Relationship between divergence of using synonymous codons in host-virus and the presence of microRNA

Franco Riberi[1], Laura Tardivo[1], Lucia Fazzi[2], Guillermo Biset[3], Daniel Gutson[3], Daniel Rabinovich[2,3]

[1]Departament of Computing Sciences, Universidad Nacional de Río Cuarto, Río Cuarto, Córdoba, Argentina

[2]Instituto Biomédico en Retrovirus y SIDA-INBIRS, Buenos Aires, Argentina

[3]Fundación para el Desarrollo de la Programación en Ácidos Nucleicos-FuDePAN, Córdoba, Argentina

Background

MicroRNAs ($_{mi}$RNA) are small RNA that regulates the expression of $_{m}$RNA in the cells. They can interfere with viruses replication. In order to do this, it is necessary that the $_{mi}$RNA recognize genome target sites and that a pairing between the $_{mi}$RNA and a fragment of viral $_{mi}$RNA occurs. This recognition is more likely if the fragment is not paired (masked) in the secondary structure of viral $_{m}$RNA[1]. It is known that the genome of some human viruses has a bias in the use of synonymous codons (different codons that encode for the same aminoacid) compared with the host even though its replication would be less efficient[2].

Goal

The aim of the study is to determine if this bias could be the result of evolutionary pressure exerted by the $_{mi}$RNA. To achieve this goal massive comparisons should be made (in the order of $10e^7$) between the recognition of the virus natural genome and the "humanized" genome. The latter may be obtained by replacing codons in the viral genome, achieving a codon usage ratio similar to the host.

Materials and Methods

For each $_{mi}$RNAs, the software to be developed will do parallel "sweep" with the natural and humanized virus sequence. For each possible genome site, this program should determine the number of recognized nucleotides and whether these sites are available or masked by the secondary structure. When comparing results in homologous sites it can be determined whether $_{mi}$RNAs have a differential effect among different target $_{m}$RNAs (normal and humanized). The program will be coded using the C++ programming language and licensed under the GPLv3 software licence.

Results

For each $_{mi}$RNA and genome, a table should be produced that for each position records the matching $_{m}$RNA score, both in the original and the humanized sequence. Table 1 shows the shape structure to be generated.

	Original sequence			Humanized sequence			Score original sequence				Score humanized sequence			
Position	Matching*	Masked+	XYZ	Matching†	Masked‡	XYZ	%const=1*	cFold*	%const=1+	cFold+	%const=1†	cFold†	%const=1‡	cFold‡
1	aaTTg CacA	aaTTg Maca	AaTTg Xaca	ttAAC Gtct	ttMAC MtcM	ttYAC YtcX	0.44	0.45	0.22	0.24	0.55	0.54	0.11	0.21
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
N	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Table 1: Table structure to generate. Where (cFold constAT = 1.25) && (cFold constGC = 0.95).

Also, It will be analyzed whether the results favor the hypothesis of $_{mi}$RNA selective pressure as a cause of bias in codon usage.

Conclusions and Perspectives

The above presented shows that software can be developed as a tool for massive comparisons for the interations between $_{mi}$RNAs and alternative target $_{m}$RNA, this will be part of a software product called RNAemo. This will contribute to the development of tools to compare the possible effects of host $_{mi}$RNA in intentionally introduced viruses for gene therapy of cancers or genetic diseases. Further studies will include an estimate of the binding reaction between $_{mi}$RNA and $_{m}$RNA[3] free energy.

References

[1] Gareth M. Jenkins and Edward C. Holmes. **"The extent of codon usage bias in human RNA viruses and its evolutionary origin"**, 2003.

[2] Ulrike Muckstein, Hakim Tafer. **"Thermodynamics of RNA-RNA Interaction"**. Institute for Theoretical Chemistry, University of Vienna.

[3] Zuker, Michael. **"Computational Methods for RNA Secondary Structure".**. June 8, 2006.