Author's response to reviews

Title: Association of common variants in JAK2 gene with reduced risk of Metabolic Syndrome and related disorders

Authors:

Alberto Penas-Steinhardt (pufetin@gmail.com)
Mariana L Tellechea (marianatellechea78@hotmail.com)
Leonardo Gomez-Rosso Dr (mulag@hotmail.com)
Fernando Brites Dr. (fbrites@hotmail.com)
Gustavo D Frechtel Dr (gfrechtel@ffyb.uba.ar)
Edgardo Poskus Dr (eposkus@ffyb.uba.ar)

Version: 3 Date: 2 November 2011

Author's response to reviews: see over
Cover Letter

Dear Tim Sands

Executive Editor, BMC Medical Genetics

Please find the manuscript entitled "Association of common variants in JAK2 gene with reduced risk of Metabolic Syndrome and related disorders" (MS: 7788175805587496) attached for your reconsideration to be published in BMC Medical Genetics.

We really appreciate that the article has been revised by external reviewers and that they have found it of potential interest, and are really grateful for all criticisms. We have modified the article taking into account suggestions of the editorial staff and the referees. We provide a point-by-point response, indicating all the changes made in the manuscript or a rebuttal of each point, which you will find appended below.

We really hope the changes made will meet the expectations of the editorial staff and the referees and that you will reconsider the publication of the revised paper after reevaluation.

Sincerely yours,

Dr Penas-Steinhardt

Corresponding author
Response to Reviewers

Response to Reviewer #1 María Eugenia Sáez

Reviewer's report: 1. Please, specify how the power has been estimated. Are there 1 or 2 sided values?. Are you based this analysis in allelic effects or dominant/recessive models?. The reference does not apply.

Response. For each individual case-control study power estimations were performed for single-point allelic effects as described by Skol AD et al.\textsuperscript{1} We performed power estimations to detect effects under both dominant and recessive models with an odds ratio of 1.5 at a nominal significance level of 0.05, two sided. Power estimation to detect effects on HTG, HW, obesity or abdominal obesity was found to be between 81 and 99\% assuming a dominant model and less than 80\% assuming a recessive model for rs7849191; and between 80 and 98\% assuming a recessive model and less than 80\% assuming a dominant model for rs3780378. The power estimation was found to be between 85 and 99\% to detect effects of rs7849191 and rs3780378 on decreased HDL-C. Unfortunatelly, power estimation was found to be less than 80\% to detect effects on Metabolic Syndrome.

Reviewer's report: 2. Although the degree of genetic differentiation within European populations is small, it exits (Gayán et al. BMC Genomics 2010, 11:326). It is possible to check in allele
frequency differ between those with Spanish and Italian origin and in this case, the origin could be included as a covariable in the study. It's not mandatory but interesting.

Response. We are not able to check if allele frequency differ between those individuals with Spanish and Italian origin because most participants have both Italian and Spanish ancestors. Furthermore, we are not able to include etnic origin as a covariable because we have no detailed information about this, mainly because this study was not intended to ask or adjust for genetic admixture. Individuals who have not European ancestry, from Spain and Italy were excluded from the study. Finally, we should add that most molecular association studies did not be include ancestry origin as covariable.

Reviewer's report: 5. That's good for the qualitative analysis, but for the quantitaive analysis, specific medications must be taking into account since they affect the means. At least, the use or not of specific medications must be entererd in the linear model as a covariable. Vieweng the means, perhaps an additive model will fit.

Response. We stated that lipid lowering medication has been taken into account to ask for the presence of MS. Each subject was assessed for the presence of MS using the American Heart Association/National Heart, Lung, and Blood Institute (AHA/NHLBI) 2005 criteria.² This statement allow for triglycerides, HDL-C levels, and blood pressure to be counted as abnormal when a person is taking drug treatment for these factors. We agree that for the quantitaive
analysis, medication must be taking into account since they affect the means. Data on
demographic characteristics, lifestyle risk factors, and specific medication were obtained using a
standard questionnaire administered by trained staff. In the sample under study only few
participants were taking lipid lowering medication. We entered medication in the linear model as a
covariable and results did not change substantially.
Response to Editorial Staff

Editorial staff comment 1. The Welch tests is not well standard in the community of association study scientists. Justification of its use and a reference should be given.

