

Spring 2014 – Lane Center for Computational Biology  
Research Projects for Masters and Undergraduate Students

**Class Year Desired** - MS

**Project Description**- The project relates to performing cancer detection from histopathology images. Specific tasks consist of image segmentation, and classification of cell phenotypes. A comparison of multiple approaches is one possibility for a project, as are others.

**Skills** - Proficiency with matlab. A machine learning course (either MS or Ph.D. level) is a requirement.

**Hours/week expected to work** - 12 hours/week

**Posted by** - Gustavo Rohde - gustavor@cmu.edu

**Class Year Desired** - MS/Undergrad

**Project Description** - CellOrganizer is an open source system (<http://CellOrganizer.org>) for using machine learning/computer vision methods to learn generative models of cell morphology and structure directly from microscope images. Students will have the opportunity to work on expanding/extending the system to provide new capabilities, such as models for new organelle types, models of cellular dynamics, or new visualization methods.

**Skills** - Working knowledge of matlab or python. Good math and statistics skills.

**Hours/week expected to work** - 12 hours/week

**Posted by** - Robert F. Murphy - murphy@cmu.edu

**Class Year Desired** - MS/Undergrad

**Project Description** - We have developed systems for automatically analyzing protein subcellular patterns in images of human tissue with the goal of identifying proteins that change their locations during the development of specific cancers. Current work is on extending and testing the system in collaboration with local pathologists.

**Skills** - Working knowledge of matlab or python. Good math and statistics skills.

**Hours/week expected to work** - 12 hours/week

**Posted by** - Robert F. Murphy - murphy@cmu.edu

**Class Year Desired** - MS/Undergrad

**Project Description** - Machine learning to improve healthcare. The Clinical Genomic Expert project (developed here at CMU) is a software system that will make machine-learning-based predictions of disease risk, personalized drug efficacy, and treatment

outcomes for individual patients based on their genomes and clinical history and communicate these predictions to clinicians treating the patient. We are seeking one or more people to help continue to develop this software for personalized medicine. Work would consist of finding efficient, simple, and generalizable ways to implement and extend the existing architecture to new types of genomic and clinical data and to incorporate additional machine learning approaches.

**Skills** - C++ proficiency is required.

**Hours/week expected to work** - Approximately 6-10 hours/week

**Posted by** - Carl Kingsford - carlk@cs.cmu.edu

**Class Year Desired** - MS/Undergrad

**Project Description** - We are looking for a student to work on simulation of macromolecular assembly processes, especially virus assembly. The project will potentially involve improving and porting algorithms for fast simulation of complex reaction systems, developing tools to facilitate design of rule sets describing how different systems behave, and applying these to learn assembly models from various kinds of experimental data on complex assembly.

**Skills** - Different aspects of this work require different skills, but in general experience with Java and C++ is needed. Knowledge of chemical kinetics, physical chemistry, and probability are all pluses.

**Hours/week expected to work** - 10-12 hours/week

**Posted by** - Russell Schwartz - russells@andrew.cmu.edu

**Class Year Desired** - MS/Undergrad

**Project Description** - We are looking for a student to work on studies of cancer evolution. The project will involve analyzing various kinds of data on cancer genomics and expression to identify mechanisms by which tumors evolve and build evolutionary tree models of tumors. One aspect of the work is concerned with developing methods for analyzing cancer genomic sequence data. Another concerns learning evolutionary models of tumor development from experimental data.

**Skills** - Experience with programming and algorithms/data structures is needed and experience with phylogenetics would be a plus for either direction, as well as potentially some basic machine learning.

**Hours/week expected to work** - 10-12 hours/week

**Posted by** - Russell Schwartz - russells@andrew.cmu.edu

**Class Year Desired** - MS/Undergrad

**Project Description** - I would like to have a student to help identify clinical markers which can best distinguish different groups of asthma patients. An undergrad or MS student can compare different strategies to find out the best markers, e.g., whether it is

the best to find markers which can distinguish one-vs-all other groups, or markers which can distinguish multiple groups simultaneously. An MS student can further explore different feature selection algorithms to identify such markers.

**Skills** - Students are ideally compute science major, or they aim for that if they haven't declared their major yet. I would expect them to be proficient with at least one programming language, e.g., Java, C, or C++, and have taken some statistics/math courses. R will be used for coding in the projects.

**Hours/week expected to work** - Grad students will expect to work 9-12 hours/week, and undergrads 5-10 hours/week.

**Posted by** - Wei Wu - [weiwu2@cs.cmu.edu](mailto:weiwu2@cs.cmu.edu)

**Class Year Desired** - MS

**Project Description** - I would like to have a student to compare different clustering algorithms to find out which algorithm is the best for grouping asthma patients who are usually characterized by mixed data types.

**Skills** - Students are ideally compute science major. I would expect them to be proficient with at least one programming language, e.g., Java, C, or C++, and have taken some statistics/math courses. R will be used for coding in the projects.

**Hours/week expected to work** - 9-12 hours/week

**Posted by** - Wei Wu - [weiwu2@cs.cmu.edu](mailto:weiwu2@cs.cmu.edu)

**Class Year Desired** - MS

**Project Description** - The origin of open reading frames is a central question in evolutionary biology. However, recent analysis of *S. cerevisiae* ribosome occupancy data suggests that ORFs can be generated de novo (Carvunis et al., Nature 2012). The McManus lab is looking for a student to examine de novo gene birth in several species of yeast by analysis of ribosome profiling data.

**Skills** - Good candidates will have taken 03-727 and have proficiency in a scripting language (python or perl). Knowledge of R is also desirable.

**Hours/week expected to work** - 9-12 hours/week

**Posted by** - Joel McManus - [mcmanus@andrew.cmu.edu](mailto:mcmanus@andrew.cmu.edu)

**Class Year Desired** - MS/Undergrad

**Project Description** - Development of course materials for an online course in Machine Learning/Data Sciences/Statistics.

**Skills** - Some background in one of these topics is preferred, but students will be given the time learn the material as they go.

**Hours/week expected to work** - 12

**Posted by** - Chris Langmead - [cjl@cs.cmu.edu](mailto:cjl@cs.cmu.edu)

**Class Year Desired** - MS/Undergrad

**Project Description** - Web and Mobile (iPhone and Android) development.

Programmers needed to implement several websites.

**Skills** - Some programming experience is necessary, but students will be given the time to learn the development technologies as they go.

**Hours/week expected to work** - 12

**Posted by** - Chris Langmead - [cjl@cs.cmu.edu](mailto:cjl@cs.cmu.edu)

**Class Year Desired** - MS/Undergrad

**Project Description** - Cloud Computing. Programmers needed to implement back-end services (databases etc).

**Skills** - Some programming experience is necessary, but students will be given the time to learn the technologies as they go.

**Hours/week expected to work** - 12

**Posted by** - Chris Langmead - [cjl@cs.cmu.edu](mailto:cjl@cs.cmu.edu)