Journal of Bioinformatics and Computational Biology

CONSERVED OLIGONUCLEOTIDES IN NONCODING REGIONS OF SARS-COV-2 VIRUS AND THEIR POTENTIAL ROLES IN THE VIRAL PATHOGENESIS

--Manuscript Draft--

Manuscript Number:	
Full Title:	CONSERVED OLIGONUCLEOTIDES IN NONCODING REGIONS OF SARS-COV-2 VIRUS AND THEIR POTENTIAL ROLES IN THE VIRAL PATHOGENESIS
Article Type:	Research Paper
Keywords:	Untranslated regions; viral miRNAs; RNA recapping; antagonistic mechanism; delayed immune response; SARS-CoV-2 infection
Corresponding Author:	Long Dang, Ph.D University of Da Nang: The University of Danang Da Nang, Da Nang VIET NAM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Da Nang: The University of Danang
Corresponding Author's Secondary Institution:	
First Author:	Long Dang, Ph.D
First Author Secondary Information:	
Order of Authors:	Long Dang, Ph.D
	Kieu D.M. Nguyen, Ph.D
	Nguyen Hoang Minh Khang
	Tran Anh Khoi
	Nguyen Thien Luong
Order of Authors Secondary Information:	
Abstract:	Here we identify the existence and possible roles of conserved motifs in untranslated regions (5' and 3') of SARS-CoV-2's genome. The first discovered motif (5' CGATCTCTTGT 3'), which was detected near the terminal of 5'UTR of the genome, has similar characteristics as 5' terminal oligopyrimidine (TOP) motif in eukaryotic mRNAs. This motif could play an essential role in recapping of the viral RNAs, hence enhancing their expression. The second motif (5' GGAAGAGC 3') detected in 3'UTR was shown to be likely the seed region of two virus-encoded miRNAs. Using bioinformatics predictions of the miRNAs' targets, we showed that the viral miRNAs could inhibit the cellular immune response and contribute to the phenomenon of delayed innate immune response to SARS-CoV-2 infection. These predictions of the study may offer new avenues in investigating the pathogenic mechanisms of the virus and in devising new treatments for the disease.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Table S4. Predicted viral pre-miRNAs of SARS-CoV2 and scores by Vmir

Viral pre-miRNAs	Orientation	Precursor sequence (Vmir)	Score
V-md1	direct	CTAGGGAGAGCTGCCTATATGGAAGAGCCCTAATG TGTAAAATTAATTTTAGTAGTGCTATCCCCATGTG ATTTTAATAGCTTCTTAG	71
V-md2	direct	CCTAATGTGTAAAATTAATTTTAGTAGTGCTATCCCCATG TGATTTTAATAGCTTCTTAGG	52.8
V-mr1	reverse	ACATTAGGGCTCTTCCATATAGGCAGCTCTCCCTAGCATT GTTCACTGTACACTCGATCGTACTCCGCGTGGCCTCGGTG AAAATGTGGTGGCTCTTTCAAGTCCTCCCTAATGT	63.5

(Note: Sequences in red and blue are octanucleotide motif and s2m motif respectively)

Viruses	Sequence of s2m motif
Conserved motif	CGNGG(N)CCACG NN GNGT(N)A NN A N CGAGGGT(N)ACAG
SARS-CoV-2	CGAGG CCACG CG GAGT A CG A T CGAGTGT ACAG
SARS-CoV	CGAGG CCACG CG GAGT A CG A T CGAGGGT ACAG
	1 2 3 4 5 6 7 8 9 10 1112 13141516 17 1819 20 21 22232425262728 29303132

Figure S5. s2m motif of SARS-CoV-2 and SARS-CoV
Nucleotide which is underlined is mutated from Guanine-26 into Thymine-26 in s2m motif of SARS-CoV-2 while Guanine-26 is conserved in s2m of SARS-CoV.

