Balding (2006). A tutorial on statistical methods for population association studies. Nature Reviews Genetics 7: 781-.

1. Name a few statistical analyses that should precede a genuine genetic association analysis
2. What are the differences between multiple SNP analysis and haplotype analysis?
3. What is population stratification?
4. What is the multiple testing problem and how can it be overcome?
5. What does HWE mean? How can it be tested? Why is it important to test? Which conclusions can be drawn from the test results and what not?
6. What is an “exact test”? When is it appropriate?
7. Name a few missing data approaches and list the pros and cons.
8. How can missing data jeopardize your analysis? Is it important to find out more about the missing data process itself? Why or why not?
9. What is an indirect association?

10. What is the common disease common variant problem? Does assuming a different hypothesis have implications for current genetic association designs?
11. How are measures of LD related to recombination fraction?
12. What is the difference between a candidate gene and a genome wide study?
13. How does fine mapping fit into genetic epidemiology?
14. Is phasing difficult? Can you name a few approaches/packages that are able to do the haplotype phasing for you?
15. What does tagging refer to and what are the main purposes for using tagging SNPs? Does the use of tagging SNPs have consequences for your genetic association analysis testing?
16. What is a trend test? Is it a useful test?
17. What are the differences between a linear regression analysis and a logistic regression analysis?

18. What do structured association methods refer to? Under which circumstances is it useful? Can you give pros and cons?
19. What is genomic control? Under which circumstances is it useful? Can you give pros and cons?
20. What are spurious associations? In which contexts is the term generally used?
21. What are null SNPs?
22. Is a haplotype test always more powerful than a single marker test?
23. What are the problems with using a Bonferroni correction?
24. What is FDR?
25. To what extent are frequentist and Bayesian approaches different?
26. What is a good genomewide significance level?
27. Given what you know about single markers and haplotypes, would you first try a haplotype analysis or a multimarker analysis (for instance,

adopting a multiple linear (or logistic) regression analysis)?

28. What is a haplotype block and how can this information be used in a genetic association analysis?
29. What are the advantages / disadvantages of targeting epistatic effects directly?
30. What is ascertainment? Is it important to have a closer look at this?