Reviewer's report

Title: An assessment of the portability of ancestry informative markers between human populations

Version: 1 Date: 23 March 2009

Reviewer: Indrani Halder

Reviewer's report:

- Major Compulsory Revisions

The authors are commended for drawing attention to an important topic. The paper itself however needs extensive reorganization and clarifications to provide the impact the paper wishes to make. The general methods used are adequate, but could be improved upon. The writing shows a lack of understanding of some of the basic concepts, but that could partly be due to the writing style, which can be easily edited. Specific comments are provided below:

1) In the Introduction section the authors have argued that AIMs discovered in one population may not be the best ones for use in a different population, but have provided no examples. Since most AIM panels and analyses reported have been done in continentally admixed groups like Hispanics and African Americans, where portability is less likely to be an issue, it is vital that the authors support their assertion that AIMs used in one population have been used in another population without much testing, with specific examples. 3 examples appear at the very last paragraph of the discussion (46, 47 and 48). It will make more sense to cite these in the introduction to make the case stronger for the paper.

2) Page 4, second para: Authors state that a small number of AIMs can be used to infer individual ancestry and cite review by Tian et al., 2008 in support of this assertion. How “small” is the panel that will infer accurate individual ancestry estimates? References 22 and 23 both identified thousands of markers for detecting stratification in Europe. The smallest marker panel mentioned in Tian et al., review includes 96 AIMs for distinguishing within Europe. The Kosoy et al., 2009 paper, which is referred by Tian et al., 2008, contains a minimum of 28 AIMs for distinguishing between continents and 192 AIMs for detecting North-South gradient within Europe. Previous simulation studies by Pfaff et al., 2001 and Tsai et al., 2005 both showed that substantial number of markers are required for accurate individual ancestry estimation, even for between continental admixture. Hence, why do the authors assume that 10 SNPs are sufficient for providing accurate individual admixture estimates in a relatively homogeneous sample? Since the authors are claiming that portability of AIMs across populations is an issue using the panel of AIMs they have selected, it is important that they first demonstrate that this panel does indeed provide accurate estimates of individual ancestry within the population that they were originally
Additional modeling/simulations are needed to show this. If one needs at least 192 AIMs for detecting stratification related to ancestry in Europe, how is it possible that 10 AIMs are sufficient for accurate ancestry inference within Britain (which, the authors argue is more homogeneous) and why would these AIMs be useful at all in any other population (closely related European or otherwise)? Given the numbers of AIMs already available, it is more likely that these 10 AIMs would be augmented considerably when being used in another population.

3) Page 5, para 2: Authors have cited differences in population from Dresden and Munich to argue that stratification exists even within closely related groups, which is a valid argument. But they have then argued that this is a reason why AIMs identified in one group may not be appropriate in another closely related group. Are the differences between more closely related populations necessarily due to differences in genetic ancestry (and hence amenable to ancestry analysis) or due to drift? Heterogeneity within a population can arise due to several reasons: drift, mutation and migration being some common ones. How can one distinguish between genetic heterogeneity caused by drift and one caused by systematic differences in ancestry? The rationale used here is confusing and needs further clarification. The authors assume that all heterogeneity in a sample can be attributed to ancestry differences. That is certainly possible, but how far back in time should one look to determine shared ancestry within a more homogeneous sample? Why would other methods that lead to population stratification not be more effective in such cases? The authors appear to contend that AIMs are most suitable markers for determining stratification using any method, but do not explicitly say so. Maybe that is something to comment on.

4) Page 6, Methods Para 1: Why were only 13 SNPs selected? Was there a specific cut point? “Lowest P values from a chi square test” is too vague and needs some context in addition to the cited reference.

5) Page 7, Para 2: Description of HWE in different populations. Incongruous sentences. 2nd sentence of the paragraph says no loci were out of HWE. Next sentence says 7 loci deviated from HWE in a population. If a locus is found to be in HWE after Bonferroni correction, which lowers the threshold P value significantly, how does it violate HWE when considering a P<0.5 cut off? Please explain or check for errors.

6) Page 7, Para 3: In a previous paragraph authors mentioned 952 individuals in the CEPH-HGDP, but describe empirical distributions in 927 individuals. Which individuals were excluded and why?

7) Methods: If the authors main premise is that AIMs selected in a relatively homogeneous group are not likely to be portable even within more closely related populations, why should the analyses not be restricted to populations of European ancestry? What additional information is gained by testing these samples in Asian or African populations? If authors describe these markers as “BritAIMs”, why would one assume that they would be used in an Oceania or African population without additional markers to represent the genetic variation in Europe as a whole? Please clarify in the Discussion section.
8) NO Tables were attached to the text. Only supplementary materials and Figures in the original manuscript.

9) Page 8, Results paragraph 2: What is the rationale for examining whether these 10 SNPs show wide variation within continental populations? What is the likelihood that polymorphisms which show wide variation within a defined geographical region that is assumed to be more homogeneous, are also highly differentiated globally?

10) Page 9, Second Paragraph: What do authors mean by "unusually differentiated"? unusually differentiated in comparison to what? Is there a criterion that you would use to infer "usual differentiation"? In the same line: What do you mean by "as a set"? Do you expect all the 10 SNPs to behave similarly? Based on what you have already shown that is an unlikely expectation, i.e. all the SNPs will not show similar pattern of differentiation. Why is a mean Fst a good measure for the “set of AIMs”? median is a specific value and some values will be lower than that. The use of a median value effectively ignores the SNPs which have Fst lower than the median. So, how can one infer how the “set of AIMs” behave by comparing median Fst? It is also unclear how this experiment was performed. How were the 10 random SNPs chosen? Was only a single set of 10 SNPs used in 10,000 replicates? Did the authors test any other set of 10 random SNPs? More useful information would be gleaned if multiple sets of 10 random SNPs were used in multiple replicates for this experiment. Since the characteristics of these random SNPs have not been provided, it is unclear if any characteristics of any of these random SNP may have influenced the results. How do these random SNPs behave within Europe? More details needed on how the replicate samples were generated for the Worldwide samples and the continental samples. How many individuals from each worldwide population sample was used for the replication studies. If an algorithm was used to generate replicates, please provide a brief description or a reference (as supplementary material) so this experiment can be validated. Finally, in addition to comparing median value distribution, consider looking at range and mean Fst values.

11) Page 9, Last paragraph: These studies provide more information than the previous studies. But what they also show is that only 4 of the AIMs make significant contributions and thus cast doubt on the utility of the “set of AIMs” that the authors refer to in the previous paragraph. How can one reconcile these results to the results for the set of AIMs?

12) Page 10, Discussion Paragraph 1: While the premise of the argument regarding portability of AIMs is very valid, authors need to examine whether the evidence they have used (refs 12, 13, 21, 22, 25) are proposing that those reported AIMs should be used to detect stratification within a narrower geographical region. The continental AIMs clearly are not meant to be used for within continental differentiation, and although they may sometime show within continent differentiation, very rarely are they proposed for specifically detecting ancestry within continents. Second, the number of AIMs required to detect more fine scale differences in ancestry is substantially higher than that required for continental differences. AIMs represent an ancestral group and defining a continental ancestor can be very difficult, so previous studies have sometimes
used multiple populations within a continent to capture ancestral allele frequencies representative of that continent.

13) Please provide definite examples where AIMs detected in a broader sample has been used to detect stratification in a more homogeneous group.

14) Discussion Paragraph 2: Sentence construction is confusing and unclear. Need more information on the 10,000 samples. What do you mean by "expectation at random"?

15) Discussion Paragraph 2: Since the authors state that there is no a priori reason to expect that the allele frequencies of the 10 BritAIMs would show differences in worldwide samples, this argument and set of experiments could be presented differently. i.e., perhaps start with what one can expect, that the SNPs are less likely to show a lot of differences in worldwide samples, and then show that these results prove this expectation. The way these results are presented now detract more than they add to the paper.

16) Page 11, first paragraph: It is incorrect to call these 4 AIMs "outliers" since you have used a P of 0.05 as a cut off to select these AIMs. An "outlier" would be one which has a very high value with no other values near its vicinity. rs7696165 is the only one that could be considered an outlier. The three other SNPs just have values greater than 0.05 but there are several other random, unselected SNPs near those when you look at the empirical distribution. Hence, the finding that only 4 of the 10 SNPs are actually portable is not a new finding either. In stead, these results cast doubt on the effectiveness of the set of 10 BritAIMs as a “set” as the authors have previously tried to show. The difference in median Fst when compared to median Fst values in Europe could have all been entirely due to these 4 SNPs. This issue could be cleared up by looking at the Fst after excluding those which the authors identify as “outliers” and comparing them to the empirical distribution.

17) General organization of the paper can be substantially improved upon. Sections of the discussion (e.g. Page 11, paragraph beginning “Previous studies have provided evidence...”) Along with refs 40, 41 and 42) should be moved to the introduction section, as should references to specific studies in the last paragraph of the discussion. The Introduction as it is now written does not do justice to the rationale for undertaking these studies. The paragraph dealing with the study by Heath et al., can be shortened considerably and the second last paragraph of the Discussion follows naturally after the paragraph ending with the sentence “Thus the BritAIMs may be useful as AIMs in other groups of populations, but the patterns are often not systematic and their effectiveness in other samples would be difficult to predict a priori.”

-Minor Essential revisions:

1. Tables 1 and 2 are missing from the main text.
2. Definition of AIMs not clear. Please define AIMs clearly and state how and why they can be used to measure population stratification that is not related to ancestry. I.e. why would AIMs still be the markers of choice for detecting stratification related to drift or migration, especially in a homogeneous population.
3. Page 6, Methods Para 1: Use of postcodes to determine geographical regions in Britain: Explain whether this is a standard method for determining the basis for expected population stratification in Britain. One or two sentences would be very helpful to orient the readers.

4. Page 8; Results section, first paragraph: Sure, Fst is one measure of ancestry informativeness that is used commonly, but as Rosenberg et al., pointed out that other measures of ancestry like marker informativeness, In, is perhaps a better measure. Could authors provide In measures of these SNPs?

-Discretionary Revisions:

1. The authors introduce the topic by saying that population stratification is a significant issue to be dealt with the GWAS and state that PCA and SA are the two methods of choice for handling stratification issues. However, they then directly start talking about ancestry, which implies that variation in genetic ancestry is the sole cause of stratification. Is that a correct assumption? What about drift, migration and mutation, the “classical” sources variation that could lead to stratification that is probably more prevalent in more homogeneous populations? Sure, at some level one might argue that all these forces lead to or result in differences in ancestry, but the use of the term “ancestry” and “ancestry informative markers (AIM)” need to be used with caution and should not be a catchall term for all sources of genetic variation.

2. SA and PCA do not make inferences of ancestry directly, rather they both try to infer genetic relatedness. This genetic relatedness need not be due to similar genetic ancestry. Such spurious relatedness can happen by chance. That said the reviewer understands that both these methods are practically useful in partitioning genetic variation on some level and is being used successfully for such issues. But, the subtle distinction needs to be made in order to stress why AIMS should be/are used for detection of population stratification.

3. Page 8; Results section, first paragraph: Please start with a brief description of the BritAIMs.

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests.