

# TmiRUSite and TmiROSite scripts: searching for mRNA fragments with miRNA binding sites with encoded amino acid residues

Olga Berillo<sup>1\*</sup>, Mireille Régnier<sup>2</sup> & Anatoly Ivashchenko<sup>1</sup>

<sup>1</sup>National nanotechnology laboratory, al-Farabi Kazakh National University, Almaty, Kazakhstan; <sup>2</sup>INRIA/LIX team AMIB 1 rue Honoré d'Estienne d'Orves 91120 Palaiseau, France; Olga Berillo – Email: devolia18@mail.ru; \*Corresponding author

Received July 03, 2014; Accepted July 11, 2014; Published July 22, 2014

## Abstract:

microRNAs are small RNA molecules that inhibit the translation of target genes. microRNA binding sites are located in the untranslated regions as well as in the coding domains. We describe TmiRUSite and TmiROSite scripts developed using python as tools for the extraction of nucleotide sequences for miRNA binding sites with their encoded amino acid residue sequences. The scripts allow for retrieving a set of additional sequences at left and at right from the binding site. The scripts presents all received data in table formats that are easy to analyse further. The predicted data finds utility in molecular and evolutionary biology studies. They find use in studying miRNA binding sites in animals and plants.

**Availability:** TmiRUSite and TmiROSite scripts are available for free from authors upon request and at <https://sites.google.com/site/malaheenee/downloads> for download.

**Keywords:** python script, evolution, mRNA, miRNA.

## Background:

microRNAs (miRNAs) are short protein non-coding RNAs [1]. They participate in translation repression [2]. Previously, miRNA binding sites were predicted only at the 3'UTR [3], but recently it is known that these sites are also located at the 5'-untranslated regions (5'UTRs) [4] and coding sequences (CDSs) of mRNAs [5]. Therefore, it is of interest to identify miRNA encoded oligo-peptides that are conserved across evolutionary distance. This was manually studied in previous report and was time consuming [6]. Hence, it is of interest to develop automated computer scripts for easy and robust analysis of such data. Thus, we describe TmiRUSite and TmiROSite scripts for the robust and fast extraction of nucleotide sequences for miRNA binding sites with their encoded amino acid residue sequences.

## Methodology:

### Software development

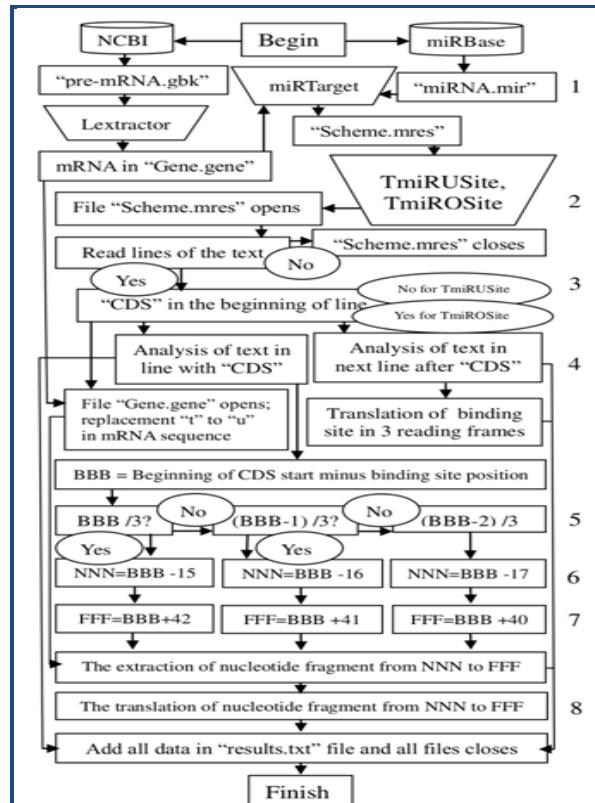
Scripts described in this article are programmed using Python and starts up through terminal. TmiRUSite and TmiROSite

scripts allow to search for binding sites in “scheme.mres” file of the miRTarget program. There are eight stages (**Figure 1**) that describe whole procedure of work with the scripts. mRNA sequence with binding site and its encoded for oligopeptide in open reading frame (ORF) can be defined by TmiROSite. The script treat the binding sites located in CDS and TmiRUSite script work with sites of untranslated regions. The oligopeptides contain five amino acids before binding site that is equivalent to 15 nucleotide (sixth stage of **Figure 1**) and some amino acids after them. It translate 42 nucleotides after start position of binding site. The quantity of additional nucleotides can be increased if necessary.