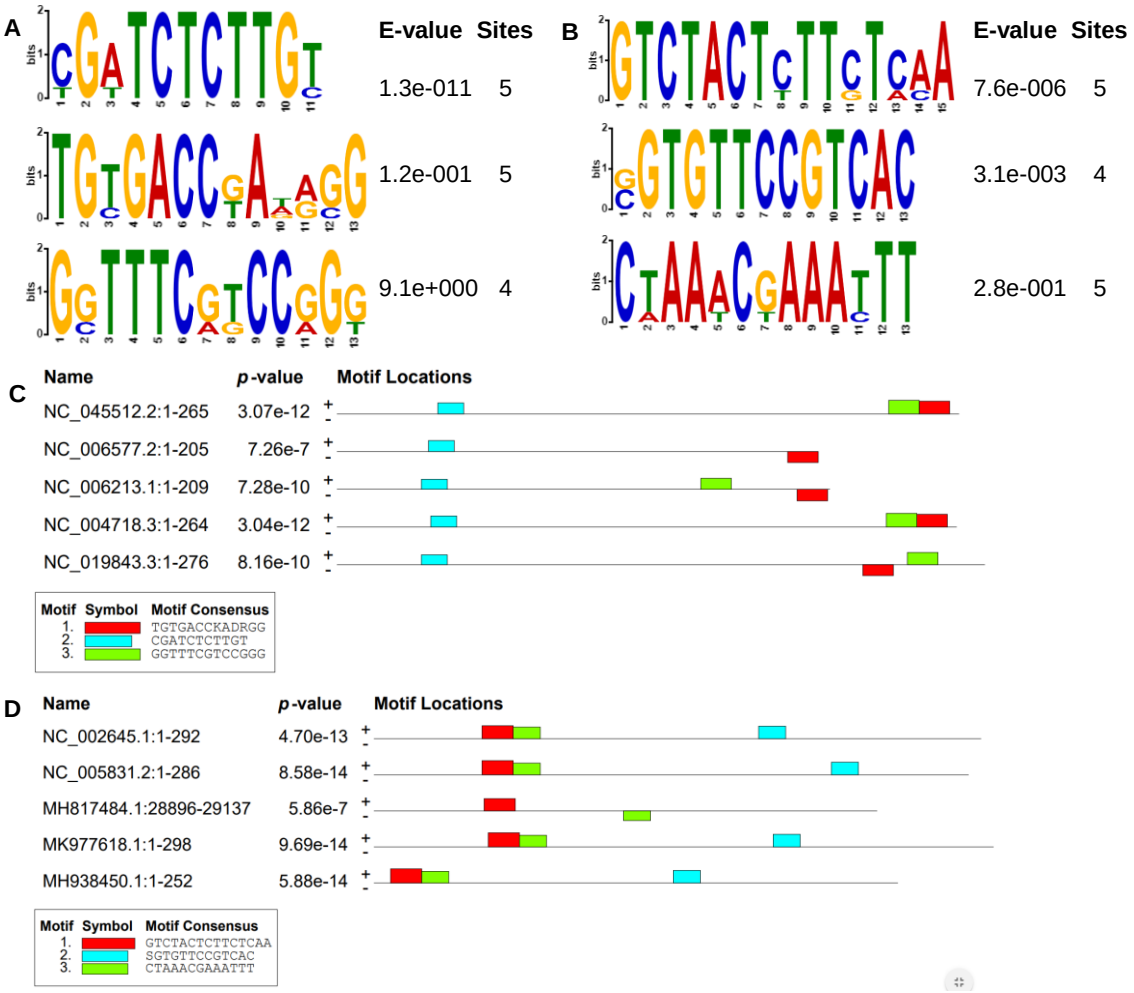
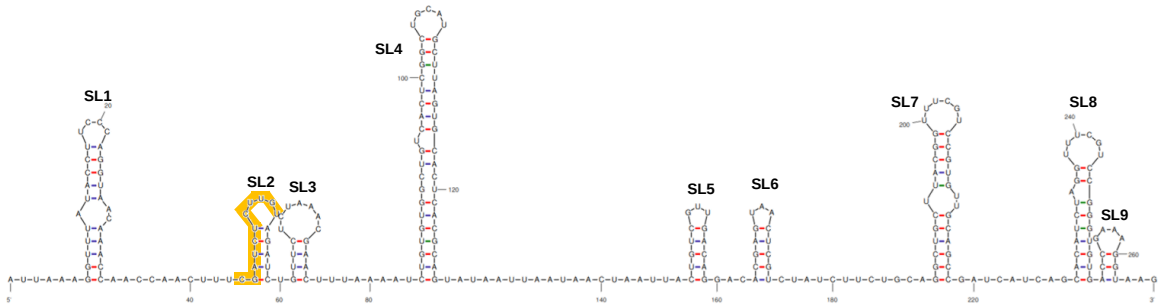
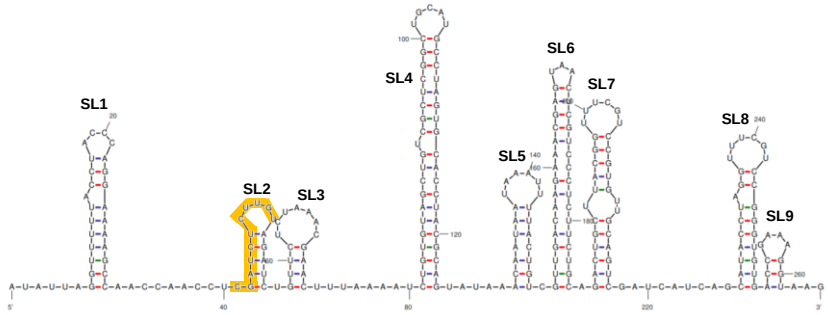


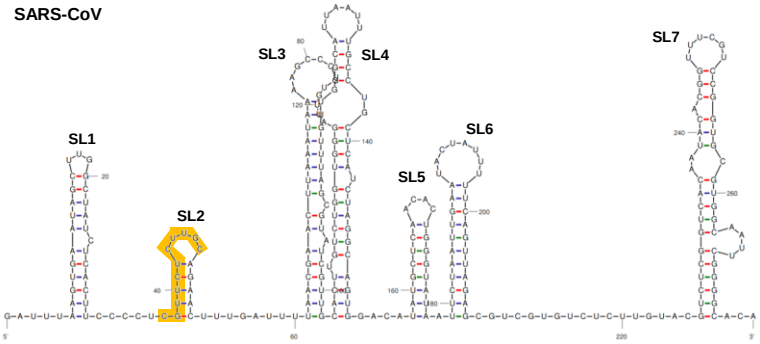
Figure S1. Motif discovery in 5'UTR of betacoronavirus (n=5) and alphacoronavirus (n=5) by MEME
(A) Motifs were discovered in 3'UTR of human-infecting betacoronaviruses. **(B)** Motifs were discovered in 5'UTR of two human-infecting alphacoronaviruses and three other alphacoronaviruses. **(C)** Locations of discovered motifs of betacoronavirus 5'UTRs **(D)** Locations of discovered motifs of alphacoronavirus 5'UTRs.
NC_006213.1 (HC0V-OC43); NC_006577.2 (HCoV-HKU1) NC_019843.3 (MERS-CoV); NC_004718.3 (SARS-CoV) NC_045512.2 (SARS-CoV-2), NC_002645.1 (HCoV-229E), NC_005831.2 (HCoV-NL63), MH817484.1 (FCoV-SB22), MK977618.1 (SeACoV-p10), MH938450.1 (Bat-CoV/P)



