

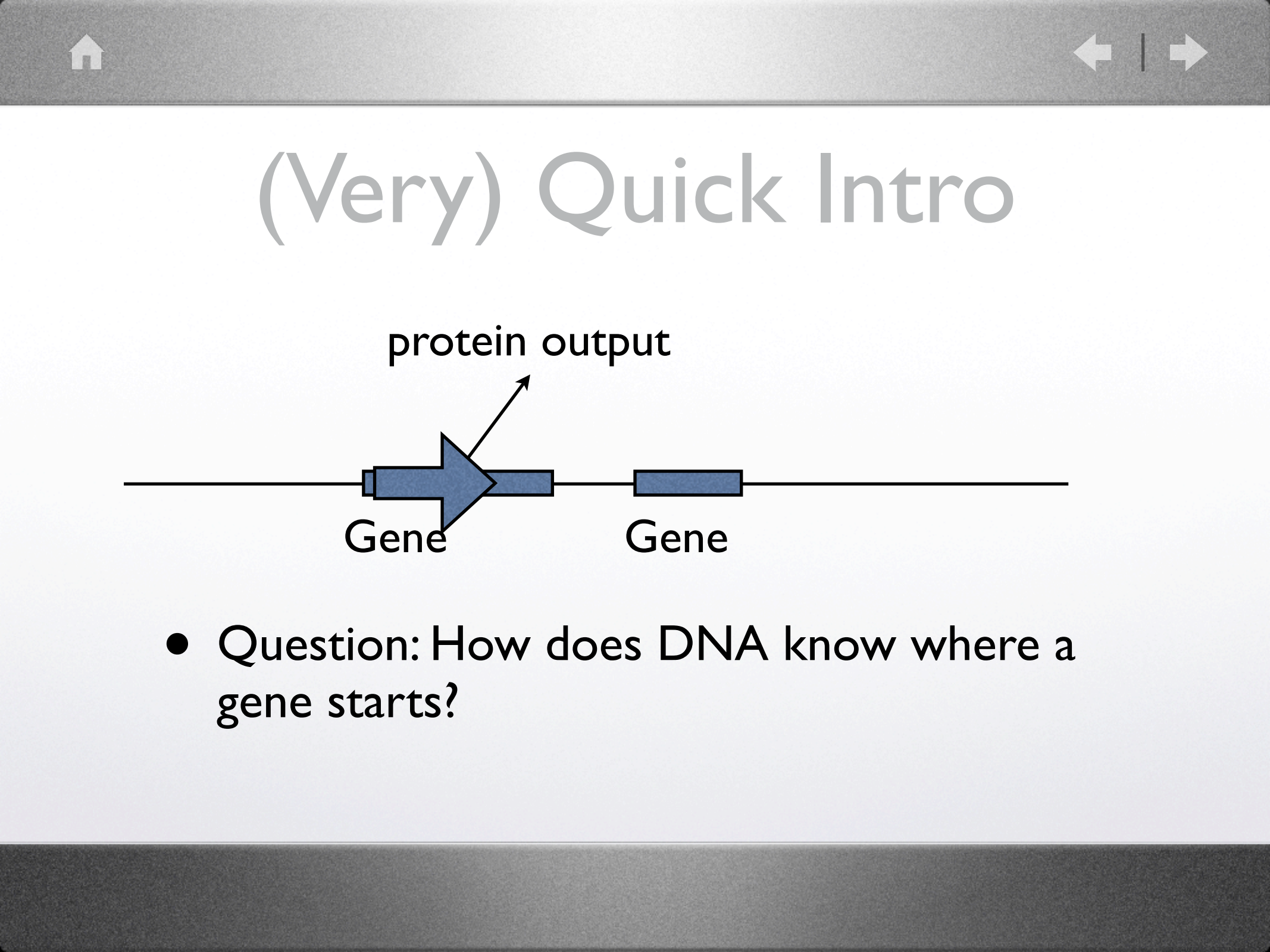

TFBS Prediction Set Visualization

Andrew A Carbonetto

CS533c Project Update


TBFS Prediction

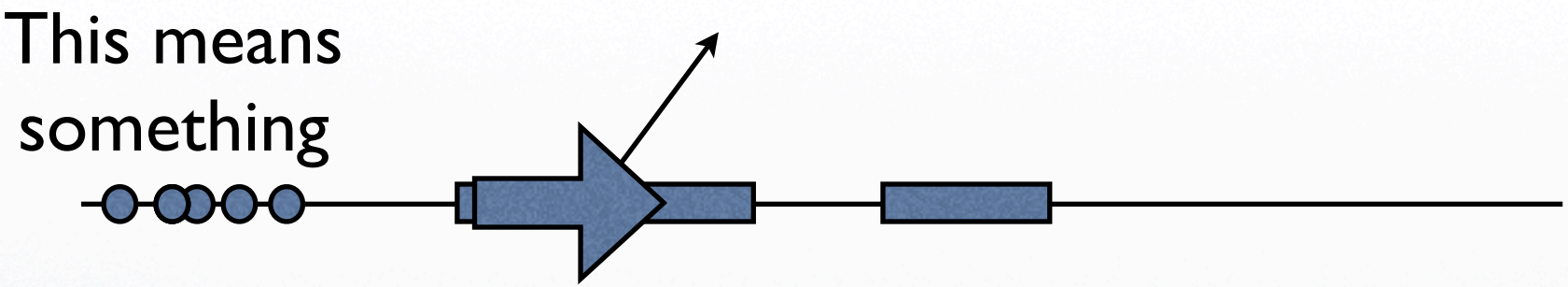
- Quick (very) introduction to problem
- How to tackle the problem
- Project Solutions



(Very) Quick Intro

