Topology
Colloquium

A shape-based method for determining protein binding sites in a genome

Valerie Hower

University of California, Berkeley

Abstract: We present a new algorithm for the identification of bound regions from ChIP-Seq experiments. ChIP-Seq is a relatively new assay for measuring the interactions of proteins with DNA. The binding sites for a given protein in a genome are "peaks" in the data, which is given by an integer-valued height function defined on the genome. Our method for identifying statistically significant peaks is inspired by the notion of persistence in topological data analysis and provides a non-parametric approach that is robust to noise in experiments. Specifically, our method reduces the peak calling problem to the study of tree-based statistics derived from the data. The software T-PIC (Tree shape Peak Identification for ChIP-Seq) is available at http://math.berkeley.edu/ vhower/tpic.html and provides a fast and accurate solution for ChIP-Seq peak finding.

Monday, January 31, 2011, 4:30 pm
Mathematics and Science Center: W201

Mathematics and Computer Science
Emory University