

The Final Project

The capstone to BIO/CSC 295 is a collaborative, interdisciplinary *Final Project*. Your goals are to

- Form a mixed-background team that includes some students with biology knowledge and some students with programming/math knowledge. Ideally, each team would have four students.
- Pick an interesting (preferably unsolved) problem in bioinformatics.
- Gather data (typically from the Internet).
- Write a program that helps you analyze the data.
- Analyze the results.

Ideally, the project will require significant contributions from both biology experts and computer science experts.

Deadlines:

Thursday, 10 November 2011: *Preliminary Project Proposal*

This project proposal, which should be one or two pages long, gives an overview of your research question. It should describe the kind of data you expect to use. (Preferably, you would identify the data set precisely, but we will understand if you cannot identify a data set by this time.) It should describe how you plan to analyze the data. And it should explain what you expect to learn by analyzing the data in that way.

- * Bring six printed copies of your proposal to class.
- * We will also ask you to give an “elevator presentation” of your proposal in class.
- * We will assign each student a project proposal to review.
- * Sam and Vida will also review each proposal.

Tuesday, 15 November 2011: *Proposal Reviews*

Short responses to the the proposals from peers and profs.

Thursday, 17 November 2011: *More Short Presentations*

We will have an alumna doing research in bioinformatics visit class and provide feedback on your approaches.

Tuesday, 22 November 2011: *Revised Proposals*

A chance for you to show us that you have thought about the comments on your proposal. By this time, you should have clearly identified your data sets and resources from the literature.

Tuesday, 22 November 2011; Tuesday, 29 November 2011; Thursday, 1 December 2011: *Work, Work, and More Work*

Class time to work on your projects.

Thursday, 1 December 2011: *Even More Short Presentations*

Slightly extended elevator presentations for our second guest speaker. At this point, you may be able to include some preliminary results in your presentation.

Tuesday, 6 December 2011: *“Long” Presentations*

In-class presentations, including process and results. Six groups: Ten minute presentation, 5 minute Q&A. Not everyone in the group needs to present.

Thursday, 8 December 2011: *Wrapup*

Reports due. Standard scientific paper format: Introduction, including a literature review; Methods; Results; Discussion. Reports should include supplementary material: Code, Data, and Instructions.

* Reflections due. The group should prepare a *joint reflection* on the overall project, including what went well, what went poorly, and what you would do differently if you were to do it again. Each individual should prepare a personal reflection on his or her contributions to the project. Individuals may reflect on other issues, too.

* In-class debriefing.

Sample Projects

Finding olfactory receptor genes. In order to write a program to find such receptors, one would first determine typical patterns of DNA that make it likely that the code for a receptor is nearby (biology expertise). One would then write a program that detects such patterns, handling possible variations and approximation (computer science expertise). Next, one would run the program to ensure that it identifies known olfactory receptor genes and to see if it also identifies new areas of the genome that may code for olfactory receptors (mixed expertise).

Trans-membrane sequences. In addition to identifying alpha helices and beta sheets, can we identify other interesting structural characteristics? In particular, can we determine which part of a protein is likely to be a trans-membrane sequence? Biology expertise lets us identify and analyze data sets (and develops the question). CS expertise lets us adapt Chou-Fasman to this problem.

More generally, you may find it useful to try to adapt one of the techniques you've learned this semester to a new domain or data set. The last time we taught the course, one particularly ambitious group applied the Kellis technique to micro RNA sequences. You might find other applications for the Kellis approach, for Chou-Fasman-style techniques, for the Potti approach, and more.

Copyright © 2009-2011 Vida Praitis and Samuel A. Rebelsky. This work is licensed under a Creative Commons Attribution-NonCommercial 2.5 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/2.5/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.