Reviewer’s report

Title: Determining genomic kinship coefficient seems unhelpful in distinguishing consanguineous couples with a high versus low risk for adverse reproductive outcome

Version: 2 Date: 26 December 2014

Reviewer: Alan Bittles

Reviewer’s report:

As reported by the authors, the findings of their study are both convincing and unexpected and they effectively undermine the hypothesis that, in consanguineous families with no family history of inherited disease but one or more children with an autosomal recessive disorder, the parents would have more DNA that is IBD than comparable consanguineous families with at least three unaffected, ‘healthy-only’ children.

Some points could usefully be clarified and where appropriate changed:

Major compulsory revisions

1. lines 218-222 and Figure 1: the MDS two-dimensional plot shows a very striking difference between the combined results obtained with the Saudi samples and those of the other ethnicities/nationalities. To what extent might the overall results obtained have been dependent on, or influenced by, the different analytical methodologies adopted (lines 134-139)?

2. lines 337-340: there appears to be an implicit assumption that all the affected cases were homozygotes for the causative mutation. Is this known to be so, including the 8/73 cases that were diagnosed biochemically or clinically?

3. The authors consider whether different results might be obtained in future case-control studies that concentrate on specific regions of the genome. Could a further possibility be the actual disorders that principally were studied? According to Supplemental Table S2A just four disorders: AR deafness (n = 8), Bardet-Biedl syndrome (n = 11), Congenital adrenal hyperplasia (n = 12) and Familial Mediterranean Fever (n = 9) accounted for 40/47 (85.1%) of cases from Tunisia and hence 40/73 (54.8%) of cases overall. Is there an explanation for the preponderance of these four disorders, especially in the Tunisian sample which accounted for 63.6% of all cases? For example, might their high prevalence be due to directed sampling, or simply to the accessibility of patient records for the four disorders in question? Or are AR deafness, Bardet-Biedl syndrome, Congenital adrenal hyperplasia, and Familial Mediterranean Fever all quite common in the general Tunisian population?

If the causative mutations for the four disorders are indeed present at high frequency in the general Tunisian population (and hence in the overall case sample), a significant excess in parental DNA IBD of the cases versus the
controls would not necessarily be expected, which could influence the results obtained. Is it possible that quite different results might have been obtained if the case parents were carriers of disease genes known to be genuinely rare in the 10 constituent study populations, and particularly in Tunisia?

Minor compulsory Revisions

1. lines 113-115: on what basis were the targets of 100 cases and 100 controls set? To what extent did the inability to attain these targets hinder the study and influence the findings?

2. lines 171-173: the King robust estimates are said to 'account for population stratification', which is important since significant genetic sub-division would be expected in a large majority of the study populations.

Besides the computer simulations described by the authors of the method, is supporting information available on the successful, prior application of the King robust method in the study of actual populations?

3. Table 2: to what extent can the lower estimates of genomic pairwise relatedness attained with the King robust method (and King homo) be ascribed to control for population stratification? Why was a King robust estimate not derived for the Tunisian cases only, which comprised 63.6% of the total?

4. line 214: ‘Substructuring in the population’: does this subheading refer to substructuring within and/or between the different ethnicities?

5. lines 226-228: in the King robust estimates the SD of the cases is given as 0.35, whereas in Table 2 it is 0.035.

6. line 244: refers to ‘a few couples with high R-values’. The actual number or percentage of these couples should be cited.

7. line 322: ‘intrafamilial’ rather than ‘intrafamiliar’?

Discretionary revisions

1. Should the title of the paper be altered slightly to read: ‘Determining the genome-wide kinship coefficient seems unhelpful….’?

Declaration of competing interests:

I have no competing interests to declare.