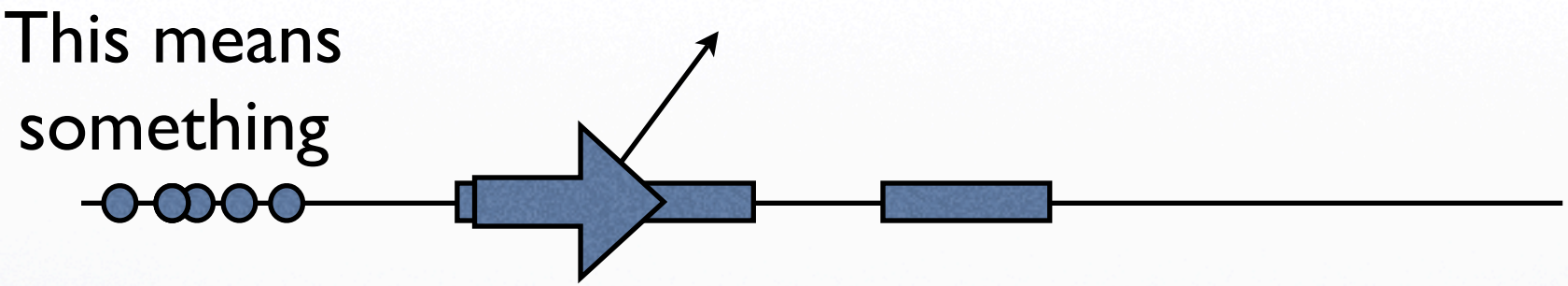
- Question: How does DNA know where a gene starts?

(Very) Quick Intro



- Question: A set of sites (where proteins bind to) on the DNA triggers this.

(Very) Quick Intro



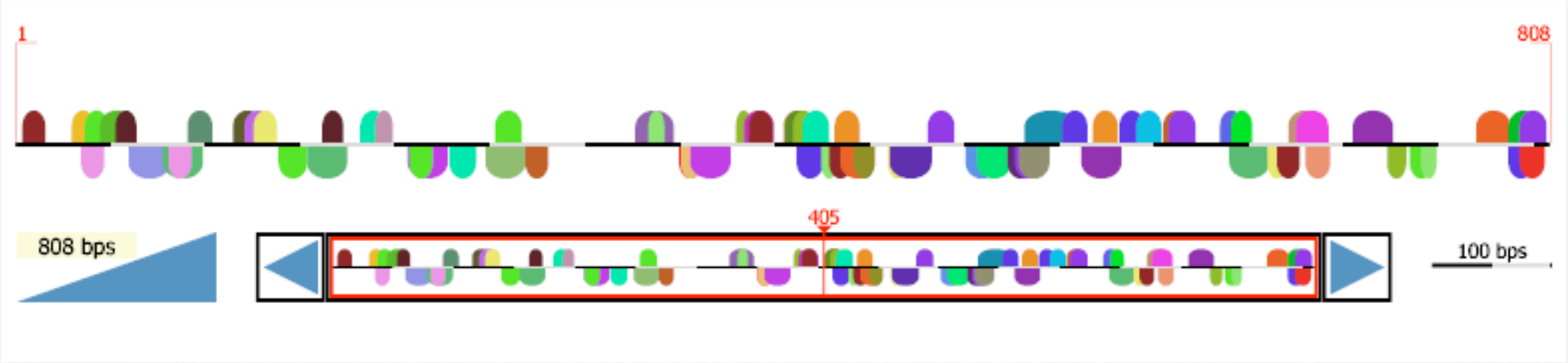
- We can predict these sites... badly.


