Circular RNA design report

Etienne Rolland
Master of Molecular and Cellular Biology
etienne.rolland.cdk@gmail.com

Abstract

Circular RNAs have been proved to be a large class of post-transcriptional regulators, decoying microRNAs from their original targets [1]. This documents is a output report of a script designing such a circular sponge against microRNAs passed as arguments, the goal being to use this circular sponge to upregulate genes originally targeted by those microRNAs. This script is the final step of a workflow relying on the TriplexRNA database [2] to identified those microRNAs to be targeted, but can be also used alone.

Chapter 1

Some details about the design

1.1 Distance of collaboration

There is experimental evidences of microRNA cooperativity in the post-transcriptional repression of the gene expression level when their seeds are in close proximity [13-35 nt] [3]. The design of the seed use by default the 17 nt distance, since the 13 nts distance, even if sufficient to enhance the repression, was less effective than the 17 nt distance. No significant statistical differences was found for outer distance compare to the 17 nt distance. However, this distance use by the script is adjustable in the eventuality of new scientific evidences, using the -d option.

1.2 U-rich design

The micro-RNA binding site with hight local content of A and U perform the best, over any other variable or predicator [4]. This is probably correlated to binding site accessibility, and to take advantage of this feature, and to reduce the potential to form a self-interacting secondary structure on the localisation of the microRNA binding site. A cognate problem is that some protein target and bind some AU rich sequence [known as AREs] [5], however, since this sequences require some A [5], the safer option is to specify a U in the sequence when no specific other nucleotides for complementarity is required.

1.3 Binding site

The design of the binding site rely on features which algorithms of microRNAs binding site prediction are looking for in sequences and highlight as biologically relevant. For example, Target Scan reward a A on the first position for scoring a potential microRNAs binding site [6].

The rules used by the circular RNA design script to design a binding site are the following :

- a A is systematically put in first position of the binding site [6].
- the seed (2-8nt) is a build by inserting complementary nucleotide to the microRNAs', forming an 8-mer site. See [4, 7] for further details and representation.
- a A is put in the ninth position [6].
- the 12th from 10th nucleotides are not complementary to the microRNAs' in order to avoid any Ago2 mediated cleavage of the circular RNA. This nucleotide are U, or A if not possible.
- the nucleotides from 13 to 16 (include, so four in total) are complementary to the - microRNAs'. The reason is that Watson-Crick pairing to four contiguous nucleotides produce pairing in 3' and that most downregulation is associated to this pairing if it start at the 13th position [7].

1.4 Clusters

The binding sites for one microRNA are organized in cluster by the design script, considering the observation that Ago2 shuttles between adjacent target and then that neighboring sites could cooperate to retain the Ago2-miRNA complex. [8].

The cluster are wrote from 3' to 5' to make comparaision with microRNAs sequence easier. The clusters are the following :

1.4.1 hsa-miR-25

Sequence microRNA : 5' cauugcacuugucucggucuga 3'

3' AUAACGUGAUUUGAGCUUUUUAAUAACGUGAUUUGAGCUUUUUAAUAACG

UGAUUUGAGCUUUUUAAUAACGUGAUUUGAGCUUUUUAAUAACGUGAUUUGA
GCUUUUUAAUAACGUGAUUUGAGCUUUUUAAUAACGUGAUUUGAGCUUUUUA
AUAACGUGAUUUGAGCUUUUUAAUAACGUGAUUUGAGCUUUUUA 5’

1.4.2 hsa-miR-21

Sequence microRNA : 5’ uagcuuaucagacugauguuga 3’

3’ AUCGAAUAAAUAGACUUUUUUAAUCGAAUAAAUAGACUUUUUUAAUCGAA
UAAAUAGACUUUUUUAAUCGAAUAAAUAGACUUUUUUA 5’

Chapter 2

Quality control

In this section are detailed some quality control. A alignement is perform againt all human microRNAs using miRanda [9] to check if whether or not the best alignement is performed by the microRNA passed as argument. If not, please reuse the design script using the -q option. This option allow the user to specify a binding site for the microRNA, and recreate a circular RNA and redo the quality check with this sequence. The fact that another microRNA can bind better the binding site is mostly cause of the G:U wobble between the microRNA and the circular RNA. In that case, the most relevant modification is to substituate this U by A [4].

2.1 Calculation of the minimum free energy of the circular RNAs

In this section is report the calculation of the minimum free energy of all the circular RNAs outputed by the design script

The calculation have been made using RNAfold of the package ViennaRNA 2.3.5.

