



Uri Keich

The University of Sydney

Estimating the statistical significance of sequence motifs

The identification of transcription factor binding sites is an important step in understanding the regulation of gene expression. To address this need, many motif-finding tools have been described that can find short sequence motifs given only an input set of sequences. Our talk concentrates on the statistical analysis of the significance of the motifs reported by those motif finders. Somewhat surprisingly, this statistical analysis has lagged considerably behind the extensive development of the finders themselves. Nevertheless, this analysis is often the only information available to biologists when deciding whether or not to invest the resources required to verify the predictions of those finders. While our statistical analysis is focused on the problem of motif finding, some of our techniques apply more generally. For example, we develop novel and more efficient algorithms to compute the exact significance of the multinomial goodness-of-fit and the Mann-Whitney tests. The talk should be accessible to people with no biological background.

About the speaker: Dr Uri Keich is Senior Lecturer at School of Mathematics and Statistics, The University of Sydney. He is interested in bioinformatics and in theoretical and computational statistics.

Time: 4pm, Friday, 28th August

Location: Room RC4082

Seminar co-ordinator: Spiridon Penev

e-mail: S.Penev@unsw.edu.au