Questions on Jun et al (2012) *American Journal of Human Genetics* 91:839-848

Detecting and Estimating Contamination of Human DNA Samples in Sequencing and Array Based Genotype Data.

1. The authors say that some unusual features (for example, excess heterozygosity) in samples from a type 2 diabetes sequencing study prompted them to explore the possibility of sample contamination. What are some of the consequences you might expect when analyzing a contaminated DNA sample?

2. The authors note that cross-species contamination is usually easier to resolve. How would you go about resolving cross-species contamination of a DNA sample? Do you expect the process to be perfect? Why?

3. The authors note that p-values and confidence intervals estimated using their maximum likelihood method may not be accurate. Why? Any ideas on how to improve the situation?

4. Intuitively, can you describe what some of the expected differences in distribution of sequenced bases between a contaminated and non-contaminated sample will be?

5. Do a little research on how array intensity values are usually converted into genotypes. Can you summarize how the process works? What are the advantages and disadvantages of carrying out these analyses one sample at a time versus one marker at a time?

6. The authors explore evidence for contamination in 3 different settings: with array data alone, with sequence data alone, and with sequence and array data together. Can you summarize some of the strengths and weaknesses of each option?

7. Checking for contamination using data from a single chromosome can be a lot faster. However, the authors urge some caution when doing this with related samples. Why?

8. What experiments did the authors carry out to validate their method? For you, what is the main take home message from each of these experiments?

9. Consider the likelihood for detecting contamination from DNA sequence data alone. How would you modify the likelihood if you expected read-depth to vary by genotype? Why would read depth vary by genotype?

10. What struck you most about the paper?