Among all the circular RNAs present in the file named ”circular_design_output”, the one with the lower minimum free energy is >0

The minimum free energy for this design is -45.0kcal/mol

Which sequence is :

5’ AUUUUUCGAGUUUAGUGCAAUAAUUUUUCGAGUUUAGUGCAAUAAUUUUU
CGAGUUUAGUGCAAUAAUUUUUCGAGUUUAGUGCAAUAAUUUUUCGAGUUUA
GUGCAAUAAUUUUUCGAGUUUAGUGCAAUAAUUUUUCGAGUUUAGUGCAAUA
AUUUUUCGAGUUUAGUGCAAUAAUUUUUCGAGUUUAGUGCAAUAAUUUUUUC
AGAUAAAUAAGCUAAUUUUUUUCAGAUAAAUAAGCUAAUUUUUUUCAGAUAAAU
AAGCUAAUUUUUUCAGAUAAAUAAGCUA 3’

2.2 Alignements using Miranda

This section is the report of the best alignements among all the matures human micro-RNAs against the bindings site of micro-RNA on the sponge, and against the whole sequence of the circular RNA produced by the executable.

The purpose of this is to check if wether or not the best alignemenents are perform by the micro-RNAs given as arguments to the circular RNA design, and to ensure that a binding site for another micro-RNAs has not been create by mistake somewhere on the circular RNAs sequence during the design process.

The list of all the mature micro-RNAs have been download at mirBase.org on May 2017.

The alignements are done using miRanda, with the following options : ’-noenergy -strict -sc 150’, meaning :

-noenergy : Turn off thermodynamic calculations from RNAlib. If this is used, only the alignment score threshold will be used.

-strict : Require strict alignment in the seed region (offset positions 2-8). This option prevents

the detection of target sites which contain gaps or non-cannonical base pairing in this region.
-sc score : Set the alignment score threshold to score. Only alignments with scores >= score will be used for further analysis.

2.2.1 Mir_name : hsa-miR-25-3p, cluster for : hsa-miR-25

Score : of the alignement :161.000000

```
miRNA:     3′ agucuGGCUCUGUUCACGUUAc 5′
                    :|||| : ||||||||
CircRNA:   5′ auuuuUCGAGUUUAGUGCAAUa 3′
```

2.2.2 Mir_name : hsa-miR-944, cluster for : hsa-miR-25

Score : of the alignement :156.000000

```
miRNA:     3′ gaguaggcuaCAUGUUAUUAAa 5′
                         ||:||||||||
CircRNA:   5′ uucgaguuuaGUGCAAUAAUUu 3′
```

2.2.3 Mir_name : hsa-miR-92b-3p, cluster for : hsa-miR-25

Score : of the alignement :153.000000

```
miRNA:     3′ ccuccGGCCCUGCUCACGUUAu 5′
                    :|| | : ||||||||
CircRNA:   5′ auuuuUCGAGUUUAGUGCAAUa 3′
```

2.2.4 Mir_name : hsa-miR-590-5p, cluster for : hsa-miR-21

Score : of the alignement :157.000000

```
miRNA:     3′ gacgUGAAAA----UACUUAUUCGAg 5′
                   |:||||    || ||||||||
CircRNA:   5′ ----AUUUUUUCAGAUAAAUAAGCUa 3′
```

2.2.5 Mir_name : hsa-miR-21-5p, cluster for : hsa-miR-21

Score : of the alignement :152.000000

```
miRNA:     3′ aguuguAGUCAGACUAUUCGAu 5′
                     ||||  :  |||||||
CircRNA:   5′ auuuuuUCAGAUAAAUAAGCUa 3′
```

”

Bibliography

[1] S. Memczak, M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S. D. Mackowiak, L. H. Gregersen, M. Munschauer, A. Loewer, U. Ziebold, M. Landthaler, C. Kocks, F. le Noble, and N. Rajewsky. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 495(7441):333–338, Mar 2013.

[2] U. Schmitz, X. Lai, F. Winter, O. Wolkenhauer, J. Vera, and S. K. Gupta. Cooperative gene regulation by microRNA pairs and their identification using a computational workflow. *Nucleic Acids Res.*, 42(12):7539–7552, Jul 2014.

[3] P. Saetrom, B. S. Heale, O. Sn?ve, L. Aagaard, J. Alluin, and J. J. Rossi. Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Res.*, 35(7):2333–2342, 2007.

[4] David P Bartel. Micrornas: target recognition and regulatory functions. *cell*, 136(2):215–233, 2009.

[5] M. Spasic, C. C. Friedel, J. Schott, J. Kreth, K. Leppek, S. Hofmann, S. Ozgur, and G. Stoecklin. Genome-wide assessment of AU-rich elements by the AREScore algorithm. *PLoS Genet.*, 8(1):e1002433, Jan 2012.

[6] B. P. Lewis, C. B. Burge, and D. P. Bartel. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1):15–20, Jan 2005.

[7] A. Grimson, K. K. Farh, W. K. Johnston, P. Garrett-Engele, L. P. Lim, and D. P. Bartel. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, 27(1):91–105, Jul 2007.

[8] S. D. Chandradoss, N. T. Schirle, M. Szczepaniak, I. J. MacRae, and C. Joo. A Dynamic Search Process Underlies MicroRNA Targeting. *Cell*, 162(1):96–107, Jul 2015.

[9] A. J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander, and D. S. Marks. MicroRNA targets in Drosophila. *Genome Biol.*, 5(1):R1, 2003.