

## 7.344 Genomics and bioinformatics of gene expression

### General guidelines

1. Length  $\leq$  2 pages
2. Do not copy and paste from papers. It's boring for you and for us.
3. Due: Wed. 10/08/2003 at the beginning of the class or by email.
4. Be specific without getting lost in the details (e.g. it is not enough information to say, "I will make a model". You need to state what kind of model, what the inputs are, what features you use, how you train it, etc. On the other hand, we do not need to know at this stage what concentration of polymerase enzyme you will be using for this or that experiment).

### Specific guidelines

You are traveling somewhere in Africa when you suddenly discover this new exotic species that no one has ever characterized before. You name it *Serendipitum Africoli*. The first thing that comes to your mind (of course!) is where transcription starts for the genes of *Serendipitum Africoli*. Armed with your knowledge of biology and bioinformatics you take the first flight to your friend's lab at the University of Oxbridge. He has a very powerful computer as well as all the tools for experiments in molecular biology. First you call your friend Craig Venter and he sequences *Serendipitum* in two days of continuous work. In order to tell your Oxbridge friend what you are going to do, you (i) first draw a scheme of a gene, indicating the promoter, the transcription start site, the ATG, the exons and introns. Then (ii) you tell him how you are going to try to make computational predictions about where the transcription start sites are located. What kind of models are you going to use? What data are you going to use to train those models? What are the assumptions involved? Your friend is still reluctant to believe that you have made a major discovery just from the computational predictions. You then decide to go to the bench. (iii) What type of experiments would you conduct to discover where transcription starts?