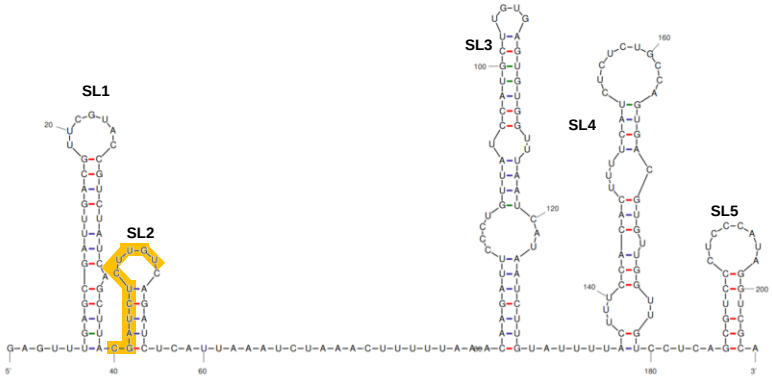
SARS-CoV-2



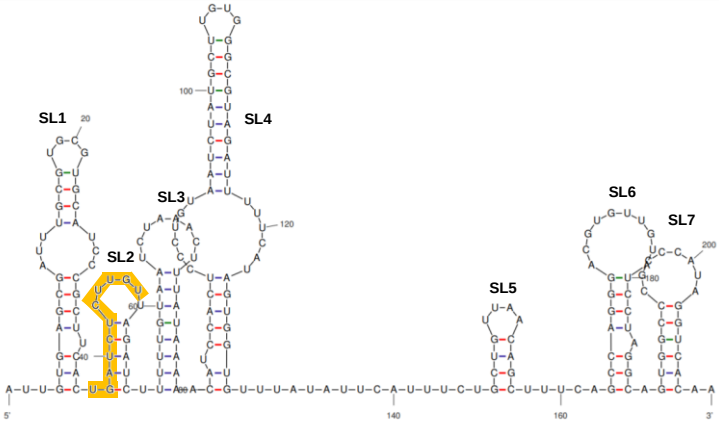
SARS-CoV



MERS-CoV



HCoV-HKU1



HCoV-OC43

Figure S2. Location of TOP-like motif on stem-loop structures of betacoronavirus' 5'UTR
TOP-like motif of five human-infecting betacoronaviruses, which is highlighted in yellow color, locates in SL2 in stem-loop structure of 5'UTR. SL is short for stem loop.

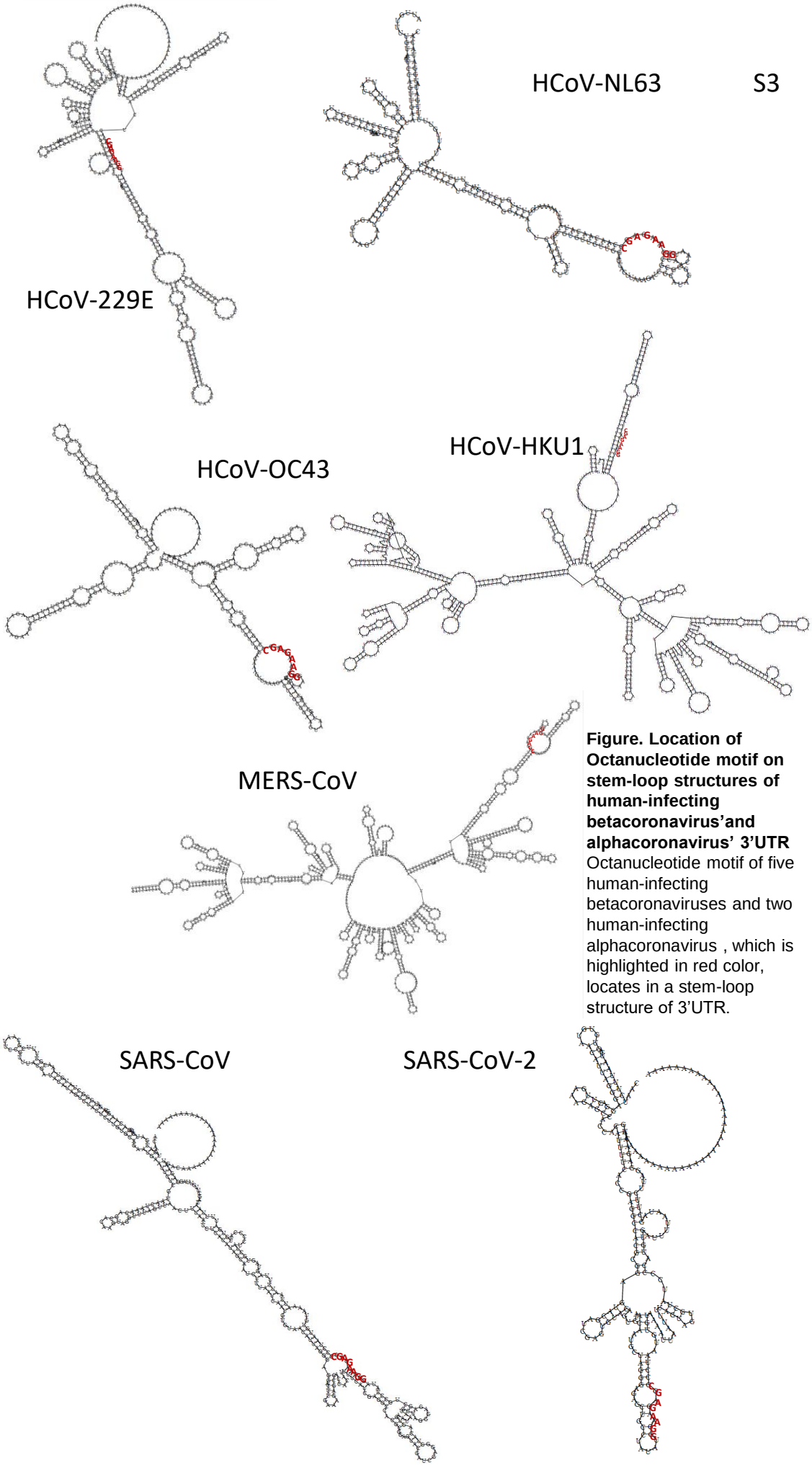


Figure. Location of Octanucleotide motif on stem-loop structures of human-infecting betacoronavirus'and alphacoronavirus' 3'UTR
Octanucleotide motif of five human-infecting betacoronaviruses and two human-infecting alphacoronavirus , which is highlighted in red color, locates in a stem-loop structure of 3'UTR.

