Reviewer’s report

Title: Genetic variations in APPL2 are associated with obesity in a Chinese population with normal glucose tolerance

Version: 1 Date: 29 November 2011

Reviewer: Francis Vasseur

Reviewer’s report:

The manuscript presented by Shan Jiang et al investigates associations between tag SNPs of the APPL2 gene and obesity related parameters. The work is based on a large population. However some questions and remarks must be drawn.

#1 At least on the pdf file provided for reviewing the manuscript, except on the cover page (with title, short title, authors, affiliations...) where lines are numbered, the lines are no longer numbered on the following pages and the remaining pages of the manuscript are not numbered.

#2 In the background section the authors wrote "APPL1 is correlated with body fat distribution in Chinese type 2 diabetic patients". As a gene per se may not be correlated with anything, do they mean that genetic variants of the APPL1 gene are associated with body fat distribution?

#3 In the beginning of the results section the authors wrote that their 5 SNPs were in modest linkage disequilibrium. However according to the HapMap version 2 Release 22 or 24, at least rs2272495 and rs1107756 display a 0.93 D’ in the CHB+JPT population. How the authors explain the strong LD reported in the HapMap project and the weak one their report for their data.

#4 As according to the HapMap data the two SNPs associated with obesity and obesity related phenotypes, disclose three major haplotypes, why the authors did not complete their associations studies with haplotype analyses?

#5 In table 1, the authors display the OR and their 95%IC together with pvalues and empirical pvalues determined following Monte Carlo permutations. Indeed it has become classical not to present the pvalue and the 95%IC when results are expressed as OR: indeed when the value 1 is excluded from the 95%IC the pvalue is always significant at the 5% risk level. Thus it is often considered as redundant to display both 95%IC and pvalue and it is now classical to only display the 95%IC. I propose to delete the pvalues from the table(s) and the manuscript. However empirical pvalues should be presented as they represent a robust estimation of the significance.

#6 A table presenting the number of patients, the clinical and anthropometric features for the 3 groups of subjects should be of great interest. This table should display (1) either mean and SD or median and Q1 Q3 according to the parameter distribution, (2) frequencies of relevant parameters for categorical variables. In the method section the authors refer to their reference [9] for the details of the population of 1808 patients included in the present study; are the 1808 patients
the one recruited as control subjects in the stage 1 of Diabetologia. 2010;53(2):290-298 Association of genetic variants of NOS1AP with type 2 diabetes in a Chinese population Hu C et al ? This Diabetologia paper refers to another Diabetologia paper (2009, 52, 451-456) regarding the control population that was only 1800 patients! In the latter Diabetologia paper the control population refers to a third Diabetologia 2007 (50, 286-292) where the control population appears at last well described with 2666 non-diabetic participants. What were the criteria to pick up 1800 (1808?) subjects among the pool of 2666 participants in the initial paper? 

#7 In the result section and in the table 1 the overweight patients have been pooled with the obese patients. What are the results of association with rs2272495 and rs1107756 when separating overweight and obese patients? 

#8 In the results section and in table 2 the authors present the results of association between rs2272495 and and obesity related phenotypes. It is not explained in the legend of table 2 what group of patients (obese? overweighted? normal weighted?) have been included in the GLM analysis. This should be clearly mentioned both in the results section and in the legend of table 2. 

#9 Regarding this table 2, when summing the patients included in the GLM analysis: 131+706+915=1752, one can imagine that the whole patients of the study have been included (obese, overweighted, normal weighted) as in the method section the authors wrote they included 1808 non diabetic patients. Are the associations with obesity related phenotypes in table 2 the authors report on the whole cohort of patients significant in separate groups (obese, overweighted, normal weighted) taken individually? This maybe of great importance as the association between rs2272495 and and obesity related phenotypes could concern only one of the subgroups of patients. 

#10 Regarding table 2 the authors report mean and SE. I suppose they mean standard error (of the mean?). This abbreviation should have been mentioned in the legend of the table 2. Moreover as for a normal distribution mean and standard deviation and for a asymmetrical distribution median, Q1 and Q3, best reflect the dispersion of the distribution the relevant parameters should be added in the table rather than the SEM, according to the distribution of the parameter under investigation. 

