Reviewer’s report

Title: Genotype-Informed Risk of Coronary Heart Disease Based on Genome-Wide Association Data Linked to the Electronic Medical Record

Version: 1 Date: 2 April 2011

Reviewer: Vijay Nambi

Reviewer’s report:

Ding et al report on the effect of adding a genetic risk score in CHD risk prediction. The following are the major comments related to the manuscript:

1. The biggest limitation is the lack of outcomes which the authors acknowledge.

2. The authors use the Framingham risk score as published by Wilson et al. However, the clinically used ATP-III based risk prediction score is a modified version of this. Diabetes, for example, will be considered a “high risk equivalent” but is used as a variable in the risk prediction tool per their methods. Why did the authors choose to use the FRS (original) rather than that which is clinically used (ATP III)? Were the results similar if the ATP III version was used?

3. Better characterization of the study population is needed: The authors state that these were 1262 controls. How were they chosen? Were they consecutive? Was absence of cardiovascular disease based on history alone or history + ECG or...???

4. Most of the SNPs presented were likely associated with CHD in Whites alone. No information is provided regarding the racial make up, i.e. were all subjects included Whites? If different ethnicities were enrolled how were ORs/RRs estimated for them?

5. The authors need to explain why different strategies were used in estimating the Odds Ratio and 95% CI (i.e. one strategy for 4 and another for the remaining 11). Uniformity in general would be preferred.

6. The quality control for the genotypes in eMERGE is referenced; however a couple of key findings could be included as this would be important to the paper as not all readers will be able to access the reference.

7. The construction of the odds ratio/genotype effect from multiple SNPs is a little difficult to follow. Would suggest formal statistical review of the same as this is a very important part of the paper. The overall risk score ranged from 6-to 18 suggesting that all individuals had at least 6/22 of the risk alleles. The authors also suggest that having <12 risk alleles will be associated with a lesser relative risk than the population which is a bit confusing. What this suggests is that most (all) of the population have some risk alleles which in turn then throws up the question related to the value of these risk alleles when, even in their presence,
your risk could be less than that of the population.

8. The authors state in the results and again in discussion that they only chose SNPs associated with CHD but not with CHD risk factors? Is this correct? Looking at the SNPs on Table 2, surely LDLR and PCSK9 are associated with LDL-cholesterol. So am not sure if what the authors state is correct. If the argument is that LDL-c is not part of FRS, by considering both Total Cholesterol and HDL-c in the FRS, the LDL-c is accounted for. Further, LDL-c is clearly a CHD risk factor

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I work on some genetic risk scores as well and our group has used similar platforms in gentotyping but i have no financial arrangements etc (ui.e. no true conflict)