Author's response to reviews

Title: Common variation in EMSY and risk of breast and ovarian cancer: a case-control study using HapMap tagging SNPs

Authors:

Patrick R Benusiglio (pbenusiglio@yahoo.com)
Fabienne Lesueur (lesueur@cng.fr)
Craig Luccarini (craig.luccarini@srl.cam.ac.uk)
Joan McIntosh (joan.mcintosh@srl.cam.ac.uk)
Robert N Luben (robert.luben@srl.cam.ac.uk)
Paula Smith (paula.smith@srl.cam.ac.uk)
Alison Dunning (alison.dunning@srl.cam.ac.uk)
Douglas F Easton (douglas.easton@srl.cam.ac.uk)
Bruce AJ Ponder (baip@mole.bio.cam.ac.uk)
Paul D Pharoah (paul.pharoah@srl.cam.ac.uk)

Version: 2 Date: 15 May 2005

Author's response to reviews: see over
Common variation in *EMSY* and risk of breast and ovarian cancer: a case-control study using HapMap tagging SNPs. Benusiglio et al., submission to BMC Cancer.

Here is how we have addressed the comments made by the three referees.

Referee 1: NELSON TANG

**Minor revisions**

1. More details have been given on htSNP selection page 7.

The paragraph was edited as follows: We selected htSNPs from the HapMap database (www.hapmap.org, public releases up to September 2004) with the TagSNPs program, including 5 kilobases upstream and downstream the gene, aiming for a minimum $r^2_h$ of 0.8. $r^2_h$ is a measure of correlation between haplotypes defined by all SNPs and haplotypes defined by the selected htSNPs.

2. Our primary purpose was to establish if htSNPs in *EMSY* were associated with cancer risk. This is what we have described in the introduction, methods and results sections. However, the leave-one-out procedure was used as a means to assess the validity of our study (how well we have covered genetic variation, how reliable our negative results are). We therefore strongly feel that its place is in the discussion.

3. We have now introduced all SNPs with conventional rs numbers in the abstract, results (page 9), and table 1, before giving the less conventional names (5’up t>g, for example) used in the remainder of the article.

The paragraph was edited as follows: In order to screen the gene promoter and coding regions for polymorphisms, we performed denaturing high performance liquid chromatography (DHPLC) using the Wavemaker detection system (version 4.1, Transgenomics, Crewe, United Kingdom), followed by sequencing (3100 Genetic Analyser, Applied Biosystems, Warrington, United Kingdom) on genomic DNA from 48 random controls.

5. Table 4 has been updated, haplotype frequencies are now given for cases and controls

Referee 2: THOMAS SELLERS

Minor revisions

1. See reply to Nelson Tang, minor revisions, point 4.

2. There is little to be gained by adjusting for other co-variates in the context of a genetic association study. Such adjustment should only be carried out for true epidemiological confounders, and it is difficult to envisage any such factor which might confound a genetic association. A confounder is a factor that is associated with both exposure (genotype) and outcome of interest (breast or ovarian cancer), but is not a factor on the causal pathway (an intermediate factor). Any of the known risk factors for breast cancer, if associated with genotype of interest are more likely to be on the causal pathway. Adjusting for an intermediate factor biases results toward the null. Family history would certainly be an intermediate factor and not
a confounder and so adjusting for it would be inappropriate. Age is a possible confounder - for example if the genotype influences survival through a disease process different to the one of interest. Under these circumstances there would be an association of genotype with age in controls. We do not observe such an association in our controls (if we did it would be of interest in itself). We have added the sentence below to the results section (page 9) to make this clear.

"There was no association of genotype with age in controls and, as expected, age adjusted risks were close to the unadjusted risks (data not shown)."

3. None of the 22 SNPs available in HapMap are in coding regions.

We added the following to the first sentence of the Results section, page 9:

Genotypes for 22 common EMSY SNPs were available in HapMap, none of the SNPs were in coding regions.

4. The sample size required to detect a common disease susceptibility allele is typically reduced by two- to fourfold when familial cases are selected. However, since only about 15% of our cases are familial, statistical power would not be increased by restricting our analyses to such cases.

Discretionary revisions

1. We did carry out the analysis in breast cancer cases; no association was seen between genotypes and age of onset. These negative results were not included in the manuscript for the following reasons:

- The biological hypothesis behind the analysis is debatable: it is true that relatives of young breast cancer (or ovarian cancer) patients are at a particularly increased risk of developing the disease (the genetic influences are strong). Nevertheless, the contribution of genetic factors remains evident even
in later-onset breast cancer cases, as relatives of such cases are also at increased risk of developing the disease. If EMSY genotypes influenced disease risk, we would have observed a main effect in our all cases-all controls comparisons.

- This was a post hoc analysis and as the number of post hoc analyses is almost unlimited, the potential for false positives is huge.

2. The proportion of non-Caucasians in our study is so small (around 2%) that removing them would not influence the results, should the genotype-phenotype association vary in different ethnic groups (which is itself an unlikely hypothesis, see Ioannidis JP, Ntzani EE, Trikalinos TA: 'Racial' differences in genetic effects for complex diseases. Nat Genet 2004, 36: 1312-1318).

3. See reply to minor revisions, point 2.

Referee 3: JENNIFER HU

Major revisions

1. 

a. We have added the age characteristics for cases to the Patients and controls section, page 6:

Median age at diagnosis was 51 years for breast cancer cases (age range 25 to 69) and 55 for ovarian cancer cases (age range 16 to 74).

For controls, we added: The EPIC-Norfolk cohort comprises 25,000 individuals resident in Norfolk (East Anglia), ages 45-74 years.
b. For age of onset and genotypes, see reply to Thomas Sellers, discretionary revisions, point 1. For genotype-disease associations by family history, see reply to Thomas Sellers, minor revisions, point 4. We did not perform more sub-group analyses, as we believe that it is inappropriate to carry out detailed sub-group analyses (stage for example) particularly in the context of a null result for the main effect as we have for our data. The problem of the Type I statistical error in association studies is well recognised, as the prior probability of association is very small, and uncritical, post hoc, sub-group analyses will only make this worse. The difficulties of interpreting sub-group analyses in clinical trials is widely recognised (see Schulz KF, Grimes DA. Multiplicity in randomised trials II: sub-group and interim analyses. Lancet 2005;365(9471):1657-61.) and we do not believe that we should over-analyse our data either.

c. Regarding adjustment, see reply to Thomas Sellers, minor revisions, point 2.

2. Although the 4 putative coding SNPs mentioned by Dr Hu are included in the dbSNPs database, none of them has actually been confirmed (there is no frequency data). We have screened the gene coding sequence with DHPLC, a highly sensitive technique and have not identified any of these putative coding SNPs. We are therefore confident that these 4 coding SNPs are not present in our white British population. To avoid confusion, we have removed the sentence “No coding SNPs have been identified in EMSY so far” from the introduction.

**Minor revisions**

1. Dr Hu is raising an important issue. However, the number of cases carrying BRCA2 mutations in our study is small (n = 16). Power to detect BRCA2-EMSY genotype interactions would therefore be very low. We believe that breast cancer families with mutations in BRCA2 would be more appropriate for such analyses.
We hope you will be satisfied with our revisions. We uncovered, after submission, an error in the command file used to carry out the leave-one-out procedure (discussion section, 3rd paragraph). We had in fact strongly underestimated the ability of htSNPs to tag unidentified SNPs. Appropriate changes have been made to $r^2$ values.