Author's response to reviews

Title: Analysis of common PTPN1 gene variants in type 2 diabetes, obesity and associated phenotypes in the French population

Authors:

Claire Cheyssac (claire.cheyssac@good.ibl.fr)
Cecile Lecoeur (cecile.lecoeur@good.ibl.fr)
Aurelie Dechaume (aurelie.dechaume@good.ibl.fr)
Amina Bibi (a.bibi@imperial.ac.uk)
Guillaume Charpentier (guillaume.charpentier@ch-sud-francilien.fr)
Beverley Balkau (balkau@vif.inserm.fr)
Michel Marre (michel.marre@bch.ap-hop-paris.fr)
Philippe Froguel (philippe.froguel@good.ibl.fr)
Fernando Gibson (fernando.gibson@imperial.ac.uk)
Martine Vaxillaire (martine.vaxillaire@good.ibl.fr)

Version: 2 Date: 28 March 2006

Author's response to reviews: see over
Dear Editor,

We have revised our manuscript according to the comments raised by the reviewer, each point has been addressed in details in the response letter. We hope the amendments made to our revision means that you find that the paper is now suitable for publication.

Best regards,

Martine Vaxillaire
Response to the reviewer:

Major compulsory revisions

1. LD structure of the PTPN1 gene.
The authors state that the LD structure is consistent with previous reports that the PTPN1 gene encompasses a single haplotype LD block. It is therefore puzzling that figure 1 presents a haploview block structure that clearly has two haplotype blocks-- this needs to be explained.

We agree with the reviewer that the LD structure of the PTPN1 locus, defined by the four gamete rule, showed two haplotype blocks when we analysed the 14 SNPs genotyped in a group of 736 normoglycaemic non-obese French subjects. From our data (Figure 1), we can see that most of the SNPs in the PTPN1 region were strongly correlated, confirming previously published observations. Moreover, we observed a strong linkage disequilibrium between the two blocks (D’= 0.98). When considering haplotypes with a frequency ≥ 2%, which is rare in our sample, only one block was observed illustrating that the two block structure is due to a rare haplotype (frequency = 1.5%).

We have added more detail in the result part on page 9:
"The linkage disequilibrium (LD) structure at the PTPN1 locus was evaluated from the fourteen SNPs genotyped in a group of 736 French normoglycaemic non-obese subjects (C2). As shown in Figure 1, the LD analysis defined two strongly correlated haplotype blocks (D’ = 0.98). Moreover the LD matrix showed high pairwise LD between 12 of the 14 SNPs genotyped. The SNP rs3787335 presents a moderate LD (0.55 ≤ D’ ≤ 0.70) with 7 SNPs leading to a two block structure covering ~91 kb at the PTPN1 locus and SNP rs914458, located 10 kb downstream of PTPN1, showed no LD with the other 13 SNPs genotyped. Thus, our data indicates that PTPN1 lies in a region of strong LD, in agreement with previous reports [14, 16]."

2. The abstract is not a very good reflection of the data reported in the body of the manuscript. For example it states that none of the 14 SNPs show evidence of association in the case-control study, however rs6020563 does have a P-value of 0.04 D1/C1. It is also puzzling that the results for rs914458 are presented for the dominant model with T2DM in the abstract, but these data are not shown in the tables. In addition, in the abstract the summary of associations with HOMA-B, HDL, etc. are reported as two P-values. It should be clarified that the P-values are under dominant and recessive models.

To better reflect our data we have modified the abstract as following:
"From the 14 SNPs investigated, only SNP rs914458, located 10 kb downstream of the PTPN1 gene significantly associates with T2D (p = 0.02 under a dominant model; OR = 1.43 [1.06-1.94]) in the combined sample set. SNP rs914458 also showed association with moderate obesity (allelic p = 0.04; OR = 1.2 [1.01-1.43]). When testing for association with metabolic traits, two strongly correlated SNPs, rs941798 and rs2426159, present multiple consistent associations. SNP rs2426159 exhibited evidence of association under a dominant model with glucose homeostasis related traits (p =0.04 for both fasting insulin and HOMA-B) and with lipid related variables (0.02 ≤ p ≤ 0.04). Moreover, risk allele homozygotes for this SNP have an increased systolic blood pressure (p = 0.03). No preferential transmission of alleles was observed for the SNPs tested in the family sample sets."
As suggested by the reviewer, we have also added the results of the test under dominant and recessive models for the combined analysis in Table 2.

3. Haplotypes are reported numerically (1221). This presents a major challenge to even a motivated reader to ascertain which base is actually making up the haplotype. This makes it especially difficult to assess the haplotypes from different studies. The base designations should be used.

As proposed by the reviewer, we have now included the base designation in Tables 2 and 3.

4. It is also puzzling that the next states that no haplotypes show evidence of association with diabetes or obesity but in the tables the 2112 haplotype does show evidence of association with both T2D and obesity.

Indeed, the CACG haplotype (previously referred to as 2112) showed significant nominal p-values for T2D and moderate obesity association, but the overall p-value was not statistically significant. We tested the relevance of these associations using 1000 permutations with the Cocaphase software. For both T2D and moderate obesity, the empirical p-value of the original association for the CACG haplotype was not statistically significant suggesting that such “associations” can be observed by chance (empirical p-values = 0.22 and 0.30 for T2D and obesity respectively).

Accordingly, we have added comments in the result section on pages 11:
"In the combined sample, haplotype CACG, including the G "protective" allele of the previously associated SNP rs914458 was more frequent in the control subjects than in the diabetes cases (Table 2) suggesting a potential "protective" effect. In order to better estimate the significance of this effect, we computed the p-value for the association in 1000 permutations of the dataset. The empirical p-value obtained (p = 0.22) was not statistically significant, suggesting that the association could have been observed by chance."

and on page 12: "The effect of the haplotypes was also investigated in both groups of obese subjects. In spite of a non-significant overall p-value, the CACG haplotype was significantly more frequent in the control group compared to the moderately obese subjects group (Table 3). However, the empirical p-value estimated through permutation testing was not significant (p = 0.30) suggesting that the association observed with moderate obesity was probably due to chance."

5. The follow-up assessment of the polymorphisms in the French Canadian families that originally gave evidence of linkage on 20q is appropriate, but as presented is somewhat misleading. In this study the total sample did not provide evidence for linkage in the PTPN1 gene region and this should be noted. A subset of the families (I believe 55 early onset sib pairs) appeared to drive the evidence of linkage in Zouali et al. Have these SNP adjustment analyses been performed in this subset of the family collection?

Indeed the association with T2D was also investigated in the group of 55 families with early onset diabetes (age-of-onset before 45 years). No preferential transmission to affected offspring was observed in this subset and this is now added in the Results section on page 13:
"Three non redundant SNPs (-7077 G/C, rs941798 and rs914458) presenting trends or associations with T2D in the case-control study, or with quantitative traits, were further investigated in the entire family sample set comprising 148 French families and in a subset
of 55 sib-pairs presenting with an early age-of-onset of T2D and having shown linkage at the 20q13 locus. No preferential transmission of alleles was observed for the SNPs tested in either of the samples (using the FBAT software; data not shown).”

Minor Essential Revisions

1. There are significant number of typological errors in the text

2. There are quite a few places where results for T2D analysis with the dominant or recessive models is quoted in the text, but these data are not shown in tables. The text should state “data not shown” or the authors should include this data in tables.

We prefer to show the full data and we have added in Table 2 the results of the test for dominant and recessive models in the combined analysis.