Reviewer’s report

Title: Validation of genotype imputation in Southeast Asian populations and the effect of single nucleotide polymorphism annotation on imputation outcome

Version: 0 Date: 23 May 2017

Reviewer: Ali Torkamani

Reviewer's report:

The authors present an analysis of imputation accuracy in Southeast Asian populations. The manuscript is quite straightforward and very similar to a manuscript published by Krithika et al. "Evaluation of the imputation performance of the program IMPUTE in an admixed sample from Mexico City using several model designs" - which utilizes essentially an identical study design for samples from Mexico City. While not necessarily novel, the manuscript provides some useful metrics in terms of imputation strategies for Southeast Asian populations. I have a few minor clarification points:

1. The authors state that 5% of SNPs from each population were randomly selected and masked. Is this the same set of SNPs per population or were different SNPs selected per population. It is important to clarify this point as the random selection of SNPs could potentially have an impact on the results - especially given that random masking was only performed 5 times. This could lead to some uncertainty in comparing the performance of imputation across populations.

2. It would be nice to provide the imputation accuracy for the set of SNPs imputed by both HMII and 1000G (the overlapping set - presumably the HMII set). As written now the results give the impression that imputation accuracy is lower for the 1000G reference whereas it is more likely that imputation accuracy is comparable (maybe superior) for 1000G when limited to the set of SNPs interrogated by the HMII imputation. 1000G is probably allowing for the imputation of a larger set of rarer and more difficult to impute variants.

3. This is a discretionary revision: the GWAS results really do not add much to the manuscript, and as written are a little confusing. The abstract states a significant association with Dengue severity was found, but the body of the results report a significant association with platelet count. Is this platelet count during the course of fever - and is that a measure of disease severity? It is unclear whether platelet count is in anyway related to Dengue fever - perhaps some underlying assumption about the relationship is made by the authors and not stated here. Regardless, I do not believe these results add much to the manuscript. Is the association with platelet count novel? I would prefer to remove these results entirely and expand further on the imputation comparison.
Finally some suggested thresholds for allele frequency, or other features that could be used to pick out SNPs that can be assumed to impute accurately would be well received. This manuscript should act as a resource for other researchers in the field - it would be nice to have guidance beyond the simple fact that HMII is potentially the better reference set for this task.

**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:

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.