Author's response to reviews

Title: TCF7L2 variant genotypes and type 2 diabetes risk in Brazil: significant association, but not a significant tool for risk stratification in the general population

Authors:

Guilherme F Marquezine (g.marq2@gmail.com)
Alexandre C Pereira (alexandre.pereira@incor.usp.br)
Andre GP Sousa (agpsousa@iq.com.br)
Jose G Mill (jgmill@npd.ufes.br)
Whady A Hueb (mass@incor.usp.br)
Jose E Krieger (krieger@incor.usp.br)

Version: 2 Date: 7 October 2008

Author's response to reviews: see over
Dear Melissa Norton,

Editor in Chief,

BMC Medical Genetics

Response to Referees

Please find attached a revised version of our manuscript entitled “TCF7L2 variant genotypes and type 2 diabetes risk in Brazil: significant association, but not a significant tool for risk stratification in the general population” that we would like to re-submit for publication in BMC Medical Genetics as an original contribution.

We would like to thank the referees for their insightful and helpful comments and suggestions. We have considered all suggestions and made changes accordingly. We believe these modifications have improved the overall quality of our work.

We hope our manuscript is now suitable for publication in BMC Medical Genetics and that this improved revised version will be interesting to BMC readers.

Sincerely,

Alexandre C Pereira,

Lab Genetics and Molecular Cardiology

Heart Institute – University of Sao Paulo, Brazil
Referee 1 comments:

Regarding the suggestion “…the authors analyze rs7903146 assuming a recessive genetic model which might both not be the correct one or which might be less powerful than the (log-) additive model – so I strongly recommend using the latter for the basic as well as for the ROC analyses”, we have recalculated the statistical analysis and, indeed, observed that the log-additive model has a better fit than the recessive model. However, there were no changes in results (basic statistics and ROC analyses), because the relative importance of the T allele in heterozygosis was insignificant (when compared to obesity, hypertension and age, for example) and even in TT individuals it was minimal. Nevertheless, this modification was performed and included in manuscript.

About the comment “the authors should give all genotype counts (CC,CT,TT) for each sample which offers the option to compare results more easily across studies”, we certainly agree and have included this information as text in the “Results” section.

As regards to the request “please clearly distinguish for which statistical tests hypotheses exist from the literature (a-priori information) and if the reported data point to the same direction”: all statistical tests calculated were performed as a two-sided test. In accordance to the literature, the rs7603146 T allele was associated with T2DM (pointed towards the same effect direction). We included this information in the “Methods” section.

Regarding the suggestion “for all statistical tests please report unadjusted p-values as well as point estimators and confidence intervals for the genetic effects”, we made changes in Table 3 in order to accommodate this request.

We also agree with the comment that “in case of sparseness of the data the authors should use exact tests (e.g. Fishers test instead of Chi-squared test)” and we performed these modifications (Fisher’s test was used whenever appropriate). This information was added into the new version of the manuscript.
Regarding the question “what is the purpose of including the highly selected MASS II data?”: the MASS II population is at high risk for type 2 diabetes, with a prevalence that exceeds 30%. As such, it is the closest to a case-control design as we could get (as opposed to the cross-sectional design of the MONICA-Vitoria population) and, therefore, we consider it instrumental in replicating the association results observed in studies from other populations.

The analyses of both populations jointly can be observed as following:

<table>
<thead>
<tr>
<th>Genotype (MASS2 + VIT)</th>
<th>DM N (%)</th>
<th>Non-DM N (%)</th>
<th>OR (CI95%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC + CT</td>
<td>246 (85.1%)</td>
<td>1543 (90.2%)</td>
<td>1.61 (1.12 - 2.32)</td>
<td>p= 0.0087</td>
</tr>
<tr>
<td>TT</td>
<td>43 (14.9%)</td>
<td>167 (9.8%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Although there is a significant association when both populations are merged, we prefer to analyze them separately because the populations have important demographic and clinical differences (prevalence of cardiovascular disease, for example). It is our opinion that analyzing both populations together would introduce heterogeneity in our analysis and led to the probably wrong conclusion of an overestimated OR associated with the presence of the TT genotype and diabetes risk in the general population. As a result, we prefer to maintain the results of two separate analyses.

Regarding the question “what is the practical purpose of the prediction model and to what population should it be applied (e.g. if it is a screening tool for T2DM it’s application to MASS II would make little sense and a variable like “obesity” should be assessed at the time of screening while T2DM status should be assessed later in time)?”, we believe that the prediction model is aimed at the general, low-risk, population (such as the Vitoria population sample). The MASS II population was used as a confirmation of case-control studies in a Brazilian population.

