doi 10.18699/bgrs2024-9.2-25

Searching for conservative targets for siRNA-therapeutics in human rotavirus type A genome

Makashov A.[1]*, Brodskaya A.[1, 2]

[1] *Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

[2] *Smorodintsev Research Institute of Influenza, St. Petersburg, Russia*

** makashov_aa@spbstu.ru*

Key words: rotavirus; siRNA; comparative virology

Motivation and Aim: Developing effective vaccines and treatment for viruses of various strains and subtypes relies on search of the conserved regions of viral genes and proteins. Our project investigated four rotavirus genes: VP4, VP7, NSP1, and NSP4 as potential targets for RNA interference as an antiviral therapy. Finding conserved gene regions for RNA interference could be a universal solution to counteract multiple virus serotypes. The first two genes encode the main rotavirus surface antigens, which play a key role in the interaction of the virus with the infected cell. The other two genes, NSP1 and NSP4, encode nonstructural rotavirus proteins involved in the interaction of virus and immune system, and are critical for self-assembly of new virions as well.

Methods and Algorithms: 9218 viral genomes presented in the NCBI GenBank database as of February 2024 were analyzed. Sequences of human rotavirus group A strains circulating from 1974 to 2022 were included in the analysis.

Multiple alignments of the obtained sequences were performed using the MAFFT tool [1]. We chose the G-INS-i algorithm as the most suitable for sequences with similar lengths and a high percentage of homology. Fasta file with multiple alignment results was converted into a nexus-file of the seqmagick program for further analysis. Bayesian reconstruction of the selected sequences was performed using the MrBayes program [2]. The number of generations was chosen based on a previously performed computational experiment, where after all generations, the standard deviation of cleavage probabilities was below 0.01 (value chosen based on published data). A cladogram with posterior probabilities for each cleavage and a phylogram with average branch lengths were generated and written into a nexus file. The trees were visualized and edited with the FigTree tool [3].

Results: Nonstructural protein genes NSP1 and NSP4 are characterized by greater homogeneity across all sequences. However, we can note the presence of indel polymorphism and small insertions in the reading frame of 6-9 nucleotides for several strains isolated in the United States and Australia in the last decade. Several extended indels have been shown for NSP4, predominantly in South Asia. Strains circulating in India in 2012–2013 are characterized by specific long stable insertions of up to 50 nucleotides, but are not evolutionarily conserved and are not found anywhere else.

Consensus sequences were constructed using the cons program from the EMBOSS package. Further analysis of short conserved regions satisfying multiple miRNA selection requirements was performed using this consensus sequence.

We also performed a comprehensive analysis searching for conserved regions of 20 to 25 nucleotides each with optimal GC composition and certain nucleotide patterns in order to satisfy the rules of selection of effective inhibitory miRNAs [3].

Conclusion: The trees were visualized and edited with the FigTree tool. Most circulating human RVAs are reported to belong to three evolutionary lineages that differ in genome constellations: Wa-like strains, DS-1-like strains and AU-1-like strains [4].
Two potential conserved regions in the NSP1 gene that could be targets for antiviral miRNA inhibition were selected: 260–300 and 675–750 (positions in the consensus nucleotide sequence obtained by alignment); and 3 potential conserved regions for NSP4:160-210, 360-395, 460-490. In the protein-coding region of the VP4 and VP7 genes, it is not possible to find a 20-nucleotide conserved region. The maximum length of the sequence with stable conservatism and optimal GC composition does not exceed 15 nucleotides. Nevertheless, these regions were also identified for further miRNA selection namely 333–352 and 600-650 for VP4 and 548–665 for VP7. Further in these regions were determined target sequences for RNAi. Experimental approbation of RNAi effectiveness is holding on now.
Funding: The study is supported by RSF Project (No. 22-74-10117).

References

1. Katoh K., Misawa K., Kuma K., Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, *Nucleic Acids Res.* 2002;30(14):3059-3066. doi 10.1093/nar/gkf436
2. Ronquist F., Teslenko M., van der Mark P. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539-542. doi 10.1093/sysbio/sys029
3. Rambaut A. FigTree v1.3.1. Institute of Evolutionary Biology. Edinburgh:University of Edinburgh, 2010
4. Matthijnssens J., Ciarlet M., Heiman E. et al. Full genome-based classification of rotaviruses reveals a common origin between human Wa-Like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. *J Virol*. 2008;82(7):3204-3219. doi 10.1128/JVI.02257-07
5. Laganà A., Veneziano D., Russo F. et al. Computational design of artificial RNA molecules for gene regulation. *Methods Mol Biol*. 2015;1269:393-412. doi 10.1007/978-1-4939-2291-8_25