Choose 3 human genes of interest, for each of which you can find sequenced orthologous upstream regions in at least 4 (and preferably more) other vertebrates. (Avoid the genes listed in the FootPrinter slides on the course web: you are embarking on novel discovery.) The upstream sequences should each be about 1000-2000 bp long, and preferably should have their 3' end at the transcription (rather than translation) start site, assuming that site is known.

Use the FootPrinter web server (available at http://bio.cs.washington.edu/software.html, where you will also find pointers to papers describing FootPrinter) to identify well conserved motifs in each of these 3 sets of orthologous genes. This will involve some experimenting with FootPrinter parameter settings.

Optional: If you have the time, do some literature search to determine the known regulatory elements of these genes. Which of FootPrinter’s predictions correspond to known regulatory elements?

For each of your 3 data sets, send the following items to tompa@cs.washington.edu:

1. an explanation of why you are interested in the genes you picked and the procedure by which you found the orthologous upstream sequences;

2. a FASTA file containing the upstream sequences;

3. the URL of your best FootPrinter result (the result pages aren’t discarded, so you don’t have to worry about your results disappearing);

4. anything you learned about known regulatory elements for this gene.