Searching for SNPs disrupting RNA secondary structures (Keynote)

Jan Gorodkin

Center for Non-Coding RNA in Technology and Health, University of Copenhagen, Danemark

Single Nucleotide Polymorphisms (SNPs) can have large impact on diseases as well as phenotypic traits. Traditionally, SNPs have been studied in protein coding sequence and lately also in regulatory elements such as transcription factor binding sites. Since phenotypic SNPs are widespread in the genome it is of equal interest to search for their impact everywhere including in RNA structure in transcriptomic sequence. Studying the potential impact of, for example, SNPs in coding sequence takes outset in non-synonymous changes and these have then further been used to study structure disruptions which then again are used to imply functional changes. In contrast, studying SNPs for structure disrupting potential in RNA is more complex, because longer range base pairings often are involved.

A number of strategies have been employed to address this, but they have mainly considered the RNA sequence globally, and thus local changes in large sequence can be harder to detect. We address this by constructing a computational approach, called `RNAsnp`, which considers the sequences locally from globally computed base pair probabilities in either the full sequence or in sliding windows. Our approach compares the wild-type and mutant sequences and search for the region which maximizes the difference in base pair probabilities using a given distance measure. Furthermore, we compute mutation effects by empirical p-values.

On the analysis of disease-associated SNPs in UnTranslated Regions (UTRs) we obtain substantially more candidates (20 vs. 3) than obtained by a global strategy on a set of 501 diseases associated SNPs. In a further study of cancer associated Single Nucleotide Variants (SNVs), we combined prediction of disrupted local RNA secondary structure and microRNA targets. We analyzed existing transcriptome data from patients with non-small cell lung cancer (NSCLC). In the original set, aimed at finding non-synomous SNVs, 40% of the in total (somatic and germ-line) 73,717 SNVs overlap UTRs. Of 29,290 SNVs in UTRs of 6,462 genes, we predict 962 (408 associated with local RNA structure; 490 to miRNA targets) disruptive SNVs in 803 different genes. Of these 188 (23.4%) were previously known to be cancer associated, which is significantly higher (p=0.032) than the ratio of 1,347 of 6,462 in the full data set. This analysis can furthermore be used for network analysis indicating where disruptive SNVs appear.

RNAsnp is available as standalone software and as webserver at:
`http://rth.dk/resources/rnasnp`

DOI: 10.15455/CMSR.2014.0008