

You are cordially invited to a talk in the Edmond J. Safra Center for Bioinformatics Distinguished Speaker Series.

The speaker is **Prof. Pavel Pevzner**, Dept. of Computer Science and Engineering, University of California, San Diego.

**Title:** "SPAdes genome assembler and its applications to emerging NGS technologies"

**Time:** Thursday, **October 22** 2015, at **11:00** sharp (refreshments from 10:50)

**Place:** **Holtzblat Hall 007**, entrance floor, Exact Sciences Faculty

**Host:** Prof. Ron Shamir, [rshamir@tau.ac.il](mailto:rshamir@tau.ac.il), School of Computer Science

**Abstract:** The lion's share of bacteria in various environments cannot be cloned in the laboratory and thus cannot be sequenced using existing technologies. A goal of single-cell genomics (SCG) is to complement gene-centric metagenomic data with whole-genome assemblies of uncultivated organisms. Assembly of SCG data is challenging because of highly non-uniform read coverage and highly elevated levels of chimeric reads/read-pairs. We describe SPAdes, an assembler for both SCG and standard (multicell) assembly that incorporates a number of new algorithmic ideas. We demonstrate that recently developed single-cell assemblers not only enable single-cell sequencing, but also improve on conventional assemblers on their own turf. We further describe (i) TrueSPAdes that assembles accurate and long (10Kb) reads generated by the recently released Illumina TrueSeq technology, (ii) transSPAdes for transcriptome assembly, and (iii) dipSPAdes for assembling highly polymorphic diploid genomes. Finally, we show that the de Bruijn graph assembly approach is well suited to assembling long and highly inaccurate SMRT reads generated by Pacific Biosciences.