Problem

- Everything in Biology has a certain level of uncertainty:
  - There are hundreds of types of sites
  - There is uncertainty to what the sites might look like
  - and which sites work together (patterns)

Problem

- So we tend to get a mess:



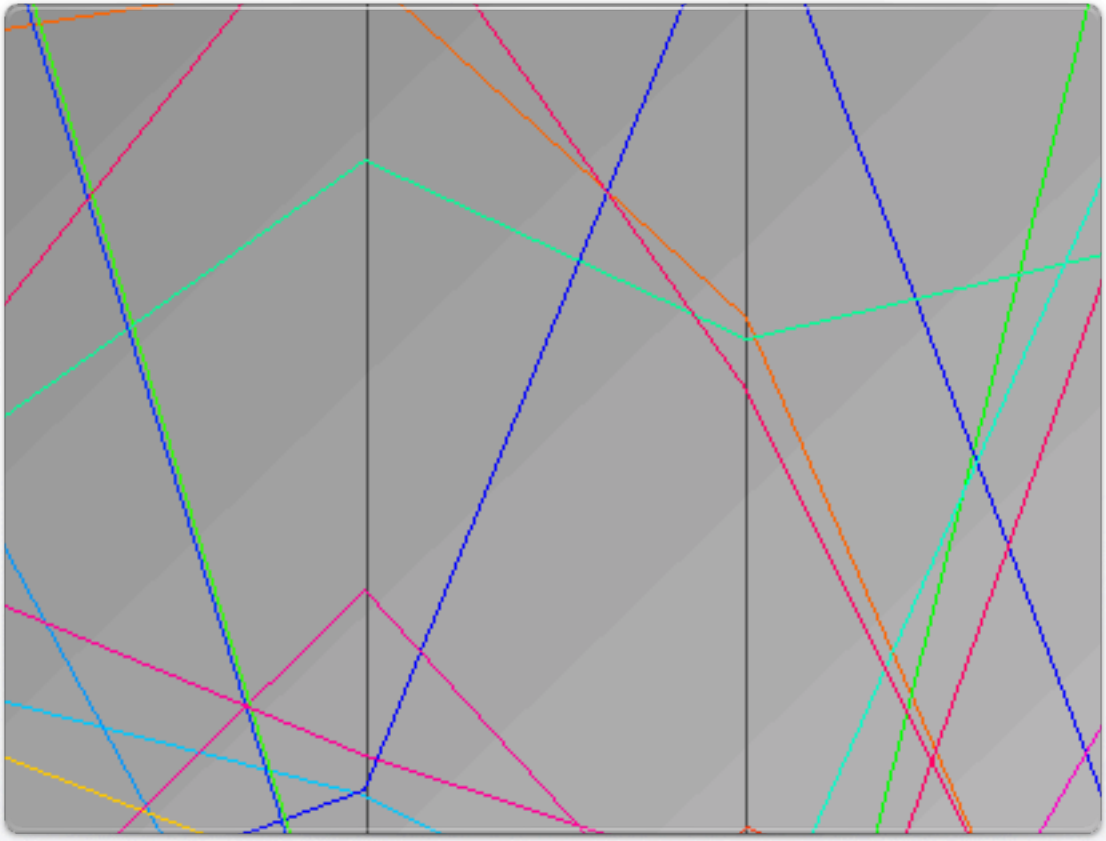
MatInspector at www.genomatix.de


Parallel Coordinates

- Build a Visualization so that multiple sequences of DNA can be put side-by-side to view the listing of predicted sites on each sequence.
  - idea: these patterns are conserved between several closely related species


Parallel Coordinates

- Then filter the results to get a more concise visualization (there are several methods of comparison between the sites)



InfoVis Toolkit