Speaker: Shay Zakov, Dept. of Computer Science and Engineering, UCSD

Title: Reconstructing Breakage Fusion Bridge architectures using noisy copy numbers

Abstract:

Determining genomic sequences still poses a major challenge, which is addressed by both technological
and computational approaches. The difficulty increases when sequenced regions are complex, i.e. variable regions whose architectures deviate significantly from that of previously sequenced reference regions, and which contain many repetitions of common substrings.

Breakage Fusion Bridge (BFB) is a spontaneous mechanism that produces such complex genomic regions. The BFB process operates on a specific chromosomal arm in cycles, where in each cycle some suffix of the arm is duplicated, inverted, and concatenated at the end of the arm. Lately, it was speculated that BFB can be a cause of certain kinds of cancers. While the process itself was observed during the late 1930's, little is known about the extent of BFB in tumor genome evolution. This can be attributed to natural BFB instances being rarely observed as they occur, and the lack of rigorous assays for identifying BFB-modified genomes after the process has ceased.

This talk will address the challenge of computationally detecting BFB, given modern available data such as Next Generation Sequencing (NGS) and Array Comparative Genomic Hybridization (array CGH) data. We will discuss some combinatorial properties of BFB, the manner these properties can be exploited for BFB identification, and show some analyses results over real biological data. Interestingly, some of the formulated computational problems can be solved in linear time, while others have sub-exponential and exponential time algorithms. If time permits, we will also discuss an Informed Search approach that significantly improves the computational efficiency in practice for otherwise inefficient algorithms.

Joint work with Marcus Kinsella and Vineet Bafna.