An Integrated Approach to RNA Regulatory Motif Discovery

Daron Standley*[1], Taeho Kim[1], Takashi Mino[2], Alexis Vandenbon[1], Yutaka Suzuki[3], Osamu Takeuchi[2]

[1]Laboratory of Systems Immunology, WPI Immunology Frontier Research Center (FReC), Osaka University, Japan
[2]Laboratory of Host Defense, WPI Immunology Frontier Research Center (FReC), Osaka University, Japan
[3]Department of Medical Genome Sciences, The University of Tokyo, Japan

*Corresponding author (standley@ifrec.osaka-u.ac.jp)

We describe an alignment method for discovering regulatory motifs in mRNA 3' untranslated regions (UTRs), given a small number of known "seed" motifs. First the seed motifs are aligned to a set of candidate 3' UTRs using integrated sequence-structure alignment. In this stage the secondary (2D) structure of the seed and candidate UTRs are predicted and the base pairing propensity is used in the alignment similarity score. Local motifs in the candidate sequences with sequence and secondary structure similarity score above threshold values are retained. Next, all-against-all alignment of the high-scoring candidate motifs and hierarchical clustering are used to assess co-occurrence with the seeds. In this study, seed motifs consisted of 2 experimentally confirmed mRNA 3' UTR fragments known to be targeted by Zc3h12a, an RNase that is essential for regulating several inflammatory cytokines, including interleukin 6. We constructed a training set of 100 true and 100 false mRNA 3' UTR targets of Zc3h12a, based on RNA immunoprecipitation (RIP) sequencing data, and optimized the alignment and clustering parameters. We then constructed a test set of 13 true and 15 false targets of Zc3h12a, based on luciferase reporter assays with 3' UTRs of target mRNAs that were not in the seed or training sets. The overall accuracy, as defined by the area under the receiver-operating characteristic curve for co-occurrence of true test cases with the 2 seed clusters, was 76%. Seed-like 2D structures that were conserved in human/mouse were observed in 11/13 of the positive UTRs and 8/15 of the negative UTRs. The free energy of the positive stem-loops was much lower, on average, than that of the negative stem-loops (-4.0 and -1.4 kcal/mol, respectively). We conclude that the seed-like structure is likely to be necessary for Zc3h12a RNase activity.