Response. For comparison of continuous variables we conducted one-way ANOVA with Levene's Test for equality of variances. In the first submission of the present manuscript, if the data failed to meet the Levene's test criteria, we used the non-parametric Kruskal–Wallis procedure. We described mean and SD in Table 3 and, if the data failed to meet the Levene's test criteria, Kruskal-Wallis p value was given. Reviewer #1 suggested to describe median and range rather than mean and SD for Kruskal-Wallis analyses. We then entirely agree that, for Kruskal-Wallis analyses, median and range should be described rather than mean and SD, but we really believe that it is worthy for the reader to have a look to the mean and SD values. For this reason, we conducted Welch tests, as stated in Statistical Analysis section and we reported Welch test p values in Table 3 of the revised manuscript. We believe that reporting both median (and range), and mean (and SD) would make Table 3 long and probably difficult to understand. The Welch correction, is a widely used solution to testing for the difference in the means of two populations when the variances are unequal. Furthermore, many association studies included Welch tests in their statistical analysis. Moreover, means (± standard deviation) of carriers and non-carriers were compared using Welch's modified t test in at least one study recently published in BMC Medical Genetics (Ley SH et al., BMC Med Genet. 2011).
**Editorial staff comment 2.** PHASE program has been used for the haplotype analysis. This is an efficient method for inferring haplotype frequencies and reconstructing haplotypes. However, in presence of low LD, it is well admitted that assigning haplotypes to ambiguous multi-heterozygotes individuals should not be done when conducting association studies. It seems that this is what has been done. By contrast, computing posterior haplotype probabilities and incorporating them in weighted regression analysis (see eg haplo.stats package in R) or using multiple imputation haplotype programs (eg WHAP, THESIAS) are advocated.

**Response.** According to Editorial Staff suggestions, we detected the haplotype effect size using the Thesias program, which is based on the maximum likelihood model and is linked to the stochastic version of the expectation maximization algorithm. However, conventional statistical methods were performed as well. A Fisher exact test was used to compare haplotype frequencies. The effects of a particular haplotype load (0: no copies of the particular haplotype; 1: 1 copy; and 2: 2 copies) on continuous variables were tested using a linear regression. The haplotype effects analysis performed with Thesias software yielded similar results related to cathegorical variables than conventional methods. However, we found an association with HW that was not identified using convensional statistical methods. In the revised version of the manuscript we show Thesias software´s results.

The haplotype analysis of quantitative metabolic traits performed with Thesias software showed some little differences with the previous analysis. We found association of haplotype CT with TG/HDL-C ratio and triglycerides. Association of haplotype CT with LAP became significant only after age and BMI adjustment and; a not previously observed asociation of haplotype CC
with LAP become apparent. In the revised version of the manuscript we show Thesias software’s results.

We reconstructed the haplotypes using PHASE 2.1. PHASE 2.1 implements Bayesian statistical method to infer phase and reconstruct haplotypes.\textsuperscript{10} The haplotype frequencies for JAK2 SNPs were estimated using the programs PHASE 2.1 and Thesias. The haplotype frequencies did not change substantially.

\textit{Editorial staff comment 3. The authors have studied quite a large number of traits (MS, HG, BMI, HDL, insulin, HOMA ). Even if these traits are not independent, the authors did not take this into consideration for the multiple testing issue. This should be discussed and their results tempered.}

\textit{Response.} We agree that all genetic association studies must address the issue of multiple testing and comparisons.\textsuperscript{11 12 13} While it is widely considered that the Bonferroni correction is a too conservative correction for most purposes, the standard \( p \) value of 0.05 is similarly too liberal. Approaches to this problem include a formal permutation analysis or (somewhat arbitrarily) choosing a more stringent \( p \) value for significance (e.g. 0.01 or 0.001).\textsuperscript{14} As stated in Statistical analysis section, we follow the recommendations of van den Oord and Sullivan\textsuperscript{15}, which suggest that a level of significance of \( p = 0.01 \) on average control the false discovery rate at 0.10. Furthermore, after applying the Bonferroni correction for multiple tests, the significance level was \( p< 0.025 \) (0.05/2 for 2 loci). Since our results represents a basic replication of previously reported findings, \( p \)-values presented in the current study are two-sided but not corrected for the number of test performed. Of note, the same criteria was applied by Hertel JK et al. in a study
recently (2011) published in BMC Medical Genetics.\textsuperscript{16} We point out in the Statistical analysis section of the corrected version of the manuscript that “since our results represents a basic replication of previously reported findings, p-values were not corrected for the number of test performed”; furthermore, we clearly stated in the discussion section as one of the limitations of the study that “p-values were not corrected for the number of tests performed”.

Multiple testing correction based on the number of test performed but not on number of SNPs are frequently found in the literature of association studies [e.g. “…remaining significant even after correcting for multiple testing (Bonferroni correction for testing 5 SNPs, p = 0.015)...”] (Rees SD et al., BMC Proceedings 2009).\textsuperscript{17}

Interestingly, there are evidence that false-positive rate is influenced by MAF (Tabangin ME et al., BMC Proceedings 2009). Common SNPs (MAF 25-50\%) may result in significantly fewer false positives than expected under the null $\chi^2$ distribution, suggesting that for common SNPs the current thresholds may be too conservative in genome-wide association studies \textsuperscript{\textcopyright}.

