

Title: Characterization of Somatic Mutations in Cancer Genomes

Speaker: Ben Raphael

Affiliation: Brown University, Department of Computer Science & Center for Computational Molecular Biology

Abstract:

Cancer is a disease that is driven by somatic mutations that accumulate in the genome during an individual's lifetime. Recent advances in DNA sequencing technology are enabling genome-wide measurements of these mutations in numerous cancers. I will discuss algorithmic approaches for two problems that arise in cancer genome analysis. The first problem is the inference of somatic mutations from the short sequences produced by current DNA sequencing technologies. Somatic mutations in cancer occupy a continuum of scales ranging from single nucleotide mutations through structural rearrangements of large blocks of DNA sequence. I will describe an algorithm for classification and comparison of structural rearrangements using paired-read DNA sequencing data. The second problem is to distinguish functional mutations that drive cancer progression from neutral "passenger" mutations. Recent cancer sequencing studies have shown that somatic mutations are distributed over a large number of genes. This mutational heterogeneity is due in part to the fact that somatic mutations target cellular signaling and regulatory pathways, and that a mutation in dozens of possible genes might be sufficient to perturb a pathway. While some of these pathways are well characterized, many others are only approximately known. This approximate information is represented as an interaction network, a graph whose nodes are genes and whose edges represent biological interactions between genes. I will describe *HotNet*, an algorithm to identify subnetworks of an interaction network that are mutated in a significant number of cancer genomes. *HotNet* models mutations as heat sources and employs a diffusion process on the interaction network to find "hot subnetworks." We also derive a statistical test to rigorously assess whether the number of hot subnetworks is significant under a suitable null hypothesis. I will illustrate applications of these algorithms to data from The Cancer Genome Atlas, a project that is characterizing the genomes of thousands of samples from dozens of cancer types.