A

```
v-mr1                 9  gcucuuccauauaggcag  26
                         ||||||||||||   |||
hsa-miR-1270         18  gcucuuccauaucuccag  1
                                                e-value= 8.7,
score= 63/100
-----------------------------------------------------------------------------------------

v-mr1                62  acuccgcguggccucggu  79
                         |||| ||||||| |||||
hsa-miR-1307-3p       1  acucggcguggcgucggu  18
                                                e-value= 1.5,
score= 72
                             |||| ||||||| |||||
cfa-miR-1307          1  acucggcguggcgucggu  18
                                                e-value= 1.5,
score= 72
                             |||| ||||||| |||||
ptr-miR-1307          1  acucggcguggcgucggu  18
                                                e-value= 1.5,
score= 72
                             |||| ||||||| |||||
bta-miR-1307          1  acucggcguggcgucggu  18
                                                e-value=
1.5, score= 72
                             |||| ||||||| |||||
ssc-miR-1307          1  acucggcguggcgucggu  18
                                                e-value=
1.5, score= 72
                             |||| ||||||| |||||
ppy-miR-1307          1  acucggcguggcgucggu  18
                                                e-value=
1.5, score= 72
                             |||| ||||||| |||||
eca-miR-1307          1  acucggcguggcgucggu  18
                                                e-value=
1.5, score= 72
                             |||| ||||||| |||||
chi-miR-1307-3p       1  acucggcguggcgucggu  18
                                                e-value= 1.5,
score= 72
                             |||| ||||||| |||||
pal-miR-1307-3p       1  acucggcguggcgucggu  18
                                                e-value= 1.5,
score= 72
                             |||| ||||||| |||||
dno-miR-1307-3p       1  acucggcguggcgucggu  18
                                                e-value=
1.5, score= 72
                             |||| ||||||| |||||
ocu-miR-1307-3p       1  acucggcguggcgucggu  18
                                                e-value= 1.5,
score= 72
                             |||| ||||||| |||||
dma-miR-1307          1  acucggcguggcgucggu  18
                                                e-value= 1.5,
score= 72
                             |||| ||||||| |||||
pha-miR-1307          1  acucggcguggcgucggu  18
                                                e-value=
1.5, score= 72
-----------------------------------------------------------------------------------------
v-mr1            79 ugaaaauguggguggcucuuu 98
                    | ||| ||||| ||| ||||
cja-miR-371a      3 ucaaacugugggggcacuuu 22
                                                score: 64,
evalue: 7.2

v-mr1            55 cgaucguacuccgcguggcc 74
                    |||| | || ||||||||||
eca-miR-9083     20 cgauagcacaccgcguggcc 1
                                                score: 73,
evalue: 1.3

v-mr1            79 ugaaaauguggguggcucuuu 98
                    | ||| ||||| ||| ||||
ssc-miR-371-5p    3 ucaaacugugggggcacuuu 22
                                                score: 64,
evalue: 7.2

v-mr1            29 cucccuagcauuguucacugu 49
                    || || || ||||||||||||
eca-miR-9021      4 cugccgagaguuguucacugu 24
                                                score: 69,
evalue: 2.7

v-mr1            85 uguggugcucuuucaa 101
                    | ||||| |||||||||
cja-miR-9978     17 uuuggugccucuuucaa 1
                                                score: 67,
evalue: 4.0
```

B

```
v-md1            11  cugccuauauggaagagc  28
                     |||   ||||||||||||
hsa-miR-1270      1  cuggagauauggaagagc  18
                                                e-value= 6.3, score =

v-md1            18  uauggaagagcccu  31
                     |||||| |||||||
hsa-miR-6769a-5p  8  uauggaggagcccu  21
                                                e-value= 9.1, score =

v-md1            61 uccccaugugauu 73
                    |||||||||||||
cja-miR-9961      1 uccccaugugauu 13
                                                e-value= 4.2, score = 65/100

v-md1            10 gcugccuauauggaag 25
                    ||||||| | ||||||
eca-miR-9007      2 gcugccuuucuggaag 17
                                                e-value= 7.5, score =  62/100
```

Figure S6. Alignment with mammalian mature miRNAs of SARS-CoV-2 viral precursor miRNAs (v-mr1, v-md1) by miRBase

(A) Alignment of v-mr1 with mature miRNA of mammalia. Sequence in red is predicted to be viral mature miRNA and octanucleotide sequence highlighted in yellow box. Human mature miRNA sequence is bold.

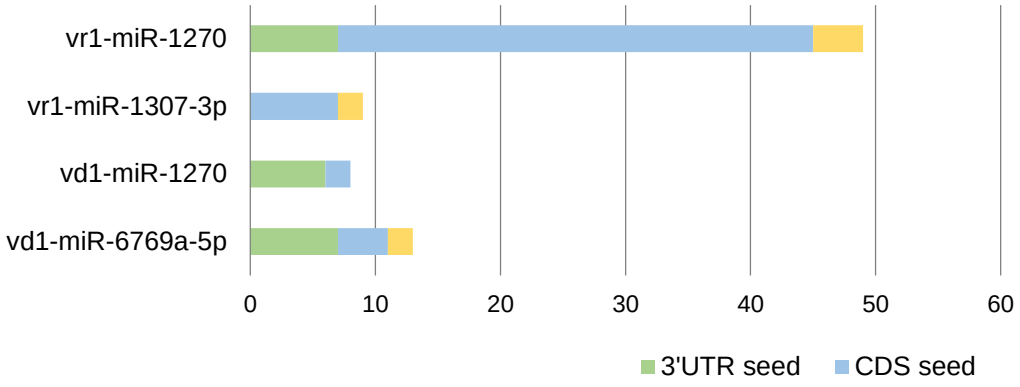
(B) Alignment of v-md1 with mature miRNA of mammalia. Sequence in red is predicted to be viral mature miRNA and octanucleotide sequence highlighted in yellow box. Human mature miRNA sequence is bold.

(Hsa: human, cfa: dog, ptr: chimp, bta: cow, ssc: pig, ppy: bornean orangutan, eca: horse, chi: domestic goat, pal: Collared flycatcher, dno: armadillo, ocu: rabbit, dma: aye-aye, pha: baboon)

