

# Customizable detection of ncRNA genes

Marcus Lechner and Manja Marz

*Department of Pharmaceutical Chemistry, Philipps-University Marburg*

lechner@staff.uni-marburg.de

The detection of ncRNAs is an important task in genome annotation projects. In contrast to prediction of protein-coding regions, sequence based approaches alone cannot be used for a bioinformatic search of ncRNAs. Additional features such as secondary structure, synteny, promoter data, interaction sites, conserved motifs and specific distances are crucial prediction steps. Each RNA family has its own characteristic features which are more or less conserved among different species. Therefore, a number of tools (e.g. Blast[1], RNABOB[2], Infernal[3], GotohScan[4], Fragrep[5], RNASHAPES[6], RNADUPLEX[7], RNAfold[7], RNASUBOPT[7]) need to be combined in order to find and evaluate putative ncRNA genes. This combinatorial task can be time consuming and challenging, especially for scientists with limited background in computer science. Furthermore, there is only a small fraction of programs that come with a web interface to ease the application.

Flexible and effective solutions to customize the prediction of ncRNA genes and speed up the bioinformatic description of RNA families are required. To facilitate this task, we implemented a web based tool which allows to phrase descriptions using various programs, whereby advanced users still have the opportunity to use the full range of functions. The results are combined and assessed automatically. If favored, scoring parameters can be added to ease subsequent evaluation steps.

Descriptions sets can be exported and imported and thus, reused and published without much effort. A command line version for local use is available as well. Both versions provide the user with a comprehensive output about the undergone procedure as well as the necessary citations in order to offer comprehensible analysis.

## References

- [1] S F Altschul, W Gish, W Miller, E W Myers, and D J Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–10, Oct 1990.
- [2] SR Eddy. RNABOB: a program to search for RNA secondary structure motifs in sequence databases. 1992-1996.

- [3] E P Nawrocki, D L Kolbe, and S R Eddy. Infernal 1.0: inference of RNA alignments. *Bioinformatics*, 25(10):1335–7, May 2009.
- [4] J Hertel, D de Jong, M Marz, D Rose, H Tafer, A Tanzer, B Schierwater, and P F Stadler. Non-coding RNA annotation of the genome of trichoplax adhaerens. *Nucleic Acids Res*, 37(5):1602–15, Apr 2009.
- [5] A Mosig, K Sameith, and P Stadler. Fragrep: an efficient search tool for fragmented patterns in genomic sequences. *Genomics Proteomics Bioinformatics*, 4(1):56–60, Feb 2006.
- [6] P Steffen, B Voss, M Rehmsmeier, J Reeder, and R Giegerich. RNASHapes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics*, 22(4):500–3, Feb 2006.
- [7] S Wuchty, W Fontana, I L Hofacker, and P Schuster. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, 49(2):145–65, Feb 1999.