Reviewer's report

Title: Insights into the genetics of blood pressure in black South African individuals: The Birth to Twenty cohort

Version: 1 Date: 16 Oct 2017

Reviewer: Cristian Pattaro

Reviewer's report:

Dr. Hendry and colleagues present a study investigating the genetic basis of blood pressure in black South African subjects, based on the Metabochip genotyping array. The lack of African data in the panorama of genetic association studies makes this study of high interest despite the limited sample size. However, the results were described only at the surface level. While several options stand in front of the Authors, that might help with data interpretation (e.g.: result comparison with African American studies or with other ethnicities), none of them have been taken into account.

Major issues:

1. Results are extremely concise and not very informative. Important information such as the distribution of study participants (which is essential to understand the Results) and regional association plots of the key loci, should not be put in the Supplement but rather presented in the main paper.

2. The Authors should make attempts to put their results into a more general context. The first option is to check whether the main findings show some level of replication in large African American studies (e.g.: Liang et al, PLoS Genet. 2017; PMID: 28498854) or in European-based studies (for instance, an interrogation of the PhenoScanner database, reveals a marginal association of rs1345441 with DBP and SBP by the ICBP consortium, PMID: 21909115). Queries to publicly available GWAS databases could be made as well (e.g.: GRASP https://grasp.nhlbi.nih.gov/FullResults.aspx, GWAS Catalog). The PhenoScanner can be interrogated also to assess the pleiotropic involvement of the uncovered variants. Linkage disequilibrium analysis with known variants should be done. Authors cannot state "so it may be that rs4842666 is in LD with..." (page 8, lines 186-7): available data (eg: the 1000 Genomes Consortium) now allows at least approximate analyses in similar ethnic groups and such investigations are now routinely performed by similar studies.

3. Discussion is extremely long and should be shortened or made more pertinent to the results. For instance, the discussion of the NO synthesis is excessively long.

4. Depending on the prevalence of hypertension, the analysis of hypertension should be either included in the methods and results or completely omitted. It should not be presented in the discussion, without appropriate methodological support (definition of hypertension, case:control distribution, statistical model for analysis, etc.). Should the Authors decide to keep hypertension
into the paper, I would recommend LD analysis with rs4930130 and, if appropriate, a meta-analysis.

5 - In the "Study Participants" subsection, the sample selection process and phenotyping should be described in more detail. Given the analyzed sample is smaller than the original sample of the Bt20 study, did any sample selection happen to the data? How did such selection might have impacted the current results (this would be appropriate for the Discussion section). Given blood pressure was measured in different waves, which time point are current data referred to? From the study protocol paper (ref #6) it looks like blood pressure was measured at the age of 5, 7, 11/12, 13, 14, and 15 years. However, the mean age of children was of 17.9. This is unclear and should be clarified.

6 - The same issue described in previous point, applies also to DNA: to which time point does extraction refer to? If taken at different time point, how was this accounted into the models?

7 - Anti-hypertensive treatment (AHT) should not be an issue in this study, given the young age of participants. However, was AHT taken into account? How?

8 - In the Methods section, page 6, a cascade of genotype quality controls is described. However, it is not clear the consistency of these operations. I counted at least 4 steps: 1st (lines 123-125), 2nd (lines 125-127), 3rd (127-131), and 4th (131-132). Some redundancy is apparent. Why were two different missingness rate threshold used for caregivers (2%) and participants (3%)? This might have introduced false positive association. How were "population outliers" defined? On line 131: please, define "a few SNPs" with an exact number. On lines 132-134, it is unclear to what kind of data the Authors refer to: phenotype analysis and selection should be moved to the appropriate paragraph and numbers for exclusions should be given.

9 - In the Association Analysis (page 6, lines 137-etc.) how were the two samples (caregivers and participants) accounted for into the linear mixed models? In fact, such a stratification might have induced spurious results and relatedness-based correction might not be sufficient to prevent it. Did the Authors consider adding a fixed-effect term to the model or treating the analysis as multilevel?

Additional essential issues:

10 - Results section, page 7, lines 159-160: how data were merged and how the number of SNPs for analysis was indentified should be moved to the Methods section.

11 - Results section, page 7, lines 162-163: what do the Authors mean with "Regions that contained two or more SNPs associated with either SBP or DBP or that were associated with both traits were examined further"? From the Methods section, I would understand that all SNPs passing the chip-wise significance threshold would be considered as significant. If the Authors introduce a new rule that there must be at least two SNPs for the locus to be considered valid, this implies the real significance threshold to be lower than that reported. It is also not clear
which "further analyses" the Authors mean, as no analysis was performed except reporting the top loci into a table.

12 - I found the "Power calculation" section not really relevant and could be omitted. Half of it states obvious things (e.g., lines 153-155), significance alpha is not given, genetic model not described, measurement unit is not given, etc.

13 - Abstract: Page 2, lines 29-30: the sentence is disconnected: it is not contradictory per se that Blood pressure is (partially) heritable and there are few studies on Africans. Maybe the contradiction is between the large amount of non-African studies and African studies.

14 - Abstract: Page 2, line 40: principal components of what? Geneticist know this is PC of the genotype information, but general audience doesn't.

15 - Abstract: Page 3, line 52: "loci" rather than "genes"?

16 - Page 5, line 114: "a few duplicate samples": please define exactly how many.

17 - Page 6, lines 139-141: the sentence is unclear, it mixes up a series of different concepts, and reference to R in the context of PCs looks not appropriate.

18 - Page 6, line 143: does "unlinked" mean that the markers should be in linkage equilibrium or independent? Please, use appropriate wording.

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

No

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

No

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:

Acceptable

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