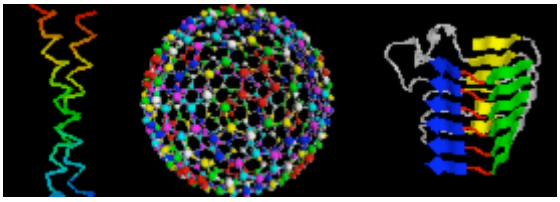
MIT
Department of Mathematics
& The Theory of
Computation Group
At CSAIL



Bioinformatics Seminar

Speaker: Voichita D. Marinescu, Ph.D., Children's Hospital Boston, Harvard Medical School.
Title: A hidden Markov model-based method for transcription factor binding site prediction
Date: Monday, 21 November 2005
Time & Location:
Refreshments: 11 am in the Theory of Computation Lab at MIT's Building 32, Stata Center Room G-575
Talk: 11:30 am the Theory of Computation Lab at MIT's Building 32, Stata Center, Room G-575
URL: http://www-math.mit.edu/compbiosem/

Abstract:
Computational prediction of transcription factor binding sites (TFBSs) in DNA sequences is challenging for several reasons. First, the models that abstract TFBS characteristics have to be trained on short nucleotide sequences usually between 6 to 20 bp in length that were determined experimentally and, therefore, are available in a small number. Secondly, searching long DNA sequences for matches to these models produces a large number of false positives that are difficult to separate from the true positives based on computational criteria or to evaluate experimentally in a large-scale manner. In this talk we present a method for TFBS prediction based on hidden Markov models that used the HMMER software package to build a large library of models for hundreds of transcription factors starting from sequences of experimentally determined sites curated in the TRANSFAC and JASPAR databases. This method served as the basis for developing MAPPER, a modular platform for TFBS identification and analysis, publicly available at http://mapper.chip.org/, that currently includes a database of putative TFBSs found in all upstream sequences of the human, mouse and *Drosophila* genomes, as well as a search engine to scan any DNA sequence or gene from the yeast, worm, fly, mouse or human genomes. We will present the results of an extensive evaluation performed on a collection of experimentally determined binding sites as well as on synthetic data that allowed us to asses the ability of our method to identify true positives, to estimate the proportion of false positives returned and to compare its sensitivity and specificity with other similar tools available. We will present several biological applications of this method and we will discuss further improvements of the HMMER modeling procedure designed to increase the sensitivity and specificity of the resulting TFBS models.

The seminar is co-hosted by Professor Peter Clote of Boston College's Biology and Computer Science Departments and MIT Professor of Applied Math Bonnie Berger. Professor Berger is also affiliated with CSAIL & HST.

Massachusetts Institute
of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

For General Questions, please contact kvdickey@mit.edu