

Announcements 5/3

- HW4 due today
- Project report guidelines posted
 - No late days for project
 - Interpret the results, do not just list what you did
 - Pay attention to the rubric
 - Separate project_report and project_supplement handin directories
 - Do not upload your input data
 - Make sure the files are readable
 - `chmod g+r <file>`

Final exam

- Final exam Wednesday May 9 at 12:25 PM
- Engineering Hall 2309
- No printed materials allowed
- Similar length as midterm exam

Final exam

- Similar style as midterm, but may need to execute some algorithms by hand
 - Specifically those algorithms not covered on homework
 - Also need to understand optimization, objectives, etc.
- Almost all emphasis is on material not covered on midterm
 - 6 themes: Noncoding variants to RNA structure
 - May need prior material for comparisons, context
- 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

Large-Scale and Whole-Genome Sequence Alignment

- topics: large-scale alignment, whole-genome alignment, suffix trees, k-mer tries, longest increasing subsequence problem, MUMmer
- required reading
 - A. Delcher, S. Kasif, R. Fleischmann, J. Peterson, O. White and S. Salzberg. [Alignment of Whole Genomes](#). *Nucleic Acids Research* 27(11):2369-2376, 1999.

- Optional reading is also helpful but not tested directly

Final exam

- Some topics are on the slides but will not be tested:
 - Phosphoproteomics
 - Classes of pathway prediction algorithms
 - Pair hidden Markov models
 - Multiple whole genome alignment
 - MLAGAN
 - Mercator
 - Details of the RNA structure energy minimization algorithm
 - CYK and inside-outside algorithms

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?
 - What does a method optimize?
 - What assumptions does a method make and what input does it require?

Final exam

- Examples

- What is the difference in the objective function for the Nussinov and energy-minimization dynamic programming algorithms? Why is one preferable?
- What is the difference between two Markov model or hidden Markov model variants?
- Given these types of data, how do we use algorithms and ideas from lecture to model them?

Final exam

- Class evaluation incentive
 - Currently 9 / 20 completed evaluations
- Will post specific sub-problem topics at two completion thresholds
 - 2 problems for 15 / 20 completion
 - 2 problems for 20 / 20 completion
- Example from midterm problem 3:
 - TF binding site prediction
 - Epigenomic data for TF binding prediction
 - Gaussian process smoothing