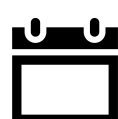# UniAligner: a new parameter-free framework for sequence comparison

Dr Pavel A. Pevzner

Department of Computer Science and Engineering,
University of California, San Diego, USA

Host: Niranjan Nagarajan

Wednesday 17 April 2024
3pm – 4pm

GIS Seminar Room (Level 2)
60 Biopolis Street, Genome, Singapore 138672

About The Speaker

Pavel Pevzner is Ronald R. Taylor Chair and Distinguished Professor of Computer Science and Engineering at University of California, San Diego. He holds Ph.D. from Moscow Institute of Physics and Technology, Russia. He was named Howard Hughes Medical Institute Professor in 2006. He was elected the Association for Computing Machinery (ACM) Fellow, the International Society for Computational Biology (ISCB) Fellow, the American Association for Advancement in Science (AAAI) Fellow, and the European Academy of Sciences member (Academia Europaea) member. He was awarded the ISCB Senior Scientist Award, and the ACM Kanellakis Theory and Practice Award. Dr. Pevzner authored textbooks "Computational Molecular Biology: An Algorithmic Approach", "Introduction to Bioinformatics Algorithms" (with Neal Jones), "Bioinformatics Algorithms: an Active Learning Approach" (with Phillip Compeau), and "Learning Algorithms through Programming and Puzzle Solving" (with Alexander Kulikov). He co-developed the "*Bioinformatics*" and "*Data Structure and Algorithms*" online specializations on Coursera with over a million enrollments in the last decade.

About The Seminar

The recent advances in "complete genomics" revealed many previously inaccessible genomic regions and enabled analysis of their associations with diseases. However, analysis of variations in centromeres, immunoglobulin loci, and other biomedically important highly-repetitive regions (HRRs) faces an algorithmic challenge since there are currently no tools for accurate sequence comparison of HRRs. Counterintuitively, the previously developed alignment approaches, such as the Smith-Waterman algorithm or minimap2, fail to construct biologically adequate alignments of HRRs. We present UniAligner, a fast parameter-free sequence alignment algorithm that prioritizes matches of rare substrings shared by two sequences. We applied UniAligner to human centromeres and primate immunoglobulin loci and quantified the extremely high rate of tandem duplications and deletions in centromeres thus demonstrating that centromeres represent the most rapidly evolving regions of the human genome with respect to their structural organization.