### Input

The complete nucleotide sequences of precursor mRNAs (pre-mRNAs) of human genes were downloaded from GenBank at NCBI [7] and the human mature miRNAs were extracted from the miRBase database [8]. The precursor were mRNAs shortened into the mature mRNA sequences by Lextractor\_002\_0020E3 script [9]. This script creates a file with

the mRNA sequence and indicates the boundaries between the 5'UTR, CDS, and 3'UTR according to the information available in GenBank. This file on the mRNA sequence was used together with a file of the miRNA sequences to identify miRNA binding sites using the miRTarget program [10]. These data are shown at first stage of (Figure 1).



**Figure 1:** The scheme of TmiROSite and TmiRUSite scripts. Note: NCBI and miRBase are databases; L extractor and miRTarget are programs; "pre-mRNA.gbk", "miRNA.mir", "Gene.gene", "Scheme.mres" and "result.txt" are created files. 1 - preparation of necessary files; 2 - choice of TmiRUSite or TmiROSite; 3 - selection of schemes with CDS for TmiROSite, but schemes with 5'UTR and 3'UTR for TmiRUSite; 4 - analysis the text after "CDS" for TmiROSite and "UTR" for TmiRUSite; 5 - definition of ORF (check of multiplicity to three); 6 - NNN is start nucleotide position of fragment with binding site; 7 - FFF is end nucleotide position of fragment; 4, 5 and 8 are specific stages for TmiROSite.

### Output features

The TmiROSite script (Translation of miRNA sites located in ORF) extracts oligopeptides encoded by nucleotide sequences of miRNA binding sites disposed in the CDS of mRNAs. It

should be noted that three possible oligo-peptides are encoded by the same binding site in three reading frames.

### Data Processing

This script processes two types of files: sequences of mRNAs in the file with the gene extension and schemes of binding sites received using the miRTarget program in files with the mres extension which contain the following data: the interaction schemes between miRNAs and mRNAs; the mRNA part where the site is located; beginning of a binding site; the ratio  $\Delta G$ ; relation  $\Delta G/\Delta G_m$ .

### Results:

The parameters used and the nucleotide sequence of binding sites are added to the result.txt file. The TmiRUSite script (Translation of miRNA sites located in untranslated regions) allows to find the oligonucleotide sequences containing binding sites with miRNA and above-mentioned characteristics of miRNA binding sites in table format (Figure 1).

### Caveat & future development:

It should be noted that additional nucleotide related information is required for the analysis of binding site conservation among evolutionary distance at yet another relatively nearby sites. It is also possible to study variability of sites for the same miRNA located in different parts of mRNAs. It is further possible to analyse conservation of amino acid residue sequence at the binding site localized in proteins. This process was traditionally done manually consuming huge processing time. The TmiROSite and TmiRUSite scripts described here are fast, simple and robust for such analysis. The accuracy of scripts is planned for improvement in future with the possible addition of binding site specific novel features.

### References:

- [1] Lee RC & Ambros V, *Science* 2001 **294**: 862 [PMID: 11679672]
- [2] Zeng Y *et al. Proc Natl Acad Sci U S A.* 2003 **100**: 9779 [PMID: 12902540]
- [3] Lai EC, *Nat Genet.* 2002 **30**: 363 [PMID: 11896390]
- [4] Da Sacco L & Masotti A, *Int J Mol Sci.* 2012 **14**: 480 [PMID: 23271365]
- [5] Brümmer A & Hausser J, *Bioessays* 2014 **36**: 617 [PMID: 24737341]
- [6] Ivashchenko AT *et al. Biomed Res Int.* 2013 **2013**: 902467 [PMID: 23957009]
- [7] <http://www.ncbi.nlm.nih.gov/pubmed/>
- [8] <http://mirbase.org>
- [9] <https://sites.google.com/site/malaheene/download>
- [10] Ivashchenko AT *et al. Bioinformatics* 2014 **10**: in press.

Edited by P Kanguane

Citation: Berillo *et al. Bioinformatics* 10(7): 472-473 (2014)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited