Questions on Wu et al (2011) *American Journal of Human Genetics* 89:82-93

Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test.

1. The authors state that a variety of burden-tests that assess the cumulative effects of multiple variants in a region had been proposed prior to their work. Can you summarize some of the differences between these burden tests? And some of their shared weaknesses that motivated this work?

2. In contrast to standard burden tests, the authors state that the C-alpha test has certain features that make it more attractive. Can you summarize those appealing features are? And what are the weaknesses in C-alpha that motivate this work?

3. In addition to proposing a new statistical test, the authors state that their work provides several other important advances. What are these?

4. The authors state that it is important to include covariates, such as principal components of ancestry, in the analysis of genetic association data. What are principal components of ancestry? Why are they important?

5. A major advantage of SKAT is that it can be implemented as a score test. Why is this advantageous? How big an advantage do you think this might provide?

6. What the factors that should guide choices of Kernel functions and weight-matrices for the SKAT test?

7. The authors note that their Q statistic can be calculated as a weighted sum of single-variant score statistics. Why is this an important result?

8. What are the most interesting features you note in Figure 1?

9. What are the most interesting features you note in Table 1?

10. What struck you most about the paper?