Reviewer's report

Title: Race-Ethnic Differences in the Association of HbA1c-Associated Genomic Loci with HbA1c Levels and Mortality in U.S. Adults: the Third National Health and Nutrition Examination Survey (NHANES III)

Version: 1 Date: 21 December 2011

Reviewer: David Meyre

Reviewer's report:

- Major Compulsory Revisions

In this report, JL Grimsby and colleagues assessed the ethnic differences in the association of HbA1c-associated loci with HbA1c levels and mortality in 3,041 black American, Mexican American and White American US adults. They conclude that minor allele frequency of some HbA1c-associated loci as well as their combined effect on Hba1c levels may vary according to the ethnic background. This is an interesting study that feeds the debate about the transferability of GWAS signals identified in Europeans to other ethnic backgrounds. The current version of the manuscript may however be significantly improved.

1-Please precise the study sample size in the abstract.

2-For the SNPs showing significant MAF differences according to the ethnic background, it may be useful to add a post-hoc test to know which specific ethnic background(s) significantly diverges from the other ones (NHW = NHB # MA, NHW # NHB = MA, NHW # NHB #MA) and to comment these results. Given the strong selection pressure by infectious diseases on erythrocyte-related genes it would be especially interesting to investigate if the MAF differences are randomly or non-randomly distributed among the different ethnic groups. Are the HbA1c-increasing alleles systematically more frequent in certain ethnic backgrounds?

3-In my opinion the fact that the HbA1c genotype risk score (GRS) was significantly associated with HbA1c level in White and Mexican Americans but was not associated with HbA1c level in Black American does not signify that the average beta-values per additional risk allele significantly vary according to ethnicity. Indeed, the significance of the association depends on statistical power and sample size considerations. To properly demonstrate an ethnic heterogeneity in the impact of the HbA1c GRS on HbA1c value, the authors need to apply the following linear regression model in the whole U.S. sample: Hba1c level (outcome) = sex, age, GRS, ethnicity, GRS x ethnicity interaction. IF a significant (P < 0.05) GRS x ethnicity interaction on HbA1c level is found, the authors may conclude that the combined impact of HbA1c-associated loci on HbA1c levels vary by ethnicity. If such interaction is demonstrated, it would be
interesting to further discuss the potential reasons why the average beta-value per additional risk allele on HbA1C is 2-fold bigger in Mexican American than in European Americans (gene x environment interactions...). If true, this result is important because it suggests that GWAS signals derived from European populations may not only be transferable to other ethnic backgrounds but may be more explicative at least in certain ethnicities.

4-Do the authors have access to data relative to incident T2D events during the follow-up? IF the answer is yes, it may be interesting to test the impact of the HbA1c genotype risk score on incident T2D risk.

5-The authors have demonstrated that the mean HbA1C genotype score was varying significantly with ethnicity. Is the dispersion of the HbA1C genotype score also different according to the ethnic background (extreme GRS values, dispersion of GRS classes...)?

6-T2D has been diagnosed using fasting glucose value but no OGTT was performed. This may introduce some misclassification in the T2D status attribution, and may be listed as a limitation of the study.

7-In order to provide a transparent quality control procedure please add a supplementary table with raw genotype counts and call rate values for each SNP in addition to HWE.

8-Please describe the number of SNPs harboring a directionally consistent effect on HbA1c value in comparison with those from the recent GWAS for HbA1c published in Diabetes, not only in the European Americans but also in the two other ethnic subgroups. Discuss the results in the context of ethnic-specific LD structures, flip flop effects...

9-Single SNP analyses: single SNP analyses in ethnic subgroups are clearly underpowered (only four out of 33 associations nominally significant) and must be discarded from the manuscript. To gain some statistical power in the analysis, I strongly recommend the following linear regression model in the whole U.S. sample: Hba1c level (outcome) = sex age genotype ethnicity genotype x ethnicity interaction. This may help to evidence inter-ethnic heterogeneity in the association of certain SNPs with Hba1c.

10-The mortality rates reported in this study are very surprising: 24.1% in White Americans, 14.2% in Black Americans and 9.7% in Mexican Americans. Are these highly divergent mortality rates in line with data from the literature? Did the authors have explanations for these results? The authors should provide the exact number of mortality events during the follow-up, in order to have a better idea about the statistical power of the mortality analyses.

11-Discussion: “Our analyses of differentiation and selection suggest that there may be some selection pressure at the ANK1, HK1, ATP11A, ABC11/G6PC2 and TMPRSS6 loci, all of which are erythrocyte-related loci.” This sentence is false; the ABC11/G6PC2 locus is related to glycemic control pathways.
12-Discussion: “Although hypothetical at this point, if race-specific selection pressures influence erythrocyte-associated SNPs, this could generate inter-race variation of SNP associations with HbA1c.” I do not necessarily share the view that selection processes may lead to inter-race variation of SNP associations with HbA1c level.

13-The authors found different alternative explanations to explain the lack of association between the HbA1c GRS and mortality, despite previous epidemiological association between HbA1c levels and mortality. In my opinion the main explanation for the lack of association of the HbA1c GRS with mortality is the lack of statistical power if analyses are done separately in the three ethnic subgroups. OR average values are directionally consistent with the epidemiological observations in the three subgroups, and a trend of association (P=0.09) is observed in the larger sample (White Americans). This alternative explanation has not been mentioned by the authors. In order to gain some statistical power, I recommend using the following logistic regression model in the whole sample:

mortality (outcome) = sex, age, GRS, ethnicity, GRS x ethnicity interaction.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests.