Author's response to reviews

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

Authors:

Sean Myles (smm367@cornell.edu)
Mark Stoneking (stoneking@eva.mpg.de)
Nic Timpson (n.j.timpson@bristol.ac.uk)

Version: 2 Date: 15 April 2009

Author's response to reviews: see over
RE: Myles et al. manuscript re-submission

Dear editor:

We thank the reviewers for their comments. We have addressed the comments and provide responses to their comments below.

We look forward to hearing from you in the near future.

Best regards,

Sean Myles

Response to Reviewer #1:
1. As authors pointed out the result of the study was foreseeable since there is no a priori reason for expecting SNPs that differ dramatically in allele frequency within Britain to differ dramatically among continental groups. I would like to see the result of how AIMs identified to distinguish within a continental group say e.g. Europe perform when inferring ancestry and for correcting population stratification in different European populations compared to AIMs that are specific to each population (e.g. BritAIMs for Britain). This analysis will provide more useful information than the analysis performed by the authors since most of the AIMs panel available in the literature are for differentiating between continental groups or within continents but not for each specific population.

Response: We agree with the reviewer that an assessment of the portability of Europe AIMs to different European populations would be highly informative. The data we have collected, however, does not allow us to address this question. Nevertheless, we feel that our analyses of the portability of BritAIMs provide at least some insight into the question of AIMs portability in general and is worthy of publication in its present re-submitted format.

2. Table 1 and 2 are missing from the manuscript.

Response: We have included Tables 1 and 2 in the revised manuscript.

Response to Reviewer #2:
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.
Response: Firstly, we’d like to thank reviewer #2 for the extensive comments and suggestions. These have been very helpful in improving the manuscript. To address comment #1, we have moved the very last paragraph of the discussion to the introduction.

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 identified in. 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.

Response: We understand the reviewer’s concern here. We have removed the word “small”. Our intention was to highlight that a “subset” of markers from GWASs can be used to infer ancestry and that dense genome-wide marker data (e.g. > 100K SNPs) is not required to accurately infer genetic relationships among populations. We do not “assume that 10 SNPs are sufficient for providing accurate individual admixture estimates in a relatively homogeneous sample”. In fact, we are fairly certain that these 10 SNPs would not be sufficient to capture all of the ancestry information required to correct for population structure in association studies within Britain or within any population for that matter. The 10 BritAIMs we investigate were identified as the most highly differentiated markers within Britain from Affy 500K genotypes from 16,000 Britons (see Methods for description). Thus, we do not feel that it is necessary for us to do “additional modeling/simulations” to demonstrate that these 10 SNPs “indeed provide accurate estimates of individual ancestry within the population that they were originally identified in”. Our point here is that these 10 SNPs are in fact BritAIMs – they are highly differentiated within Britain. We understand that they do not constitute a full set of BritAIMs – more markers would be needed to make a BritAIM panel, for example. We feel that the logic motivating our study still holds. The 10 BritAIMs, although only a subset of a full panel of BritAIMs, are informative for population structure within Britain and it remains unclear how informative they are in other populations. Essentially, we are simply asking whether highly differentiated SNPs within Britain are also highly differentiated among other populations. These highly differentiated SNPs do not need to fully capture population
structure within Britain in order to be useful for the analyses presented. To make this clear to the readers, we have added the following to the final paragraph of the introduction:

“Although these 10 BritAIMs do not constitute a complete set of AIMs that fully capture population structure within Britain, they are nevertheless useful for evaluating the portability of AIMs across geographic scales.” From your other comments, we see that this misunderstanding has generated considerable confusion. We hope that this is now clear to you and that a general readership will not be similarly misled by our description.

We recognize that the more ideal way to do this is the other way around: to take a full set of European AIMs and see how well they capture population structure in the 16,000 Britons. That analysis, however, extends beyond the scope of the present study and is currently in progress by another group. We recognize that our small set of markers limits the impact of our analyses and we understand that this study does not comprehensively address the question of AIM portability. We do feel, however, that our short cautionary note is scientifically sound and worthy of publication.

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.

Response: We do not fully understand the reviewer’s comment. Regardless of whether systematic allele frequency differences between groups are generated by genetic drift or recent differential gene flow from a distantly related population, AIMs are selected to effectively capture these differences in allele frequency. Moreover, it is unclear how to clearly distinguish between “ancestry” and “drift”. If people in Munich and Dresden were completely separated with no gene flow out or in for a hundred generations, there would be systematic differences in ancestry between the two populations caused by genetic drift. The reviewer asks: “How can one distinguish between genetic heterogeneity caused by drift and one caused by systematic differences in ancestry?” The answer is that these two things are not really comparable. “Genetic drift” is a population genetics force and “differences in ancestry” are a consequence of this force, and the other forces the reviewer mentioned (e.g. selection, migration, mutation). Genetic drift is independent assortment which essentially determines shared ancestry among individuals in a population. If the reviewer means that ancestry is determined on shorter timescales than genetic drift, then it is unclear where in time one should draw the line. We would be happy to comment on this further in the manuscript, but it is not clear to us how this is
relevant or what point exactly needs to be made. We would be pleased if the reviewer could be more explicit and provide an example of what should be included if this is still desired.

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.

Response: We used the information provided by the The Wellcome Trust Case Control Consortium and we unfortunately do not have any more information about how these SNPs were chosen. These SNPs were not chosen as AIMs, but were simply identified as being in the most highly-differentiated genomic regions in the study. They were investigated in the original publication to determine whether they may have identified false positives due to population structure. Since we did not have access to the original data, we were unable to expand this set to a full set of AIMs.

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<0.05 cut off? Please explain or check for errors.

Response: No loci were out of HWE with a Bonferroni correction, which essentially just lowers the P value cutoff for significance because of multiple comparisons. Seven loci deviated from HWE without Bonferroni correction. We feel that it is necessary to report this number and to state that we checked the cluster plots for these SNPs, as the Bonferroni correction is very strict. We have clarified this by adding “(P < 0.05 without Bonferroni correction)” to make it clear that this P value threshold was without the Bonferroni correction.

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?

Response: The following statement appears in the Methods section: “To allow for an unbiased comparison to the empirical distribution, Fst for the 10 BritAIMs was calculated from the same set of 927 individuals from which the empirical Fst distribution was calculated.” The empirical data set includes only 927 individuals because of the thresholds the authors of that study used for missing data. We don’t feel it’s necessary to mention their thresholds for missing data and it should now be clear that we are making comparisons between identical sets of individuals.

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.
Response: We agree with the reviewer that it should be obvious that highly-differentiated SNPs within Britain are unlikely to be highly-differentiated between two African populations, for example. However, as we mention in the Introduction, studies have used SNPs with high Fst between Asians and Europeans to correct for population structure within Asia (*Hum Mol Genet* 2008, 17(6):835-843; *Hum Genet* 2008, 124(2):179). The fact that this degree of transferability of AIMs is often assumed is baffling to those with even a cursory knowledge of population genetics. We feel that including a worldwide analysis provides as comprehensive a look at these SNPs as possible and we hope that it will be read by those who make assumptions like the one described above. In addition, the population pairwise plots (Figure 4 and Supplementary Figure 1) demonstrate that, although these SNPs were chosen because they are highly-differentiated within Britain, they sometimes show large allele frequency differences between populations distantly related to Britain. We feel that this is worth mentioning, at least as a side note.

In summary, we make several comments about why a comparison to European populations is most informative and why comparisons to distantly-related populations are unlikely to show anything interesting. For example, upon showing that the BritAIMs are not unusually differentiated on a worldwide scale, we state: “This result was foreseeable since there is no a priori reason for expecting SNPs that differ dramatically in allele frequency within Britain to differ dramatically among continental groups.” After identifying BritAIMs that are unusually differentiated, we state: “These 4 BritAIMs are outliers in the empirical distributions from Europe and the Middle East. These two continental groups are assumed to be more closely related to Britons than the other continental groups included in the present study, and this result therefore suggests that some AIMs may be portable within a restricted geographic range.” We feel that the point the reviewer wishes us to make has been made explicitly, but welcome additional suggestions of how to do so better.

8) NO Tables were attached to the text. Only supplementary materials and Figures in the original manuscript.

Response: We apologize. Tables have now been included.

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?

Response: The expectation is that these 10 BritAIMs will not be unusually differentiated globally. We felt that this hypothesis was worth testing. Again, we feel that we made the a priori hypothesis clear in our following statement in the Discussion: “This result was foreseeable since there is no a priori reason for expecting SNPs that differ dramatically in allele frequency within Britain to differ dramatically among continental groups.”

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.

Response: It is clear that the reviewer did not fully understand our resampling procedure and we apologize that we did not state this more clearly. We have changed the word “unusual” to “highly” and the word “set” to “group”. We have also changed the wording slightly as follows: “To test whether BritAIMs are highly differentiated as a group, we compared the median Fst of the 10 BritAIMs to a distribution of median Fst values from 10 SNPs sampled at random 10,000 times from the empirical distribution.” We have also added the following statement directly thereafter: “This allows us to assess how differentiated the 10 BritAIMs are compared to a random expectation.” We feel that these statements should make it clear to the reader what we have done and we provide additional information for the reviewer below.

To determine whether the median Fst for the BritAIMs is high, this value was compared to a distribution of median Fst values from 10 SNPs sampled at random 10,000 times from the empirical distribution. This means that we sampled 10 SNPs at random from the empirical data set (2750 SNPs) 10,000 times and calculated the median Fst value at each iteration. This generates a distribution of 10,000 median Fst values which we can use as an empirical median Fst distribution. We then look at where our observed median Fst value from the 10 BritAIMs lies along this distribution and this informs us about whether the observed value is “unusual” (i.e. high). The P value from this comparison simply tells us where the observed median Fst value lies in the empirical distribution of 10,000 median Fst values. The median is the most appropriate measure as Fst is not normally distributed. The alternative would have been to perform a non-parametric test (e.g. Wilcoxon test) at each of the 10,000 iterations of the sampling scheme. However, our goal was not to assess whether the observed and expected distributions of Fst were different at each iteration, rather we aimed to generate a “null” distribution of median values as we are testing whether the median observed Fst value is higher than what is expected at random. What we have done in the present manuscript is a standard random resampling procedure, essentially identical to the procedures performed in several of our previous publications:

Myles et al. *Hum Genet* 2007, **120**(5):613-621.
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?

Response: The result from the set of AIMs was used to demonstrate whether the median Fst of the 10 BritAIMs was unusual in a worldwide Fst distribution and for each within-continent Fst distribution. The fact that the median Fst for the 10 BritAIMs was high in the within-Europe comparison is not that surprising: the BritAIMs are highly differentiated within Britain and may therefore be expected to be highly differentiated within Europe. In addition, two of the 10 BritAIMs have significantly high Fst values ($P < 0.05$) and several other $P$ values for the BritAIMs are fairly low in the within-Europe comparison (see Table 1). From simply viewing the $P$ value distribution for the within-Europe comparison in Table 1, one may expect a median Fst across these SNPs to be high compared to random expectations.

In summary, the results are what they are. We believe that it is informative to look at whether the group of BritAIMs is highly differentiated and to determine whether each individual BritAIM is highly differentiated within and between different continental groups. The fact that 4 of the AIMs have significantly high Fst values does not in any way “cast doubt on the utility of the set of AIMs”. As we mentioned previously, these 10 SNPs are not intended to be a “set of AIMs” in the sense that they will effectively capture population structure within Britain or any other population. We evaluated them as a “set” when calculating the median Fst, but this “set” should not be interpreted as a “set of AIMs to be used to control for population structure in association studies”. We hope that by using the word “group” instead of “set” we have clarified this issue.

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.

Response: We agree with the reviewer. It is unlikely that studies that have identified AIMs that differentiate among continental populations recommend that these SNPs be used to correct for population structure for the majority of association studies which are conducted at local geographic levels. Nowhere in our manuscript do we state that the authors of these studies recommend the use of their AIMs in populations not represented in the panel used to discover the AIMs. Our point is that other researchers have misunderstood how to use AIMs. Our examples from the two Fujimoto et al. papers demonstrate that some people erroneously do interpret the data in this fashion. Moreover, it is common for association studies to use AIMs discovered to
differentiate among European populations to correct for pop structure within a single European population (e.g. Sulem et al.).

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

Response: We have added an additional example and now provide three examples of this in the Introduction. We state: "For example, Sulem et al. [31] tested for the presence of population stratification in Iceland with a set of AIMs that distinguish between European populations [24]. To correct for population stratification in association studies in Asian populations, SNPs with high Fst between Asians and other continental groups have been employed [32, 33]. Similarly, correcting for population structure in Caucasians, Hu et al. [34] use 38 SNPs that are highly differentiated between continental groups. These studies demonstrate that the underlying assumption that AIMs are largely portable across geographical scales is pervasive." We have also added a line to the first paragraph of the Discussion: “The medical genetics literature provides numerous examples where AIMs discovered at one geographic scale are used to correct for population stratification at finer geographic scales [e.g. 31, 32-34].” And we have added a line to our Conclusion: “The assumption that AIMs are portable across geographic scales is pervasive [31-34].” Additional examples can be provided upon request (there is no lack of them!), but we thought this number would be sufficient.

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"?

Response: We hope that we have clarified this issue by our response to the previous comments about the resampling procedure.

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.

Response: It is never stated that we expect the BriAIMs to be unusually differentiated between continental groups. In fact, we made great efforts to ensure that the Introduction provide to the reader sufficient evidence that an exploratory analysis of our questions is warranted. As population geneticists, we are well aware that the null hypothesis should be that markers showing large allele frequency differences within Britain should not necessarily show large allele frequency differences between continents. Our reasoning for this assumption is laid out clearly in the Discussion directly after we state that this is in fact the case. However, we feel that, since there have been a number of studies that have violated this basic assumption, it would be worthwhile to frame our study as exploratory. If the reviewer feels that we have made a severe error in our presentation style in this case, we will consider revising.
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. Instead, 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.

Response: We have discarded the word “outlier” and have changed it as follows: “These 4 BritAIMs are found within the top 5% of the empirical distributions from Europe and the Middle East.”

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.”

Response: We have changed the general organization of the manuscript and have taken these suggestions into consideration.

Minor Essential revisions:
1. Tables 1 and 2 are missing from the main text.

Response: The tables are now included.

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.

Response: We have added the following sentence to our description of AIMs in the Introduction: “AIMs are characterized by substantially different allele frequencies between populations and can be used to estimate the proportion of an individual’s ancestry that is derived from these populations.”
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.

Response: We feel that our description is sufficient: “These highly-differentiated SNPs had the lowest P values from a $\chi^2$ test of allele frequency difference between 12 geographic regions of Britain defined by postcode.” We are unsure how we can make this any clearer to a reader or how a detailed description of the geographical distribution of postcodes within Britain contributes to the manuscript.

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?

Response: We agree that additional measures may be useful, but we are unwilling to dedicate the time and effort to calculate an additional statistic that is highly correlated with Fst and that will likely provide similar results. We are confident that we have made our point with the data we have and that our cautionary note is worthy of publication as is.