

# BME 3810/6160 & CSE 3810/6800: Computational Genomics Spring 2015

**Lecture:** TuTh 2-3:15pm, Bronwell 124

**Instructor:** Ion Mandoiu

Office: ITE 261

Phone: 486-3784

E-mail: [ion@engr.uconn.edu](mailto:ion@engr.uconn.edu)

Office hours: TuWTh 12:30-1:30pm or by appointment

**Course outline:** Started in 1995 by the completion of the first genome sequence of a free-living organism, *H. influenzae*, the genomic era has led to thousands of complete genome sequences deposited in public databases and many more genome projects at various stages of completion. The large-scale availability of genome data is revolutionizing biological and medical research, with data-driven computational approaches taking a central role. This course covers fundamental computational methods for genomic data analysis, with a main emphasis on statistical methods and current applications in genomics and genetic epidemiology.

**Tentative list of topics to be covered:** Basic probability theory and statistics; statistical modeling of biological sequences; EM and Gibbs sampling algorithms for DNA motif discovery; Markov chains; profile HMMs for representing sequence families; models of DNA and protein evolution; likelihood methods in phylogenetics, bootstrapping; basic principles of population genetics; genotype phasing and haplotype frequency estimation; computation of Mendelian likelihoods, Elston-Stewart and Lander-Green algorithms; admixture mapping; association studies, next-generation sequencing data analysis. The list of topics may change according to progress and students' interest.

**Textbooks:** There is no required textbook for this course. Most of the covered material appears in the following books:

- R. Durbin, S. Eddy, A. Krogh, G. Mitchison, *Biological sequence analysis: probabilistic models of protein and nucleic acids*, Cambridge University Press, 1998.
- R.C. Deonier, S. Tavare, M.S. Waterman, *Computational genome analysis: an introduction*, Springer Verlag, 2005.
- K. Lange, *Mathematical and Statistical Methods for Genetic Analysis*, 2<sup>nd</sup> ed., Springer Verlag, 2002.
- R. Schwartz, *Biological Modeling and Simulation*, MIT Press, 2008.

**Prerequisites:** Undergraduate-level courses in biology, programming, and statistics. However, required background in these areas will be provided as needed via lectures and supplementary readings.

**Grade breakdown:** Grading will be based on:

- Theoretical homework assignments (30%),
- Programming assignments (30%), and
- Final project (40%)

*Homework assignments* will consist of theoretical problems reinforcing the material covered in lectures and must be submitted in electronic format via HuskyCT.

*Programming assignments* must be submitted electronically via the Rosalind site at (you must first enroll for the class using <http://rosalind.info/classes/enroll/0202538a60/>). Solutions can be prepared using any high-level programming language. You will be asked to process a dataset

generated by Rosalind on your own computer and then upload or copy-paste the solution to Rosalind along with your source code. Each submitted solution is automatically checked for correctness, allowing you to fix potential problems before the due date.

*The final project* will give you the opportunity to study a computational genomics problem in more depth. You are encouraged to devise your own final project topic; suitable topics include surveys of computational genomics topics not covered in the lectures, design and implementation of novel algorithms, and empirical evaluation of existing methods. Project requirements will include submitting two intermediate progress reports and a written final report of 15-20 pages. You will also be required to give a short presentation on your project at the end of the semester. When appropriate, projects can be done in teams of 2-3 students.

**HuskyCT & Piazza:** We have a HuskyCT site for the class; you can access it by logging in with your NetID and password at <https://learn.uconn.edu>. Please check this site regularly for lecture materials, assignments and solutions, grades, etc. For electronic class discussions we will be using Piazza. Rather than emailing questions to the instructors you are encouraged to post them on Piazza at <http://piazza.com/uconn/spring2015/bme38106160cse38106800/home>. The system is highly catered to getting you help fast and efficiently from both the instructor and classmates.

**Academic honesty:** You are expected to adhere to the highest standards of academic honesty. All submitted solutions must be your own work. For homework assignments and programming projects you may discuss ideas and concepts with others, but *must not share written solutions or code*. Use of published materials (including web resources) is allowed, but *all sources should be explicitly acknowledged in your solutions*. Violations will be reviewed and sanctioned according to the University Policy on Academic Integrity.

**Students with disabilities:** If you have a documented disability for which you are or may be requesting an accommodation, you are encouraged to contact the instructor and the Center for Students with Disabilities or the University Program for College Students with Learning Disabilities as soon as possible to better ensure that such accommodations are implemented in a timely fashion.