Reviewer’s report

Title: Disease-associated alleles in genome-wide association studies are enriched for derived low frequency alleles relative to HapMap and neutral expectations

Version: 3 Date: 27 October 2010

Reviewer: Nic Timpson

Reviewer’s report:

Review 3
Joseph Lachance
"Disease-associated alleles in genome-wide association studies are enriched for derived low frequency alleles relative to HapMap and neutral expectations, BMC Medical Genomics."

Comments to the Author:

Since the last submission of this paper, there has been considerable improvement. Most of the points raised have been addressed, although there are still a series of concerns which need to be addressed. These are outlined below:

General comments re. response to review:

(i) In your response to point 1, you give reasonable justification as to not pursuing a look up in the 1000G data at this stage. Given this the paper needs to be explicit about the part of the genetic spectrum (i.e. type of variation - common here) that is examined by this paper. This is a theme that has to be consistent throughout. Re. the point that rarer variants will need to have larger effects to allow detection - this is of course true, but they tend to (and are only evolutionarily viable if rare).

(ii) In your response to point 6, you mention that the work does not take into account differences between study populations, but that this should not be a major problem as the frequency of associated SNPs has been shown to be relatively consistent. This is fine, but should now be acknowledged in the paper.

(iii) Other points re. the last review are well addressed.

Comments from the new version of the paper:

(i) In the opening paragraph of the abstract (background), there is reference to GWAS "becoming increasingly common". I think that it is fair to say that they are now established.

(ii) In the methods (abstract), reference is given to the "disease-associated alleles" - it should be made explicit early on what type of variants these are - i.e. common. Other variants of differing frequency may not adhere to the patterns
seen here.

(iii) In the results section of the abstract (and in the conclusion of the paper), reference is given to the disease associated variants being "low frequency derived alleles...". It is understood that this is relative to neutral expectations, however this should be worded as elsewhere in the paper - i.e. these common variants were still common, but just lower in frequency than would be expected against neutral expectations.

(iv) In the conclusion of the abstract, the opening line is too general - these variants may well be different to others in the same sampling frame in the genome (i.e. those mined by GWAS), but their difference from the "rest of the genome" has not been tested for here.

(v) Please replace "genomic" with "genomewide" in the opening section of the background.

(vi) This issue of saying something about the frequencies of the disease associated alleles should be kept in the variant sampling frame throughout the paper - i.e. in the background, please reiterate this in the phrase "... there are now enough genome-wide association studies to say something about the frequencies and ancestral or derived state of disease associated alleles" - i.e. in comparison to what?

(vii) Please refrain from commenting on results in the final paragraph of the introduction.

(viii) The sensitivity analysis mentioned in the null expectations section of the methods is great, but please only comment on its methods here and retain results for the results section.

(ix) Try to remove subjective assertions such as "interestingly" or "conveniently".

(x) In the empirical data section, it is mentioned that 92 of 1143 GWAS alleles were implicated in disease associations not in Europeans. It is not clear, however, what was done with these?

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

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.