Predicting the structure of RBP recognition elements

Francois Major[*][†][1]

[1]Univerisité de Montréal – Canada

Abstract

Improvements in genomics give us access to an unprecedented amount of RNA sequence data, coding and non-coding. It is a general desire to determine the structure and function of these RNAs in order to discover: binding sites to proteins, DNA, other RNAs, and small molecules; search in transcriptomic and genomic data for sequences with similar structure and function; and, design new sequences with predetermined properties. The success of accomplishing these tasks depends on the representation of an RNA sequence that we use. Here, I will introduce a new representation based on a feature vector that is appropriate to machine learning, which we are using to identify overrepresented structural motifs from RNA folding prediction data . Interestingly, the overrepresented motifs are determinants of RNA function. From this representation, we derive decision-tree based RNA family classifiers, and show that mutations that affect the outcome of these trees also affect function. Finally, we use it to model and search for the structure of RNA-binding protein recognition elements.

[*]Speaker
[†]Corresponding author: francois.major@umontreal.ca