About the request “give a citation for the prediction model without genetics and give more
Detail on the variables (e.g. was “obesity” assessed at the same time of T2DM diagnosis?) and their weights for the prediction; a clear improvement of the manuscript would be an inclusion of other non-genetic prediction models (please give citations) and a comparison of their AUC (estimator and confidence interval) with and without rs7903146 genotype status”, we have another paper in submission process in which a more detailed description of methods and results can be found (Sousa, AGP et al., Derivation and external validation of a simple prediction model for the diagnosis of type 2 Diabetes Mellitus in the Brazilian urban population, submitted). This manuscript only describes the validation of a demographic model for T2D prediction, and has no information on the value of the rs7903146 marker in this scenario.). The weights (based on β coefficients) of each variable (including rs7903146 genotype, for this paper) are shown in the table below. In that study, the Vitoria population was cross-sectionally evaluated and, therefore, the variables (such as obesity) were assessed at the same time of T2D diagnosis. Finally, to use other prediction models that were derived from other populations can be inadequate, because the weight for each variable could be different.

<table>
<thead>
<tr>
<th>Variables</th>
<th>β Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>45-54 vs 35-44yr</td>
<td>0.966</td>
</tr>
<tr>
<td>55 or more vs 35-44yr</td>
<td>1.240</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td></td>
</tr>
<tr>
<td>25-29,9 vs &lt; 25</td>
<td>0.626</td>
</tr>
<tr>
<td>30 or more vs &lt;25</td>
<td>1.884</td>
</tr>
<tr>
<td>Hypertension</td>
<td>0.480</td>
</tr>
<tr>
<td>TCF7L2 Genotype</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>--</td>
</tr>
<tr>
<td>CT</td>
<td>0.023</td>
</tr>
<tr>
<td>TT</td>
<td>0.238</td>
</tr>
</tbody>
</table>

We have added more information about the prediction model to the revised manuscript in the Methods section, and a reference as unpublished data in the Results section.

Regarding the requests “please revise the statistical analysis part and use non-parametric tests instead of e.g. t-tests or ANOVA” and “please give detail on the power calculation”, we performed a review of all statistical analysis conducted. For statistical power calculation, we used the PS: Power and Sample Size Calculation from the Department of Biostatistics of Vanderbilt University, available for download at [http://biostat.mc.vanderbilt.edu/twiki/pub/Main/PowerSampleSize/pssetup.exe](http://biostat.mc.vanderbilt.edu/twiki/pub/Main/PowerSampleSize/pssetup.exe). This information was added to the manuscript in the Methods section.

We have performed all minor essential revisions.

About the comment “as your data nicely underlines the limits of predictive genetic testing, your discussion would benefit from a critical review of the recent papers of Chachi S et al. or A Cecile J W Janssens”, we included comments on both studies in the Discussion section. Actually the work of Balkau et al is very similar to ours and we reached similar results.

Regarding the suggestion “genotype distributions cannot be in Hardy-Weinberg equilibrium; correctly there can only be no evidence for (strong) deviations from Hardy-Weinberg equilibrium (report the smallest 2-sided exact p-value to underline your statement)”, we have stated the lack of HWE in the MASS II sample and the presence of HWE in the Vitoria sample. A probable explanation for this is the high *a priori* risk of diabetic individuals in the first sample (derived from individuals with multi-vessel coronary
artery disease). We have provided the 2-sided exact p-values for both tests in the Results section of the manuscript together with a brief commentary of this possible explanation.

And regarding the question “what would be the practical consequences of applying either predictive model – given an example e.g. for testing say 1,000 subjects (how many wrong decisions will there be?)”: a practical consequence of applying the predictive model can be exemplified by calculating the Needing to Additional Tests (as shown in Table 5), which means the proportion from population which would be tested for T2D according to the model. As shown, no significant difference in the model using information from this genetic marker, as compared to the model using only demographic variables, was observed[P2]. We included this question in Discussion section.

**Referee 2 comments:**

About the comment “As the power to detect the well established effect of the TCF7L2 SNP of T2DM in the population-based study group is extremely low (24%, see Results), the study group is too small to validly assess the effect of the SNP in a predictive test. Hence, these analyses should either be removed from the manuscript or the size of the study groups has to be increased considerably”, we agree with your comment. Indeed, the study is low-powered but we would like to stress that these results are an important exercise of the practical use of the genetic marker and an interesting way to display the relative impact of the genetic marker in comparison to other risk factors (as suggested by Referee 1). Therefore, we judge it is important that the analysis is shown on the description of our results.

Regarding the suggestion “The description of the population-based study group is to my mind too long. It should be shortened considerably”, we are in agreement with your opinion and performed this change.

We made all minor changes suggested.

Sincerely,
The authors