Table S7. Predicted target sites of SARS-CoV-2-encoded miRNAs by Sfold

No.	Gene group	Gene	Viral miRNA	Seed type	Seed region	ΔG total	ΔG hybrid
1	Interferon	IFNB1	vd1-miR-6769a-5p	7mer-m8	5'UTR	-7.362	-18.700
			vr1-miR-1270	6mer	CDS	-8.397	-16.200
		IFNG	vd1-miR-1270	7mer-A1	3'UTR	-8.841	-17.600
			vd1-miR-1270	6mer	3'UTR	-8.163	-21.100
			vd1-miR-6769a-5p	Offset-6mer	3'UTR	-8.713	-17.400
			vr1-miR-1270	6mer	CDS	-6.433	-20.900
			vd1-miR-1270	6mer	CDS	-5.790	-20.300
			vr1-miR-1270	7mer-A1	5'UTR	-6.934	-15.100
2	Interferon type 1-inducing genes	JUN	vr1-miR-1307-3p	7mer-A1	CDS	-4.054	-23.800
		FOS	vr1-miR-1270	6mer	5'UTR	-5.740	-20.200
			vr1-miR-1270	7mer-A1	CDS	-4.808	-15.200
		STAT2	vd1-miR-6769a-5p	Offset-6mer	CDS	-9.249	-16.700
			vr1-miR-1270	Offset-6mer	CDS	-10.335	-18.500
			vr1-miR-1270	Offset-6mer	CDS	-1.186	-17.700
			vd1-miR-6769a-5p	7mer-m8	3'UTR	-12.381	-15.900
		NFKB1	vr1-miR-1307-3p	offset-6mer	5'UTR	-8.621	-31.700
			vd1-miR-6769a-5p	offset-6mer	CDS	-0.636	-16.800
			vd1-miR-6769a-5p	6mer	CDS	-1.743	-15.800
			vr1-miR-1270	offset-6mer	CDS	-4.196	-15.400
		IFR3	vr1-miR-1270	7mer-m8	CDS 520-526	-6.731	-25
			vr1-miR-1270	7mer-m8	CDS 890-896	-1.643	-25.500
		PCBP2	vd1-miR-6769a-5p	offset-6mer	3'UTR	-9.587	-16.300
			vr1-miR-1270	7mer-m8	CDS	-20.400	-12.501
			vd1-miR-1270	7mer-m8	CDS	-6.801	-20.700
		STAT1	vr1-miR-1270	offset-6mer	CDS	-10.358	-18.900
			vr1-miR-1307-3p	7mer-A1	CDS	-9.726	-20.700
			vr1-miR-1270	offset-6mer	CDS	-5.685	-20.700
		STAT5A	vr1-miR-1270	offset-6mer	3'UTR	-7.117	-19.600
			vr1-miR-1270	6mer	CDS	-4.076	-15.300
		STAT5B	vr1-miR-1270	offset-6mer	3'UTR	-9.861	-16.600
			vr1-miR-1270	6mer	3'UTR	-12.389	-19.200
3	Antigen presentation	HLA-B	vr1-miR-1270	6mer	CDS	-14.638	-25.200
		HLA-A	vr1-miR-1270	offset-6mer	3'UTR	-13.418	-15.900
			vd1-miR-1270	6mer	3'UTR	-5.831	-23.200
			vr1-miR-1307-3p	7mer-A1	CDS	-1.323	-26.200
			vr1-miR-1307-3p	6mer	CDS	-0.438	-26.300
		HLA-C	vr1-miR-1270	7mer-m8	CDS	-6.414	-23.600
			vr1-miR-1307-3p	7mer-A1	CDS	-1.545	-26.200
		TAP2	vr1- miR-1270	offset-6mer	3'UTR	-3.136	-18
			vd1-miR-6769a-5p	offset-6mer	3'UTR	-3.023	-21.800
			vr1-miR-1307-3p	7mer-A1	CDS	-4.800	-23.900
			vr1-miR-1270	7mer-m8	CDS	-2.685	-23.900
			vd1-miR-6769a-5p	offset-6mer	CDS	-9.752	-19.400
		TAP1	vr1-miR-1307-3p	8mer	5'UTR	-17.012	-29.500
			vd1-miR-1270	6mer	3'UTR	-6.275	-18.300
			vr1-miR-1270	offset-6mer	CDS	-1.934	-20.300
			vr1-miR-1270	offset-6mer	CDS	-1.936	-20.300
			vr1-miR-1270	7mer-m8	CDS	-9.455	-23.900
			vr1-miR-1307-3p	offset-6mer	CDS	-9.021	-27.500
4	DNA methylation	DNMT3B	vr1-miR-1270	offset-6mer	CDS 1614-1619	-9.148	-20
			vr1-miR-1270	offset-6mer	CDS 1614-1619	-8.048	-18.900
			vr1-miR-1270	offset-6mer	CDS 549-554	-1.956	-16.400
		DNMT3A	vr1-miR-1270	offset-6mer	CDS 232-237	-5.493	-21
			vd1-miR-6769a-5p	offset-6mer	CDS	-2.426	-21.700
			vr1-miR-1270	offset-6mer	CDS 1587-1592	-0.276	-16.400
			vd1-miR-6769a-5p	offset-6mer	5'UTR	-2.911	-20.200
		DNMT1	vr1-miR-1270	offset-6mer	CDS	-11.852	-16.800
			vr1-miR-1270	6mer	CDS	-10.169	-20.200
			vr1-miR-1270	offset-6mer	CDS	-8.571	-18.100
			vr1-miR-1270	offset-6mer	CDS	-5.936	-15.100
			vr1-miR-1270	7mer-A1	CDS	-5.620	-17.100
			vr1-miR-1270	offset-6mer	CDS	-5.358	-17.300
			vr1-miR-1270	6mer	CDS	-4.200	-17.300
		TDG	vr1-miR-1270	offset-6mer	3'UTR	-10.128	-17.100
			vr1-miR-1270	offset-6mer	CDS	-8.028	-19.300
			vr1-miR-1270	offset-6mer	CDS	-0.585	-15.600
			vr1-miR-1307-3p	7mer-A1	5'UTR	-5.285	-25.400
5	Translation	EIF5	vd1-miR-6769a-5p	offset-6mer	3'UTR	-11.832	-15.600
			vd1-miR-1270	offset-6mer	3'UTR	-8.905	-16.400
			vr1-miR-1270	offset-6mer	CDS	-16.794	-16.200
			vr1-miR-1270	6mer	CDS	-2.627	-15.200
		EIF4E	vd1-miR-1270	6mer	3'UTR	-11.950	-19.100
			vd1-miR-6769a-5p	7mer-m8	3'UTR	-5.007	-19.300
			vr1-miR-1270	Offset-6mer	CDS	-25.600	-7.890
			vr1-miR-1270	offset-6mer	5'UTR	-10.179	-16.500
			vr1-miR-1270	7mer-m8	5'UTR	-7.004	-21.300
		EIF2	vr1-miR-1307-3p	7mer-A1	CDS	-7.852	-24
		EIF6	vd1-miR-6769a-5p	6mer	3'UTR	-12.329	-16.100
6	ER stress response	BI-1	vd1-miR-6769a-5p	7mer-m8	3'UTR	-2.040	-18.300
			vr1-miR-1270	Offset-6mer	3'UTR	-1.836	-18.800
		XBP1	vr1-miR-1270	Offset-6mer	CDS	-1.795	-17.600

