FALL 2018: CAP 5510/CGS 5166 – Bioinformatics

[Exam Review]

Problems

1. [Short Questions]

   (a) (Lec 3) What is the *Genetic Code*? Suggest one bioinformatics application where it is useful.

   (b) State one implication of redundancy in the genetic code.

   (c) (Lec 12) Write down the formula for computing entropy.

   (d) Consider the entropy plot on Slide 33 on Lecture 12. Explain the dips in the plot in terms of what we know about GPCRs.

   (e) (Lec 5) How are the entries in a BLOSUM matrix computed?

   (f) Explain the terms *introns* and *donor site* In human splice junctions.

   (g) (Lec 8) What is the consequence of using ddNTP for PCR during Sanger sequencing?

   (h) (Lec 9) In using NGS for *metagenomics*, explain how to obtain the abundance profile of a microbial communnity.

   (i) (Lec 19-21) Explain the *perfect phylogeny problem*.

   (j) True or false? Perfect phylogeny and ultrametric methods assume that a change occurs only once, while maximum likelihood method does not. Explain your answer.

   (k) (Lec 18) Give a 1 sentence description of GWAS.

   (l) (Lec 16) Give an example each of *supervised* and *unsupervised* machine learning techniques.

   (m) What is the difference between a scaffold and a contig in genome assemblies?

   (n) (Lec 10) Define N50 as precisely as possible.

   (o) (Lec 11) In a *profile* matrix (also known as *position-specific scoring matrix*), how are background frequencies incorporated into computing the matrix values.

   (p) (Lec 11) What is a *sequence logo*?

   (q) What is the *TGCA* and the *GNF Atlas*?

   (r) What kind of information is stored in dbSNP?