Questions on Browning and Browning (2007) *Am J Hum Genet* 81:1084-1097.

Rapid and Accurate Haplotype Phasing and Missing-Data Inference for
Whole-Genome Association Studies by Use of Localized Haplotype Clustering.

1. What where the analytical challenges that prompted this paper?

2. The paper refers to "multinomial methods" for haplotype estimation. Can you summarize how these methods work? What are the major weaknesses of these multinomial methods?

3. In the proposed graphical model of haplotype structure, what are the nodes and what are the edges? Once the model is built, is there a guarantee that any possible haplotype will correspond to a path through the model space?

4. The strategy for building the graphical model is detailed in a previous paper (reference 12). Can you summarize how it works?

5. As the number of individuals and markers being analyzed changes, how do expect computational cost for the method to increase? Is it possible to control these costs – for example, to reduce computational cost at the expense of accuracy or to increase computational cost to obtain improved haplotypes? How?

6. The paper states that the method makes the assumption of Hardy-Weinberg equilibrium. Why is this assumption needed?

7. Does the model allow for genotyping error? How could the model be modified to allow genotyping error?

8. The paper suggests sampling multiple haplotype pairs per individual during each iteration of the haplotyping process. Why is this beneficial? What is the impact of doing so on computational cost?

9. The authors describe their approach as a "stochastic EM" algorithm. What does this mean? Why is this a desirable feature?

10. The authors state that one can either sample the most likely haplotype configuration or a series of alternative solutions, based on probability. Can you think of a different haplotype solution that might improve upon these?

11. What struck you most about the paper?