#11 In table 2 the authors report the results of their comparisons between 3 groups of genotypes. The significant result they present for every obesity related phenotype only reject the Ho hypothesis "there are no differences between the mean values of the 3 genotypes". It would be of great interest to have the contrasts results between the genotypes; i.e. is there a significant difference between TT and CT and between CT and CC .... In this regard using the non parametric Steel-Dwass test would be appropriate; alternatively the Tukey-Kramer or the Dunnet test if used in the context of the Conover and Iman method (see #14) may be suitable too. A crude analysis of the authors data using SD=SEM*#n and (mean-i - mean-j)/(#(SDi2/nI)+(SDj2/nJ)) to estimate the significances let think that association between BMI and genotypes at rs2272495 should be under a recessive model, likewise for waist circumference and hip circumference. Under the same calculation the association with WHR was not
significant but it was only a crude approach. Thus it is fundamental that the authors be more precise on these features. Similar remarks must be drawn regarding the table with rs1107756 results in the supplemental file.

#12 A comment about the results displayed in table 2: between the CC and the TT genotypes at rs2272495 there is a mean difference of 0.6 kg/m² for BMI, 1.7 cm for waist circumference, 0.9 cm for hip circumference and 0.01 units for WHR. Except the fundamental knowledge of associations, what are the clinical and biological relevances of such small differences. This would gain to be discussed in the discussion section. These questions appear still more important as association results are probably under a recessive model, thus pooling CT and CC for the rs2272495 and this implies pooling a population whose mean BMI is 23.549 with one of mean BMI 23.231 and comparing with the CC group whose BMI is 22.876. Thus the differences would certainly be smaller. Similar remarks may be drawn for other phenotypes under investigation and for associations with rs1107756 in the supplemental file, where differences are 0.4 kg/m², 1.5 cm, 0.96 cm, respectively.

#13 The population of 1808 subjects included in the study was divided into three groups according to the 25 and 30 kg/m² BMI thresholds as stated in the method section. Thus although one can regret that the authors did not describe their population (see #6) it is likely that there were some patients with a BMI over 30 and some with a BMI under 25 kg/m². Indeed regarding the results of table 2, the mean BMIs for each genotype is around 23 kg/m² (22.876 for TT, 23.231 for CT and 23.549 for CC). In this context for a population with a BMI range from lower to 25 until over 30 kg/m² having a mean BMI value of 23 kg/m² implies a very strong asymmetrical distribution. Indeed the authors report in the method section that quantitative traits were skewly distributed. Thus it is clear that mean and standard deviations are not suitable to offer a reliable description of the population.

#14 Regarding the multivariate analyses (GLM method), unless using huge number of statistical units, such parametric analyse is well known to be sensitive to non normal distributions of the quantitative variables. In the present manuscript the minor group (TT genotypes) only had 131 patients. Although the authors performed a log transformation of the traits under investigation, according to the very asymmetrical distribution of the parameters (see #12) associated with a relatively low number (n=131) of statistical units in the smallest group it would be more convenient to use non parametric approaches as the one described by Conover WJ and Iman RL, to get confident into the statistical results. (Rank transformation as a bridge between parametric and nonparametric statistics. Am Stat 1981;35:124-133).

#15 At the end of the results section, referring to data of their table 2, the authors wrote that "similar association was found for rs1107756". Indeed as they report 4 association results in this table they should have wrote "similar associations were found for rs1107756"? Anyway this affirmation requires a small flat as for WHR it did not reach significance. In the supplemental file it is perhaps not necessary to report the 0.002 beta value for WHR as the results conclude that for this WHR trait, beta is not significantly different from zero.
#16 As the APPLs are involved in the adiponectin and insulin pathways did the authors had the opportunity to analyze insulin sensitivity indexes in relation to the SNPs they investigated in the present study?

#17 Although it substitutes a Valine by an Alanine (both hydrophobic amino acids) the authors have a putative explanation regarding a possible functional aspect of the rs2272495. What about such an explanation of a putative functional variant for rs1107756?

#18 In the discussion section the authors wrote that "those carrying more C alleles of rs2272495 showed higher values of BMI, waist circumference....". This affirmation would be acceptable only if they had clearly demonstrated that the reported associations rely on a co-dominant model. However this was absolutely not demonstrated in the manuscript and crude analysis let think that the reported association would rely on a recessive model. So unless clearly demonstrated such an affirmation must be regarded as an over-interpretation of the data.

#19 In several instances in the manuscript (in the title, in the abstract, and in the discussion section) the authors claimed that there was association between SNPs of the APPL2 gene and obesity. This affirmation may not be accepted as they only report significant association using the 25 kg/m2 BMI threshold as they pool overweighted and obese patients in their analyses. They only could talk of an association between variants of the APPL2 gene and a BMI over 25 kg/m2 that is very far from an obese condition as set by the WHO criteria.

#20 In one of the last pages of the manuscript a typo mistake is noticed: "Additional materia".

Level of interest: An article whose findings are important to those with closely related research interests

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