Reviewer's report

Title: Genetic variations on chromosome 15q14-22.1 for type 2 diabetes candidate region in the Japanese population

Version: 4 Date: 15 December 2007

Reviewer: Steven Elbein

Reviewer's report:

General Comments:
This is the revised manuscript reporting dense SNP mapping on chromosome 15q14-15q22.1. The manuscript is revised, but many of the issues raised in the initial review are not well addressed. The authors now state that the methods used in this study resulted from starting the study apparently before HapMap data were available. They use a two stage design in a reasonable sample of 1794 individuals, but they now concur that power for expected effect size was in fact fairly low. This alone does not diminish the importance of a report with a dense map, but combined with essentially negative data, a SNP map that appears not to capture the full variation in the region (particularly intronic and intergenic variation), and a long, redundant, and not well written manuscript, this Reviewer's enthusiasm for the work is limited. If the authors believe that this region is important, surely they should place additional SNPs in this region based on Japanese and Chinese HapMap samples and perform the study correctly. The Revision has added considerable text, but remains highly redundant.

Major Compulsory Revisions:

1. As discussed above, this manuscript is 20 pages with multiple additional files, many figures and tables, all for essentially a negative study. The authors have provided no reason that UBR1 should be considered a diabetes candidate gene indeed, they spend no effort even telling the reader what it does. SNP2140 is only weakly associated, and in the two stage design barely meets 0.05 cutoffs at either stage. Considering 1317 SNPs in stage 1 (ie, over 200 expected to be significant by chance depending on LD), and 112 in stage 2 (20 expected by chance), the findings in this paper would properly be interpreted as negative. Using the two samples together, the p value is a bit more impressive at 0.004 and OR of 1.26, but with so little power to detect genes with this effect size, the interpretation is impossible. Apparently no other SNP in UBR1 was associated.

2. This manuscript is at least 50% too long. It should be shortened considerably given the negative study. Analyses such as EHH and REHH seem to add little.

3. The authors still confused this reader on where the SNPs for this study came from. The proper comparison is not JSNP, which has very limited utility, but HapMap, which is more comprehensive and provides the best data today on what total genetic variation is present. All the readers need to know is how much of the variation in HapMap was captured using the current SNP map of 1300
SNPs as proxies and at different values of r\(^2\) (0.7, 0.8, 0.9 for example). If most intergenic SNPs were not captured, the authors may well have missed the important variants and latched onto a marginally significant one instead.

4. The entire discussion of LD block structure seems superfluous. These data can be obtained from HapMap, but pages are spent providing methods and data. This reviewer recommends that this be removed.

5. The EHH and REHH sections remain impossible to understand. As this seems to add little, this Reviewer recommends that these be removed.

6. Please remove the paragraph on page 11 comparing the SNPs to JSNP. The proper comparison is to HapMap, which is not gene centered or biased. HapMap will incorporate JSNP. The new sections add nothing to the manuscript, except for information about very large gaps of 43 Mb with no SNPs.

7. Delete the paragraph on LD block structure on page 13; this adds very little useful data. Using a gene centered SNP map to conclude that the results provided "considerable coverage" is simply flawed logic. The correct standard is HapMap, which is not gene centered. Gene centered maps can be expected to show stronger block structure.

8. Please provide the justification for the studies on Page 13, bottom, and page 14. If UBR1 is not a strong candidate, and the association for SNP2140 is marginal, why choose additional SNPs? Why choose only these 7? Again, too much is made of block structure and LDU; the association results are never provided, which is the point of this paper. How was UBR1 resequenced "did you sequence only exons? This strategy remains very unclear.

9. This reviewer does not understand where 7 tag SNPs came from. Only 8 SNPs were typed! If LD was strong, those 8 SNPs could surely be covered with fewer than 7 tags. This section is simply impossible to understand as written. No reader is realistically going to refer to 7 additional files with a long paper. Please restrict the data to what is important for readers, and put those data in the paper. The investigators need to seriously edit this manuscript.

10. Remove the paragraph "Comparison of the landmark LD block...". This information has no bearing on the role of UBR1 in diabetes. Similarly the next paragraph in this Reviewer's opinion adds nothing.

11. Page 19, the HapMap data should have at least been used in the region of UBR1; not doing so leaves the paper with limited interest.

12. The new paragraph comparing LD between Japanese and YRI is of no real interest. Sufficient publications are available showing marked differences among populations, particularly YRI vs CEPH or Asian samples. This really contributes little but length and is recommended to be removed.

13. Table 2 can probably be removed; it provides little usable data. What matters is how well the total genetic variation was captured in this study; JSNP is not the standard, and simply counting SNPs is not helpful. Similarly Figure 1 adds very little. A more useful "B" figure would be SNP density by distance, showing gaps.
14. Did the authors look at Stage 1 + Stage 2 for the stronger SNPs in Stage 1? Did they only look at the arbitrary 0.05 cutoffs for both stages separately? Given the poor power of Stage 1, this might be an error. Please present the combined data for all SNPs with stage 1 p<0.01.

15. 8 additional files is excessive. This reviewer and all but a few readers will not have time to open and examine each file. The vast majority of these are not essential. The linkage region should be shown in the manuscript, and the manuscript should show a table of the top 20 associations from Stage 1 using a well accepted test (Armitage Trend is a good choice). They can then provide Additional File 2, but clearly only 112 actually had second stage and combined data. The current submission is overly confusing and impossible to read.

Minor Essential Revisions:

1. Change the title to Genetic variation (not plural).

2. Please add sufficient description of the two stage design (number of subjects in each stage) in the abstract.

3. The grammar throughout is problematic. For example, Background have replicated confirmed 11 genes; makes no sense; do the authors mean have confirmed 11 genes with replicated associations with T2DM in Caucasians? Similar problems are present throughout the manuscript, and require careful editing of the English. On page 5, last sentence, change was closely examined to were closely examined. On Page 8, please add 372 cases and 360 controls if this is appropriate, and change false positive to false positive rate.

4. Please provide a reference for SNPAlalyze (page 9).

5. Clarify everywhere which p values are for stage 1, stage 2, and combined. Giving a p value of 0.0043 (page 17) without explaining that this is combined is confusing.

6. No real need for Figure 3D; the text is sufficient to understand the power limitations.

7. Page 18, paragraph The density of SNP markers is unclear; this needs to be reworded, and this entire section is highly redundant.

8. Page 19, tag-SNPs were not available.

9. Page 29, the figure legend is confusing. The manuscript states 1794 subjects, but here the authors list 372 stage 1, 532 stage 2, and 904 stage 1+2. Please clarify. Was power based only on cases?

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article of limited interest

Quality of written English: Not suitable for publication unless extensively edited
Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:

I declare that I have no competing interests.