

Evolutionary conservation of RNA structure

Elena Rivas

Department of Molecular and Cellular Biology, Harvard University

January 9, 2023

Abstract

Many functional RNAs depend on particular 3D structures. A long tradition of computational methods aims to infer RNA secondary structure (that is, the collection of base pairs) from sequence. However, essentially any sequence can be folded into a plausible structure. To distinguish RNA sequences that have biologically relevant structures from those that do not, additional evidence is needed. One powerful source of evidence is evolutionary conservation of RNA sequence and structure, which induces pairwise covariations that can be observed in RNA multiple sequence alignments.

Knowing when an RNA sequence includes a conserved RNA structure is not trivial and depends on clues left behind by conservation, covariation and variation.

I will present three recent advances: (1) a statistical covariation test to identify significant covariation over background covariation due to phylogeny; including a power of covariation calculation to identify negative pairs with power (variation) but insignificant covariation unlikely to form RNA base pairs; (2) a cascading folding algorithm that combines all positive and negative evolutionary information into complex structures including all types of pseudoknots and triplets. (3) An enhanced covariation statistical test at helix-level resolution that increases sensitivity in the detection of evolutionarily conserved RNA structure without sacrificing specificity.

The efficacy of these advances has been tested by predicting to great accuracy the structures of the human non-coding RNAs MALAT1 and telomerase RNA, and by inferring that the data currently available do not support a conserved structure for the non-coding RNAs HOTAIR and XIST.

I will present further directions to expand and apply these methods for the systematic identification of novel vertebrate structural RNAs relevant to human biology, and to create novel algorithms to incorporate the prediction of RNA motifs using deep learning methods.