

**Emi Tanaka**

National Institute for Applied Statistics Research Australia (NIASRA), UoW

### **Short introduction to motifs and improving the comparison of motifs**

I will give a short introduction to the biological importance of motifs followed by my research jointly worked with William Stafford Noble, Timothy Bailey, Charles Grant and Uri Keich.

A question that often comes up after applying a motif finder to a set of co-regulated DNA sequences is whether the reported putative motif is similar to any known motif. While several tools have been designed for this task, Habib et al. (2008) pointed out that scores that are commonly used for measuring similarity between motifs do not distinguish between a good alignment of two informative columns and one of two uninformative columns. This observation explains why tools such as TOMTOM occasionally return an alignment of uninformative columns which is clearly spurious. To address this problem, Habib et al. suggested a new score [Bayesian Likelihood 2-Component (BLiC)] which uses a Bayesian information criterion to penalize matches that are also similar to the background distribution. We show that the BLiC score exhibits other, highly undesirable properties, and we offer instead a general approach to adjust any motif similarity score so as to reduce the number of reported spurious alignments of uninformative columns.

#### **Bio:**

Emi did both her undergraduate and postgraduate in statistics at University of Sydney. Her PhD projects involved applying statistics in bioinformatics (particularly in motifs). Emi recently joined NIASRA at the University of Wollongong, as a research fellow under Prof. Brian Cullis.