Anderson (2010). Data quality control in genetic case-control association studies. Nat Protoc. 5(9): 1564-1573.

1. What is a false positive (negative) association and how can a genome-wide study minimize these types of errors?
2. What is informative missingness?
3. What is the relationship between genomic coverage and the power of genetic association study?
4. How can mean heterozygosity be defined and what kind of information does it provide?
5. What is the difference between IBS and IBD?
6. What is meant by cryptic relatedness? Is it a concern when performing a population-based genome-wide association study? If not, explain. If so, how can it be dealt with in your GWA?
7. Is there a difference between QC-ing in candidate gene studies and genome-wide association studies? Explain.

8. What is the HapMap population? Can you describe it in a bit more detail? Why is it useful in the context of genetic association studies?
9. Which QC measures would you take differently (or in addition) when performing a genome-wide association study using multiple families?
10. What does genomic ancestry mean and can this information be integrated in a genome-wide association study?