analyses are expected by the reader within the manuscript. Additional IHC double-marker subgroups (ER+/HER2+, ER+/HER2-, ER-HER2+, and ER-HER2-) are of importance. It is unlikely that these were not investigated. These would alter multiple testing correction (in particular, 12 SNPs x 9 subject groupings, corresponding to a
correction of P=0.0005 to 0.054). A table that presents the number of cases and
of controls in each subgroup would also be needed. In the methods section (line
158) limited potential power is noted for one of these subtypes. The rationale of
potential increased subgroup homogeneity to improve power is the premise of
the study.

Discretionary Revisions:

Future meta-analyses would benefit from presentation of a supplementary table
like that currently included, but one for each separate comparison/subgroups that
is statistically evaluated.

The final results paragraph describing LD patterns would benefit from a typical
figure of pairwise r2 values among study controls (and correspondingly
condensed text).

Minor Essential Revision:

The investigation does not independently replicate a nominally significant
observation. False positives remain possible, even under conservative correction
for multiple testing. I recommend altering the sentence beginning on line 192 to:
“In summary, after correction for multiple testing one of the twelve CASP8 SNPs
tested in our study remained nominally significantly associated with invasive
breast cancer, specifically HER2-positive breast cancer.” The first sentence of
the discussion does not appear to consider the possibility that the observation at
rs2293554 could yet fail to independently replicate in future studies. A more
conservative presentation could use the phrase “nominally statistically
significantly.” Analogously, sentence 247 of the conclusion could also include the
word “nominally.”

Discretionary Revision:

Were any of the tested SNPs of reference 29 that are in LD with rs2293554
consistent with risk for breast cancer? At a quick glance, some may be. A given
reader is likely to wonder the same.

Minor Essential Revision:

The second sentence of the abstract an odd rationale. Cost, side effects,
efficacy. The latter would seem important, in particular here.

Discretionary Revision:

The introduction does not state whether reduced or increased caspase activity
corresponds to the risk alleles, if known.

Minor Essential Revisions:

Affected status will have been designated by the hospital in which a given subject
received the initial diagnosis of invasive cancer and thus warranted entry in the
corresponding tumor registry. Clarify whether study pathologists confirmed all
Line 197 uses the phrase “European population.” European descent was intended. Most of the text uses the phrase “non-Hispanic white.”

The paragraph at line 239 discusses potential function for the evaluated SNPs. A typical tagging strategy based upon LD would, of course, imply that each evaluated SNP is a surrogate for many others. Sometimes, the additional strategy of prioritization of a given SNP among those of an LD bin is based upon potential function. The paragraph fails to address the topic in the context of the LD bin represented by a given SNP of interest.

**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:**

no competing interest