Figure S8. The number of predicted target sites of each SARS-CoV-2-encoded miRNAs at seed region of target transcripts



Manuscript Click here to access/download;Manuscript;MANUSCRIPT.docx

CONSERVED OLIGONUCLEOTIDES IN NONCODING REGIONS OF SARS-COV-2 VIRUS AND THEIR POTENTIAL ROLES IN THE VIRAL PATHOGENESIS

Kieu D.M. Nguyen[1], Nguyen Hoang Minh Khang[2], Tran Anh Khoi[3], Nguyen Thien Luong[4], Long D. Dang[5]*

[1] *Department of Infection Biology, Faculty of Medicine, University of Tsukuba, Japan; E-mail: s2130513@u.tsukuba.ac.jp*

[2] *The ABC International School, email: khangnguyenuniapp@gmail.com*

[3] *British International School Ho Chi Minh, email: khoitran130307@gmail.com*

[4] *University of Michigan, email: luongnt@umich.edu*

[5] *VN-UK Institute for Research and Executive Education, the University of Danang, Vietnam. E-mail: long.dang@vnuk.edu.vn*

*** Corresponding author:** Long D. Dang (long.dang@vnuk.edu.vn)

Abstract:

Here we identify the existence and possible roles of conserved motifs in untranslated regions (5' and 3') of SARS-CoV-2's genome. The first discovered motif (5' CGATCTCTTGT 3'), which was detected near the terminal of 5'UTR of the genome, has similar characteristics as 5' terminal oligopyrimidine (TOP) motif in eukaryotic mRNAs. This motif could play an essential role in recapping of the viral RNAs, hence enhancing their expression. The second motif (5' GGAAGAGC 3') detected in 3'UTR was shown to be likely the seed region of two virus-encoded miRNAs. Using bioinformatics predictions of the miRNAs' targets, we showed that the viral miRNAs could inhibit the cellular immune response and contribute to the phenomenon of delayed innate immune response to SARS-CoV-2 infection. These predictions of the study may offer new avenues in investigating the pathogenic mechanisms of the virus and in devising new treatments for the disease.

Keywords: *Untranslated regions; viral miRNAs; RNA recapping; antagonistic mechanism; delayed immune response; SARS-CoV-2 infection.*

I. Introduction

SARS-CoV-2 virus, a positive-stranded RNA virus belonging to genus Betacoronavirus (betaCoV) of the subfamily *Orthocoronavirinae* of the *Coronaviridae* family, is causing a pandemic (COVID-19) on unprecedented scale [1]. Researchers around the world have poured the resources to identify molecular features that make the outbreak of this virus so dangerous. For example, the observation that spike protein attached to angiotensin-converting enzyme 2 (ACE2) with high affinity helps to explain why SARS-CoV-2 transmits more efficiently and rapidly than SARS-CoV [2]. A high rate of asymptomatic transmission cases is another factor that makes the disease difficult to contain [3]. The reason of this characteristics has not been clarified at present [4].

Noncoding (*nc*) regions consisting of 5' and 3'UTR of SARS-CoV-2 genome are being received less attention than the coding regions. Stem-loop secondary structure motifs are highly conserved in 5'UTR and 3'UTR of coronavirus [5]. The important role of those strutural motifs for viral replication and transcription has been proved [6]. Secondary structural elements in the RNA genome of coronavirus have also been suggested to contribute to the viral pathogenesis [6]; however, their specific roles in the pathogenesis require further studies. The *nc* regions can also be the source of ncRNA molecules such as microRNAs (miRNAs), which have a strong hint by the presence of the stem-loop structures in coronaviral UTRs. Although ncRNAs are most prevalent in DNA viruses, they have been proved to exist in some retroviruses as well as negative- and positive-strand RNA viruses [7,8]. Several studies *in vivo* and *in silico* have discovered virus-derived miRNAs in both negative and positive strand RNA viruses and

their host targets [9,10,11]. These ncRNAs could contribute to the regulation of cellular gene expression and host response to infection [12]. However, there is a need to know more about sequential and structural features of 5' and 3'UTR of SARS-CoV-2 genome as well as their potential biological roles.

Here we investigated SARS-CoV-2's UTRs for conserved motifs *in silico* and we found two well-conserved motifs. One discovered motif locates in the 3'UTR with many stem-loop structures, leading to a hypothesis that SAR-CoV-2 virus could encode viral miRNAs that contribute to viral pathogenesis by inhibiting expression of viral infection-responsive genes. Our results suggest three most significant miRNAs in SARS-CoV-2's 3'UTR: vr1-miR-1270, vd1-miR-1270 and vd1-miR-6769a-5p. Their predicted targets are mRNA transcripts of genes that are important for modulating various cellular processes including interferon production, antigen representation, ER stress response, and translation initiation. This could explain how SARS-CoV-2 evades host immune response and other cellular immunity mechanisms. Furthermore, unexpectedly we found in the 5'UTR of SARS-CoV-2 and other human-infecting betacoronaviruses a well-conserved motif, which shares common characteristics with 5' terminal oligopyrimidine (TOP) of eukaryotic mRNAs. 5'TOP is indicated as a signal for recapping process which restores translatability of uncapped 5'TOP mRNAs[13]. Thus, we suggest a role of SARS-CoV-2's TOP-like motif as a recapping signal that helps viral RNAs to exploit host recapping pathway for the promotion of viral protein synthesis. The identified conserved motifs and their predicted roles provide new

perspectives for further experimental investigation to confirm the pathogenesis of SARS-CoV-2.

