Algorithms for Finding Genes and Simulating RNA Splicing

Christopher Burge, Department of Biology, MIT

Abstract:

Sequencing of eukaryotic genomes presents significant challenges and opportunities for computational biology. An important challenge is to develop methods for identifying all of the genes present in a genomic sequence, using either sequence similarity to known genes or proteins, 'ab initio' prediction, or other approaches. On the other hand, aligning available cDNA sequence data to the newly sequenced genome provides large datasets of primary transcript sequences with known exon-intron structure, which provides a wealth of data about introns and splicing. This seminar will cover:

- a new algorithm for identifying genes in genomic sequences, GenomeScan, and its application to gene identification in the human genome; and

- computational approaches to understanding RNA splicing specificity - the 'rules' used by the nuclear pre-mRNA splicing machinery for defining exons and introns - in a variety of organisms with advanced genome sequencing efforts.