# CSCI 2820 (formerly 2950L) Medical Bioinformatics: Genome-wide Association Studies

# **Tentative Syllabus**

Sorin Istrail (Sorin\_Istrail at brown dot edu) Phone: (401) 863-6196 Office Hours by Appointment

Graduate TA Derek Aguiar (rap at cs dot brown dot edu) Office Hours by Appointment

Time and Place: Tuesdays and Thursdays 2:30pm-3:50pm CIT 241 SWIG Boardroom Class Page: <u>http://www.cs.brown.edu/courses/csci2820/index.html</u>

**Course Description**: This course is devoted to computational problems and methods in the emerging field of Medical Bioinformatics where genomics, computational biology and bioinformatics impact medical research. There is no prerequisite for this course and individual accommodations will be made for students of different backgrounds; we will tailor assignments specifically for Life Sciences students (Biology, Chemistry, Medical) or Computational students (Applied Math, Computer Science, Engineering).

The focus will be on three main areas: Genome-Wide Disease Association Studies, Protein Folding, and Immunogenomics.

## 1. Genome-wide Association Studies (GWAS) (85% of the class)

- 1. SNPs and Haplotypes
  - Population genetics: models, linkage disequilibrium, identity by descent (IBD), pedigrees, trios
  - Tagging SNPs: The minimum informative subset of SNPs problem
  - Haplotype Phasing: Short-range phasing (Clark-consistency graphs, Clark meth- ods, maximum-likelihood, parsimony, PHASE) and long-range phasing (deCODE method)
  - The Coalescent: the Miniciello-Durbin ancestral recombination graph reconstruction
- 2. GWAS
  - Tests of associations, hypothesis testing, and multiple comparisons corrections
  - Population substructure
  - Disease models: common-disease common-variant, Zollner-Pritchard, McClellan-King genetic heterogeneity in human disease
  - The missing heritability problem: Common variants vs. rare variants
  - GWAS and Next Generation Sequencing
  - Genome-wide graph theory algorithms
  - Polya urn game
  - Ewens sampling lemma
  - GWAS case studies: Autism, Multiple Sclerosis, Type 2 Diabetes, Schizophrenia

## 2. Protein Folding and Drug Design (10% of the class)

- Protein folding: lattice models, folding, misfolding, and disorder
- Chemical graph theory: the medicinal chemist compound tinkering problem, drug- likeness and the Lipinski's rule of five

## 3. Immunogenomics (5% of the class)

- The Immunopeptidome; Are pathogens evolving their proteins to avoid the human immune system?
- Viral genomics: codon-bias

## Structure of the Course

## Homework

Homeworks will be assigned every other week. Towards the second half of the courses, homeworks will be assigned less frequently but will also be more involved. Homework problems will consist of a mix of general problems, programming assignments, problems related to the class project, and critical readings of research articles. Homeworks must be turned in on time and late submissions may be subject to penalties.

Programming may be done in Matlab, C/C++, Java, Mathematica, Python, or R.

## Projects

The list of suggested projects will become available on the projects page.

There will be two presentations for the class projects, one during the middle of the term and one at the end of the term.

## Grading

- Projects 50%
- Homeworks 35%
- Presentation 15%

Extra credit will be given for original contributions to research projects.

#### **Course Resources**

#### Web Site

Nearly everything you will need will be made available through the course web site, including TA notes, slides, homework assignments, tests, etc. Please check the web site regularly. The web site is located at: <a href="http://www.cs.brown.edu/courses/csci2820/index.html">http://www.cs.brown.edu/courses/csci2820/index.html</a>

#### Books

• (recommended – not required) Principles of Population Genetics (Fourth Edition, 2007) Daniel L. Hartl and Andrew G. Clark

#### Prerequisites

The course is designed for graduate students and upper-level undergraduates. It is also open to Computer Science and Math students, as well as biological and medical students. Since the class will be comprised of students with a diverse background, homework and tests will involve general questions for all students as well as more in-depth questions, which you will be able to choose from in accordance with your particular background. While there are no formal prerequisites for the courses, you should have a strong background in at least one of these two areas. Please contact the professor if you are unclear as to whether you have the necessary prerequisites for the course.

#### **Collaboration Policy**

You may discuss the homework problems with other students or use other resources such as textbooks or the Internet. However, you must not obtain answers directly from anyone else. All homeworks will be submitted individually.