Reviewer’s report

Title: Association of NTCP polymorphisms with clinical outcome of hepatitis B infection in Thai Individuals

Version: 0 Date: 30 Dec 2018

Reviewer: Miguel A. Garcia-Perez

Reviewer’s report:

In the present study, Chuaypen N. et al., study the association between two polymorphisms (SNP) in the sodium taurocholate co-transporting polypeptide (NTCP) gene with the risk of infection by the hepatitis B virus (HBV) and with the clinical outcome of infection, in a population of Thai individuals. To this end, the authors genotyped these two SNPs in a sample of 1021 individuals; 205 healthy controls; 206 subjects with spontaneous HBV clearance; and in 610 patients with chronic HBV infection (CHB), 305 of them with hepatocellular carcinoma (HCC) and 305 without HCC. The controls are age and gender matched.

The authors describe a lower frequency of both the A allele and the GA + AA genotypes of the SNP rs2296651 in the CHB patients, there being no differences between the CHB and HBV clearance groups. Within the group of CHB patients, those with HCC also showed lower allelic and genotypic frequencies than patients without HCC. The CHB patients with GA + AA genotype showed lower frequencies of positivity for HBeAg and the HCC patients of the same genotype also showed lower levels of HBV DNA.

The study demonstrates an association previously described in other populations, mainly Asian. In this sense, it is not novel. However, given the high frequency of the A allele in the Thai population and the endemic character of HBV infection, the present study is relevant. However, the paper presents some issues that should be explained as they reduce the interest in this work.

Concerns:

1) Much of the possible readers of the journal BMC Medical Genetics, are not experts in the subject of this work so some data should be explained conveniently. For example, it should be explained why there is a percentage of CHB patients who are seronegative for HBeAg, if the positivity for this antigen has been used to define CHB patients (page 5, lane 39). In the same line, the HBV clearance group was defined as negativity for HBeAg (and positivity for HBV core and anti-HBs). However, Table 1 indicates that there is no HBeAg data for this group of individuals. Please correct this.

2) In the present study, 2 SNPs and several comparisons between groups are analyzed, so a test for multiple comparisons should be considered. The p-values described are sufficiently low so that the results are substantially affected by the application of this test. On the other hand, the manuscript would gain in statistical solidity.
3) The authors state that the whole cohort were not deviated from Hardy-Weinberg Equilibrium (HWE). The authors must provide the p-value for the HWE. The most common causes of HWE departure are errors in the genotyping or association of an allele to a phenotype in a subgroup or in the whole population. In this sense, what was the value of HWE in healthy controls? And in the patients?

4) The SNP rs2296651 is a polymorphism for which functionality has been studied. The authors should have more emphasis on the description of this functionality in relation to the genetic association detected.

5) The authors should discuss why the frequency of allele A for SNP rs2296651 is so high in the Thai population. They should also discuss a possible evolutionary advantage in carriers of the A allele in that geographical area.

6) Although the number of homozygotes for the A allele is very low in both patients and controls to draw conclusions, it is striking that this frequency is higher in CHB patients although the differences are not significant. That is, there seems to be no protection for this genotype. Have the authors considered or tested an overdominant inheritance model? On the other hand, the formation of homo- and heterodimers for this protein has been described. A possible explanation for this could be that the polymorphism had a kind of negative dominant effect so that the heterodimer (G / A) is less functional for the entry of the virus than either of the two homodimers G / G or A / A.

7) The authors should discuss the causes of the lack of association of rs4646287 SNP with the risk of infection, since this association has been described in other populations and rs4646287 SNP seems functional being associated with lower expression of the NTCP gene.

8) Please correct the SNP of the conclusions. Such SNP has not been analyzed in the present work!

Lesser issues:

1) Please avoid the use of non-defined acronyms in the abstract (eg HCC), or define it previously.

2) In order to gain clarity, the abstract should contain less numerical information (odds ratio, 95% CI, etc.).

3) Please indicate if the TaqMan assays were commercial or designed by the authors. In relation to this, the abstract indicates that the SNPs were analyzed by allelic discrimination. In this case it would be necessary to indicate the technique used (e.g. "... by using TaqMan probes"). Also, briefly comment on the genotyping technique since the reference provided is not easily found.

4) The aminotransferase levels of Table 1 are not discussed in the text of the manuscript. Please indicate the information provided or delete this data in Table 1.
5) The data in Table 1 have been analyzed by means of an ANOVA. However, the use of this test does not appear in the corresponding section on Material and Methods.

6) English must be reviewed throughout the manuscript

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

Yes

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Yes

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

Yes

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

**Quality of written English**
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?
4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal