Questions on Howie et al (2012) *Nature Genetics* 44:955-959.

Fast and accurate genotype imputation in genome-wide association studies through pre-phasing.

1. What prompted the need for faster methods for genotype imputation?

2. What is the major difference between the methods proposed in the paper and previously used methods for genotype imputation?

3. In the proposed Hidden Markov Model, what is the set of possible states? What are the transition probabilities?

4. As the number of individuals and markers being analyzed changes, how do expect computational cost for the standard methods to increase? What about for the newly proposed methods?

5. The paper discusses potential benefits of using larger reference panels for imputation (both in terms of the number of markers and the number of haplotypes). Can you highlight two of these possible benefits? Can you find documentation in the literature of these benefits?

6. The paper describes the benefits of a multiple imputation based strategy. Why do you think that multiple imputation improves the quality results?

7. Reviewing table 2, can you summarize the factors that influence imputation quality and their relative contributions?

8. What struck you most about the paper?