

Study Guide for Nov. 25 Exam in Bioinformatics I

1. Write a brief essay on the UCSC Human Genome Browser, Ensemble, and NCBI's resources for the human and mouse genome sequences. What kinds of questions can be answered by using them? Which sites are best for which uses?
2. Write a brief essay on computer methods, and particularly Web resources, for analyzing a given, unannotated genomic DNA sequence. What specific tools would you recommend? Include brief definitions of the important features that can be found by the methods you discuss.
3. What is an "alignment" of two sequences? What is the difference between a "local" and a "global" alignment? What is the "dynamic programming method" for aligning two sequences? Be able to apply it by hand to two short sequences for either a local or a global alignment.
4. What are the differences among blastn, blastp and psi-blast? Under what conditions can one expect psi-blast to do a better job than blastp? Briefly explain how psi-blast works.
5. Describe how "log odds" approaches can be used to determine amino-acid substitution scores that are appropriate for, say, blastp. Sketch the approaches taken for the BLOSUM matrices.
6. What are "weight matrices" for identifying, say, transcription factor binding sites, and how can they be determined from experimental data?
7. What is a "hidden Markov model"? What is meant by "the probability of generating a particular observed sequence"? What is meant by "the most-probable state-path for generating a particular observed sequence"?
8. Write an essay on the inner workings of GenScan.
9. Be able to read and interpret dotplots, PIPs and MultiPIPs.