

Analysis and identification of local structures in DNA

Author: Jan Havlík

Thesis supervisor: prof. RNDr. Ing. Jiří Šťastný, CSc.

Department of Informatics, Faculty of Business and Economics, Mendel University in Brno



Motivation

R-loops are common nucleic acid structures formed by a three-strand nucleic acid composed of an RNA-DNA hybrid and a displaced single-stranded DNA (ssDNA) loop. They are commonly occurring in the Human genome, but their structure causes genomic instability and rearrangements, which is a direct link to some human diseases. Therefore their discovery and prediction is vital.

There is an existing model^[1] using *in vitro* data^[2] for R-loop creation loci prediction, which is this tool based on.

^[1]<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4489302/>

^[2]<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2682002/>

R-loop-R

R-loop-R is a web tool for the detection and analysis of R-loops in given genomic sequence. Each analysis provides information about the creation site, length and statistical information about each R-loop.

R-loop-R was intended to be more scalable and extensible for the needs of Czech Academy of Sciences than already existing tool QmRLFS finder^[3]. R-loop-R provides multiple output formats and is capable of processing ultra long genomic sequences (whole chromosomes).

There are two RIZ (R-loop initiation zone) detection models to choose from:

- 3G cluster (at least 3 clusters of Guanines containing at least 3 Guanines in a row)
- 4G cluster (at least 2 clusters of Guanines containing at least 4 Guanines in a row)

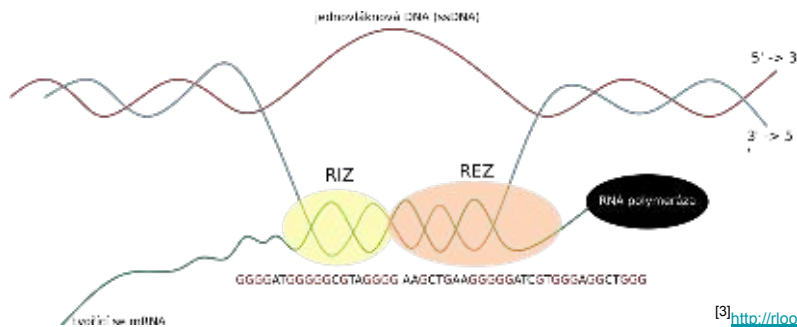


Fig 1:
Creation of a R-loop in RNA-DNA hybrid.
R-loop is defined by RIZ (R-loop initiation zone) and REZ (R-loop elongation zone)

^[3]<http://rloop.bii.a-star.edu.sg/?q=qmrlfs-finder>

Results

R-loop-R is now used for the research of R-loops and is still getting other extensions. The web application provides results of the analysis in CSV and bedGraph format, which can be used for direct integration into the [Genome browser](#). There is also a graphical output of the analysis directly in the web application. Part of the application is also an API interface.

We successfully analysed some of Human chromosomes, which are over 150 millions of bp long.

