Announcements 5/3

- Read Lathrop and Smith (1996)
  - Protein threading
- Optionally read Rohl et al (2004)
  - Rosetta

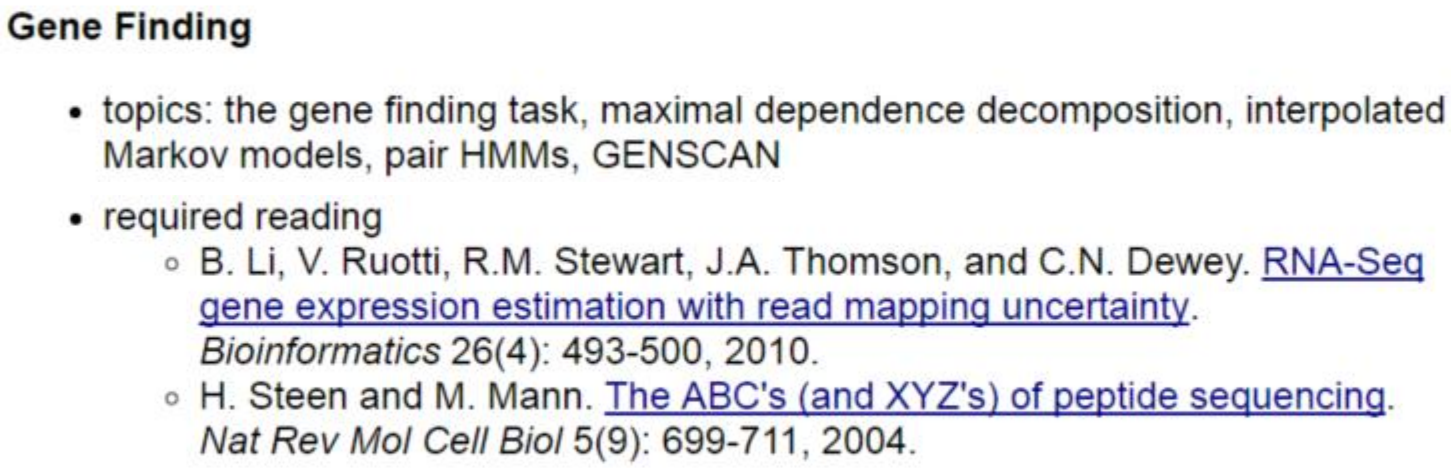
- HW2 grades posted
  - Piazza post about running the programming part
- Final exam Thurs May 12
- Extended office hours Tues May 10
  - 1:30 – 3:30 PM

Final exam

- Similar style as midterm exam
- No printed materials allowed
- Will not execute algorithms by hand
- Content from entire course, but *much heavier* emphasis on material not covered on midterm
  - 6 themes: RSEM through protein threading
  - May need prior material for comparisons, context
- *Slightly* more emphasis on material not covered on homework

Final exam

- Content includes (but is not limited to) material from lecture and required reading
  - Syllabus lists topics and require reading



Gene Finding

- topics: the gene finding task, maximal dependence decomposition, interpolated Markov models, pair HMMs, GENSCAN
- required reading
  - B. Li, V. Ruotti, R.M. Stewart, J.A. Thomson, and C.N. Dewey. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* 26(4): 493-500, 2010.
  - H. Steen and M. Mann. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* 5(9): 699-711, 2004.

  - Ignore alternative splicing, splice graphs, details of Inside/Inside-Outside/CYK algorithms, ROSETTA
- Optional reading is also helpful

Final exam

- Focus on concepts, strengths/weaknesses, algorithmic strategies
  - Why or in what cases would we use one method or experimental technology instead of another?
  - How does method X compare to method Y and what are the unique advantages of each?
  - What is important concept Z (e.g. mutual information)?
  - What does a method optimize?
  - What assumptions does a method make and what input does it require?