Probing-directed structured elements detection in RNA sequences

S.V. Vinogradova

Moscow State University, Faculty of Bioengineering and Bioinformatics, 119991, Moscow, GSP-1, Leninskiye Gory, MSU, 1-73 kintany@gmail.com

R.A. Soldatov

Moscow State University, Faculty of Bioengineering and Bioinformatics, 119991, Moscow, GSP-1, Leninskiye Gory, MSU, 1-73 solrust@mail.ru

M.A. Mironov

Moscow State University, Faculty of Bioengineering and Bioinformatics, 119991, Moscow, GSP-1, Leninskiye Gory, MSU, 1-73 mironov@bioinf.fbb.msu.ru

Transcripts often harbor RNA elements, which regulate cell processes co- or post-transcriptionally. The functions of many regulatory RNA elements depend on their structure, so it is important to determine the structure as well as to scan genomes for structured elements. The best way to do this is to use comparative genomics approach and search for evolutionary conserved structures. But a suitable set of homologous sequences with moderate sequence divergence is too often not available due to the lack of related sequenced genomes, or rugged fitness landscape resulted in extremely high or low sequence conservation of structured RNAs. In these cases, we have to deal with single RNA sequences.

Functional RNAs are more stable than genomic background and we used this fact to develop the RNASurface algorithm that detects putative structured elements in RNA sequences [1]. The sizes of regulatory RNA elements vary from tens to hundreds of nucleotides and this results in limitation for computational approaches based on sliding window. RNASurface does not restrict the search to elements of fixed size but rather detects structured RNAs of optimal lengths.

Chemical probing of RNA is an alternative source of structural information: probing reactivities strongly correlate with local nucleotide flexibility. We incorporate probing data in MFE calculation using procedure that is called ‘soft constraint’ approach [2]. It is based on pseudo-energies that favour individual positions in RNA structure to be paired or unpaired. One important advantage of our approach is the ability to incorporate any type of

experimental data (SHAPE [3], PARS [4], DMS [5], *etc.*): we deal with probabilities of nucleotides of being paired/unpaired instead of relying on arbitrary normalized data from experiments.

Incorporation of RNA probing data into computation pipeline increases the signal/noise rate of structures prediction and detects more functional structures. However, our method is still dependent on the quality of probing data. Though at the moment high quality genome-wide RNA probing data for various organisms is not available, we believe that global interrogation of RNA structure will assist computational strategies to better model RNA structure, predict RNA function and screen genomes for functional RNAs.

1. Soldatov, R.A., Vinogradova, S.V., Mironov, A.A. (2014) *RNASurface: fast and accurate detection of locally optimal potentially structured RNA segments*. Bioinformatics 30:457–463. doi: 10.1093/bioinformatics/btt701

2. Washietl, S., Hofacker, I.L., Stadler. P.F., Kellis, M. (2012) *RNA folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction.* Nucleic Acids Research 40:4261–4272. doi: 10.1093/nar/gks009

3. Lucks, J.B., Mortimer, S.A., Trapnell, C., et al. (2011) *Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq)*. PNAS 108:11063–11068. doi: 10.1073/pnas.1106501108

4. Wan, Y., Qu, K., Zhang, Q.C., et al. (2014) *Landscape and variation of RNA secondary structure across the human transcriptome.* Nature 505:706–709. doi: 10.1038/nature12946

5. Rouskin, S., Zubradt, M., Washietl, S., et al. (2013) *Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo*. Nature 1–17. doi: 10.1038/nature12894