II. Data and methods

2.1. Motif discovery in untranslated region sequences of betacoronavirus and alphacoronavirus

The untranslated region sequences in FASTA format of five human-infecting betacoronavirus and two human-infecting alphacoronavirus were retrieved from Gene bank (NCBI) (Table 1). Multiple em for motif elicitation (MEME)[14], a webserver for local multiple alignment was used to discover ungapped motifs (recurring, fixed-length patterns) in 3'UTR and 5'UTR of beta- and alpha-coronaviruses. Setting options were *classic mode, Zero or One Occurrence Per Sequence, 1 motif, 0-order model of sequences, minimum width 4, and maximum width 14 for 5'UTR and minimum width 8, and maximum width 22 for 3'UTR*. Clustal Omega, a webserver for global multiple alignment was utilized to explore conserved sequences and location in coronaviral UTRs. Setting options were *DNA, ClustalW with character counts, dealign input sequence (NO), MBED-like clustering guide (YES), MBED-like clustering (YES), number of combine (default), and max guide tree (default), max HMM (default), and order (aligned).*

Table 1. List of coronavirus candidates and accession numbers

No.	Coronavirus candidates	Accession number

Betacoronavirus		
1	Human coronavirus HKU1 (HCoV-HUKU1)	NC_006577.2
2	Human coronavirus OC43 (HCoV-OC43)	NC_006213.1
3	Middle East respiratory syndrome coronavirus (MERS-CoV)	NC_019843.3
4	Severe acute respiratory syndrome coronavirus (SARS-CoV)	NC_004718.3
5	Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)	NC_045512.2
Alphacoronavirus		
1	Human coronavirus 229E (HCoV-229E)	NC_002645.1
2	Human Coronavirus NL63 (HCoV-NL63)	NC_005831.2

2.2. Prediction of SARS-Cov-2 viral microRNA candidate precursors and mature microRNAs

Viral miRNA precursors of SARS-CoV-2 were analyzed by using Vmir program[15] (VMir Analyzer v2.3 / VMir Viewer v1.6). The software program is designed specifically to identify pre-miRNAs in viral genomes. with the following setting options: confirmation *(linear), orientation (both), window site (500), step size (10), min. hairpin size (50), max. harpin size (any), and score (any).* Then, all sequences of predictable miRNA precursors were inputted into RNAfold webserver[16] to predict their secondary structures with setting options: *fold algorithms (Minimum free energy and*

partition function, avoid isolated base pairs), dangling end options (dangling energies on both sides of helix in any case), energy parameters (RNA parameters-Tuner model 2004), slope (1.9), intercept (-0.7), rescale energy parameters to given temperature (37). RNAfold web server was used to validate structures and predict minimum free energy (MFE) of miRNA precursors produced by Vmir software. Those having MFE equal or less than -15 kcal/mol were considered potential precursor candidates. Subsequentially, we aligned viral precursor sequences with mature miRNAs of human and mammalia in miRbase program[17] to identify whether viral miRNAs conserved in human and other mammalians and obtain possible mature sequences of SARS-CoV-2 viral microRNAs with setting options: *search algorithm (BLASTN), sequence database (mature), e-value cutoff (10), max alignments (100), word size (4), match score (+5), mismatch penalty (-4), species filter (human or mammalia).*

2.3. Prediction of potential human messenger RNA targets of viral microRNAs

Mature sequences of SARS-CoV-2 miRNAs are hybridized with mRNA transcripts of interferon type I, interferon type II, interferon-inducting factors, antigen representation genes, methylation/demethylation enzymes, and translation factors in order to indicate potential targets and target sites of SARS-CoV-2 miRNAs (Table 2). We focus on those genes because previous evidences suggested that SARS-CoV, coronaviruses, and other RNA viruses modulate the following cellular processes: interferon/cytokines production[10], antigen representation[18], endoplasmic reticulum stress induction[19], DNA methylation[18], and cellular translation[19]. These target candidates are supposed to be suppressed during host immune evasion and gene

expression modulation by SARS-CoV-2. CLIP-based prediction of microRNA binding sites tool (Sfold webserver)[20] was used for hybridization with the following setting options: *V-CLIP- based model (Human) and species for prediction (Human).* Target sites at 3'UTR/5'UTR/CDS seed with ΔG total < 0 and ΔG hybrid ≥ 15 were considered significant and reported. Most significant target sides are those at 3'UTR seed with ΔG total < 0 and ΔG hybrid ≥ 15.

Table 2. List of human gene candidates and accession numbers

No.	Gene candidates	Accession number
Interferon		
1	IFNA1	NM_024013.3
2	IFNB1	NM_002176.4
3	IFNG	NM_000619.3
Interferon/cytokine production induction		
1	STAT5B	NM_012448.4
2	STAT5A	NM_001288720.1
3	STAT1	KR709942.1
4	STAT2	NM_005419.4
5	AP-1 (FOS)	NM_005252.4
6	AP-1 (JUN)	NM_002228.4
7	IRF3	NM_001197123.2

8	NFKB1	NM_003998.4
Antigen representation		
9	HLA-A	NM_002116.8
10	HLA-B	NM_005514.8
11	HLA-C	NM_001243042.1
12	TAP1	NM_000593.6
13	TAP2	NM_001290043.2
Endoplasmic reticulum stress induction		
14	BI-1	NM_003217.3
15	XBP1	NM_001079539.1
16	ECD	NM_007265.3
DNA methylation/demethylation		
17	DNMT1	NM_001130823.3
18	DNMT3A	NM_153759.3
19	DNMT3B	NM_006892.4
20	TDG	NM_003211.6
Translation initiation		
21	EIF1	NM_005801.4
22	EIF5	NM_001969.5
23	EIF6	NM_002212.4
24	EIF3A	NM_003750.4

25	EIF4G3	NM_001198801.2
26	EIF4E	NM_001968.4
27	EIF2	NM_129557.4

III. Results

3.1. Identification of a novel conserved oligopyrimidine motif in 5'UTR of SARS-CoV-2

The result showed a highly-conserved motif (5' CGATCTCTTGT 3') in 5' end of SARS-Cov-2 (nucleotides 47-53) and all other human-infecting betacoronavirus candidates (Fig. 1A). The motif was the most significant prediction among three predicted motifs by MEME (Supplementary Fig. S1A). The motif consists of a cysteine followed by most pyrimidine nucleotides, which is similar to 5' terminal oligopyrimidine (TOP) motif in 5' end of eukaryotic mRNAs. Global alignment found that TOP-like motif was also conserved in location (near position 50 from 5' end) in 5' UTR of human-infecting betacoronaviruses (Fig. 1B), but other motifs were not (Supplementary Fig. S1C). This motif was not found in the 5'UTRs of alphacoronavirus candidates but other motifs that enrich in pyrimidines were identified (Supplementary Fig. S1B) and their locations were quite conserved (Supplementary Fig. S1D). Structural analysis revealed the location of TOP-like motif was in SL2 (SL: stem loop) of 5'UTR of all betacoronavirus candidates (Supplementary Fig. S2). SL2 is the most conserved secondary structure in the coronavirus 5′UTR and has been shown to be important for subgenomic synthesis[4].

