MBB452 Genomics and Bioinformatics

Final Project:

An area of bioinformatics research that we did not cover in class is gene finding (GF). The basic problem of GF is to predict coding regions from sequenced DNA. Your final project is to either (pick just one)

(1) Write a ~1500 word research proposal focused on the problem of gene finding in *Homo sapiens*. When writing your proposal, consider potential pitfalls in human gene prediction; namely, large intergenic regions, many short exons per gene, large introns, alternative splice sites, alternative promoters, for example.   In the proposal, you must include the method(s) you propose to use, why you chose the method over others, and how you intend to verify its results. Your write-ups should be double-spaced and printed out with at least a 12pt font.

(2) Implement a GF algorithm for predicting coding regions in *Homo sapiens* (again – large introns, many short exons, etc.). You can use any programming language you want – but it must be sufficiently commented so that a reader (namely your TAs) can follow what you've done without too much difficulty and include a header paragraph describing the algorithm you've implemented. You will be graded on your code's legibility, demonstrated understanding of the algorithm, and algorithm's utility and robustness.

The project is due at the end of reading period (April 29th). You must turn in a written copy to Joann Delvecchio in Bass 432 and email an electronic version to your TA. We will post your projects on the course website sometime this summer as examples for future classes. This is project is worth 32% of your final grade. Good luck, and enjoy your summer!