\textbf{Editorial staff comment 4. It is quite intriguing that some SNP associations reported in Table completely disappear after adjusting for age and/or BMI. Do they authors have any explanation?}"

\textbf{Response.} The study was conducted in a sample of 780 young male subjects, 18–65 years of age (mean age 37). Age is a known confounding variable so we adjusted for age when appropriate. We reported a 2 df- assoacition of both SNPs rs7849191 and rs3780378 with MS (contingency
Age adjusted association between rs7849191 and MS was not significant, but association between rs3780378 and MS remained significant taking into account the effects of age. It is intriguing that 2 df- association disappear after adjusting for age. However, when we performed 1 df- association tests (contingency table 3x2) statistical significance was achieved and remained significant after age adjustment.

Similar observations can be made regarding to Table 3. Significant differences between rs3780378 genotypes (overall differences between the three genotypes) in TG, TG/HDL-C ratio and LAP are reported; however, these observations were not confirmed taking into account the effects of age. Nevertheless we found that TT genotype carriers of rs3780378 compared to allele C carriers (TC+CC) (recessive model) showed a lower TG, LAP and TG/HDL-C ratio even at levels that take into account multiple testing and even after adjustment by age (Table 3).

It is possible that genetic associations are modulated by age. It is expected that younger subjects have been less exposed to environmental influences. To further explore this issue we performed analysis in the same sample but stratified by age (< 37 years vs. ≥ 37 years) (data not shown in the manuscript).

In the group of subjects over 37 years, both rs7849191 and rs3780378 were 2 df- associated with MS. The C allele of rs7849191 or T allele of rs3780378 (under both dominant and recessive models) showed a lower risk of MS. The adjusted analysis confirmed the associations (see Table). In contrast, in the group of subjects under 37 years there was no association of rs7849191 or rs3780378 with MS (data not shown).

Table. Association study between individual SNPs and MS in the sample stratified by age

Only significant results are shown.
In the group of subjects over 37 years, rs3780378 was 2 df- associated with HTG. The T allele carriers of rs3780378 showed a lower risk for HTG (under a dominant model or a recessive model). The adjusted analysis confirmed asignificant effect on the risk of HTG. Besides, in the group of subjects older than 37 years, rs7849191 was 2 df- associated with HTG. The C allele carriers of rs7849191 showed a lower risk for HTG (under a dominant model or a recessive model). The age-adjusted analysis confirmed a significant effect on the risk of HTG.

Moreover, in the group of subjects over 37 years, rs3780378 was 2 df-associated with HW. The T allele carriers of rs3780378 (under a dominant model or a recessive model) showed a lower risk of HW. The age-adjusted analysis confirmed a significant effect on the risk of HW (see Table). In contrast, the group of subjects under 37 years there was no association of rs7849191 or rs3780378 with HTG or HW (data not shown).

**Table. Association between individual SNPs and lipid metabolism-related phenotypes in the population stratified by age**

Only significant results are shown.
In the group of subjects over 37 years, we observed significant differences between rs3780378 genotypes in triglycerides, TG/HDL-C ratio but not in LAP. The TT genotype of rs3780378 carriers compared with C allele carriers (recessive model), showed lower levels of triglycerides, TG/HDL-C and LAP (see Table). However, these differences were not observed in the group of subjects under 36 years (data not shown).

Table. Single locus analysis of quantitative metabolic traits related to lipid metabolism in the sample stratified by age

Only significant results are shown.

<table>
<thead>
<tr>
<th>Group of subjects over 36 years</th>
<th>Variable</th>
<th>(mean ± SD)</th>
<th>p</th>
<th>p a</th>
<th>p c</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT vs. CC+TC</td>
<td>TG</td>
<td>133.80±69.75 vs. 174.18±114.09</td>
<td>&lt;0.001 b</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>TT vs. CC+TC</td>
<td>LAP</td>
<td>56.41±40.71 vs. 71.93±54.27</td>
<td>0.005 b</td>
<td>0.015</td>
<td>0.005</td>
</tr>
<tr>
<td>TT vs. CC+TC</td>
<td>TG/c-HDL</td>
<td>3.67±2.69 vs. 5.05±4.46</td>
<td>0.001 b</td>
<td>0.007</td>
<td>0.006</td>
</tr>
</tbody>
</table>

a = age adjusted. b = age and BMI adjusted. c = age and BMI adjusted. b = Welch.

These results suggest that the genetic associations under study are modulated by age. In particular, the associations previously found in the global sample only were replicated when analyzing subjects older than 37 years. It is possible that, in subjects below 37 years, the association is lost because a dramatic decrease in the prevalence of MS and related traits, with the consequent “dilution” of genetic variants that confer protection to these phenotypes. The effects of the SNPs under study substantially change when taking into account age. On the other hand, the prevalence of MS and related traits clearly increases with age.
We really believe that it would be very useful to analyse the impact of the SNPs in an older cohort in relation to MS and lipid metabolism outcomes to further support our findings.

Similarly, different causes could be implicated in the observed loss of statistical significance when corrected for BMI. It is well documented that BMI increases with age. Several population studies suggest that overall BMI increases with age up to the fifth or sixth decade after which BMI declines. 19 Furthermore, when comparing young and old subjects who have similar BMIs, the older person will have a greater percentage of body weight as fat. 20


