The effect of SNPs on Boltzmann distribution for RNA secondary structure

Raheleh Salari, Teresa M Przytycka

{salarisa, przytycka}@ncbi.nlm.nih.gov

National Center for Biotechnology Information, Bethesda, MD 20894

Single-nucleotide polymorphisms (SNPs) in human genome can cause diseases, or change effectiveness of drugs and vaccines. Several studies report disease-associated SNPs, which are synonymous or map to the intergenic and non-coding regions of the genome, and propose different explanations for this association. One possible explanation is that these SNPs cause structural rearrangement in the RNA transcript that influence splicing, processing, or translational control and regulation.

To study the effect of SNPs on RNA structure, we designed an algorithm to compare the RNA secondary structure Boltzmann distribution for the SNP and wild-type RNA sequences. Our algorithm, which has $O(n^3)$ time and $O(n^2)$ space complexity, is an extension of McCaskill's algorithm [1] for computing the RNA secondary structure partition function. Given an RNA sequence and its SNP, our algorithm calculates the difference scores based on two measures "relative entropy" and "L1 distance" that quantify the distance between the RNA secondary structure Boltzmann distribution for the SNP and wild-type RNA sequences.

To evaluate the efficiency of our algorithm, we used a dataset of 514 disease-associated SNPs that map to the un-translated regions (UTRs) of human genome. The dataset has been combined by Halversen et al [2] from the Human Genetic Mutation Database [3]. We identified a set of SNPs that strongly alter the RNA structural ensemble, whereas these cases are not captured by other approaches like comparing the optimal structures or even base probabilities of the SNP and wild-type RNA. Our results demonstrate that this method can be used in similar studies to predict the effect of SNPs on RNA structural ensemble.

References

[1] **McCaskill JS (1990)** The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers; 29:1105.

[2] **Halvorsern M et al (2010)** Disease-associated mutations that alter RNA structural ensemble. PLOS Genetics; 6(8):e1001074.

[3] **Stenson PD et al (2009)** The Human Gene Mutation Database: 2008 update. Genome Med; 1(1):13.