Questions on Iqbal et al (2012) *Nature Genetics* 44:226-232.

De novo assembly and genotyping of variants using colored de Bruijn graphs.

1. What where the practical challenges that prompted this paper? What are some of the enhancements in the Cortex assembler vis-à-vis other de Bruijn graph assemblers?

2. What are the three application the authors chose to demonstrate their algorithm? For each of these, how does the use of de novo assembly present an advance in relation to previous methods?

3. How is knowledge of the reference genome sequence used in the proposed algorithm?

4. Focusing specifically on the discovery of polymorphic sequences, what are the major advantages of using de novo assembly based methods to discover these? Can you think of any disadvantages?

5. The authors discuss several distinct algorithms for identifying variant sequences … Among these are the bubble calling and path-divergence algorithm. Can you summarize key strengths and weaknesses of each of these two?

6. In the context of identifying variants, how does genotyping a population of samples help produce higher quality variant lists than genotyping a single sample?

7. A key parameter in the assembly is selecting the k-mer size. How does k influence power? How does it influence effective sequence depth? Be as precise as you can when describing the relationships.

8. Examine Figure 2. What are the most striking messages?

9. The authors identify the need for error correction as one of the open challenges for use of de Bruijn graph based assemblies. Any ideas on how this challenge might be tackled?

10. What struck you most about the paper?