3.2. Presence of a known octanucleotide motif with unknown biological function in 3'UTR of SARS-CoV-2

Our analysis revealed an octanucleotide motif (5' GGAAGAGC 3') was completely conserved in novel virus SARS-CoV-2 (location: 120-127 in the 3'UTR) and other human-infecting beta-, alpha- coronaviruses (Fig. 2A). Global alignment showed that the location of the octanucleotide sequence in 3' UTR of those coronaviruses was not conserved (Fig. 2B); however, secondary structure analysis indicated that the octanucleotide resided in a stem-loop structure of all coronaviral candidates including SARS-CoV-2 (Supplementary Fig. S3). The octanucleotide motif is previously indicated to be universally conserved in coronaviruses; however, its role in viral pathogenesis remains unclear[5]. As we realized that the octanucleotide located within the stem loop structure in 3'UTR of coronaviruses; we predicted that this sequence could be a part of viral miRNA molecules.

3.3. Prediction of the role of the octanucleotide motif in the viral pathogenesis through miRNA pathway

Hairpin prediction by Vmir program revealed three potential SARS-CoV-2 microRNA precursors (V-md1, V-md2, and V-mr1) in 3'UTR but not 5'UTR. V-md1 and V-md2 was discovered in direct orientation while v-mr1 was found in reverse orientation. The precursors (V-md1 and V-mr1) with higher scores contained the octanucleotide (Supplementary Table S4). Our analysis revealed the existence of conserved s2m element in SARS-CoV-2 (Figure S5) and this conserved element resided in a stem loop of v-mr1 (Supplementary Table S4). s2m is a mobile genetic element

found in 3'UTR of many coronaviruses. As s2m folds into a stem loop structure, it has been suggested to encode viral miRNAs[21].

It is previously showed virus could encode miRNAs as orthologs of host's miRNAs [22-24]; therefore, it is reasonable to assume that SARS-CoV-2-derived miRNAs are orthologs of hosts' miRNAs. As miRNAs in bat and pangolin, the likely natural reservoir of SARS-CoV-2, are unknown at present, we carried out the comparison of viral precursor sequences with human and other animals' mature miRNAs in order to predict mature miRNA sequence of the virus. V-md1 and V-mr1 aligned well with three human mature miRNAs while there was no homology of V-md2 in human (Fig. 3). Particularly, V-md1 precursor, which totally aligned with four mammalian miRNAs (Supplementary Fig. S6B), were homologous with mature sequence of has-miR-6769a-5p and has-miR-1270 so V-md1 was predicted to be processed into two viral mature miRNAs named vd1-miR-6769a-5p and vd1-miR-1270 (Fig. 3A). V-mr1 probably produced two viral mature miRNAs named vr1-miR-1270 and vr-miR-1307-3p as V-mr1 aligned with two human mature miRNAs (Fig. 3C). Besides, our extensive alignments showed that V-md1 alternatively aligned with other two mammalian mature miRNAs and especially a part of V-mr1 was an ortholog of miR-1307 of 13 different mammalians (Supplementary Fig. S6A). Interestingly, this part was derived from s2m element in v-mr1. Secondary structure analysis suggested that vd1-miR-6769a-5p and vd1-miR-1270 had good hairpin loop structures. In contrast, vr1-miR-1270 and vr1-miR-1307-3p was less likely because mature sequence of these predicted viral miRNAs respectively resides on a big internal loop and stem loop of v-mr1 secondary structure

(Fig. 3D). Notably, the octanucleotide located in v-md1 and v-mr1 viral miRNA precursors and even in their predicted mature sequences. The octanucleotide locates 1-8 and 4-12 in vr1-miR-1270, and vd1-miR-6769a-5p respectively (Supplementary Fig. S7), which implicates that the octanucleotide contribute to the seed region and binding site of viral miRNAs. In summary, four viral miRNAs consisting of vd1-miR-6769a-3p, vd1-miR-1270, vr1-miR-1307-3p, and vr1-miR-1270 were predicted to be produced by the 3'UTR of SARS-CoV-2. These predicted miRNAs were further searched for their potential human targets.

Sfold program[20] was used to identify the miRNAs' target sites in 3'UTR, CDS, 5'UTR seed regions of genes involving in imminity response pathways, stress response, cellular translation, and DNA methylation. Most significant target sites were at 3'UTRs of the genes. Four predicted viral miRNAs hybridized significantly with 27 human mRNAs at 79 target sites (Supplementary Table S8) and distribution of 79 sites for each viral miRNAs and different seed regions is shown in supplementary data (Supplementary Fig. S9). This showed that viral miRNAs are likely to modulate a wide range of genes with various functions including innate immune stimulation, cytokine production, antigen representation, cellular translation, DNA methylation/demethylation, and ER stress response control. Three viral microRNAs (vd1-miR-1270, vr-miR-1270, and vd1-miR-6769a-5p) produced 20 most significant interactions with 11 human mRNA transcripts of interferon gamma (INFG), interferon-stimulating genes (STAT5A, STAT5B)[26], antigen presentation genes (HLA-A, TAP1, TAP2)[27], a negative regulator of cytokine production (PCBP2)[10], a negative regulator

of ER stress response (BI-1)[28], and initiation translation factors (EIF5, EIF6, EIF4E) (Table 3). Hybridization sites revealed that the octanucleotide contributed to seed site of vd1-miR-6769a-5p and vr1-miR-1270 respectively but not of vd1-miR-1270. Notably, INF-γ were targeted at three sites at 3'UTR seed by vd1-miR-6769a-5p and vd1-miR-1270. Although vr1-miR-1307-3p was conserved in many mammalians and derived from s2m element, vr1-miR-1307-3p does not have many target sites on our human gene candidates (Supplementary Table S7).

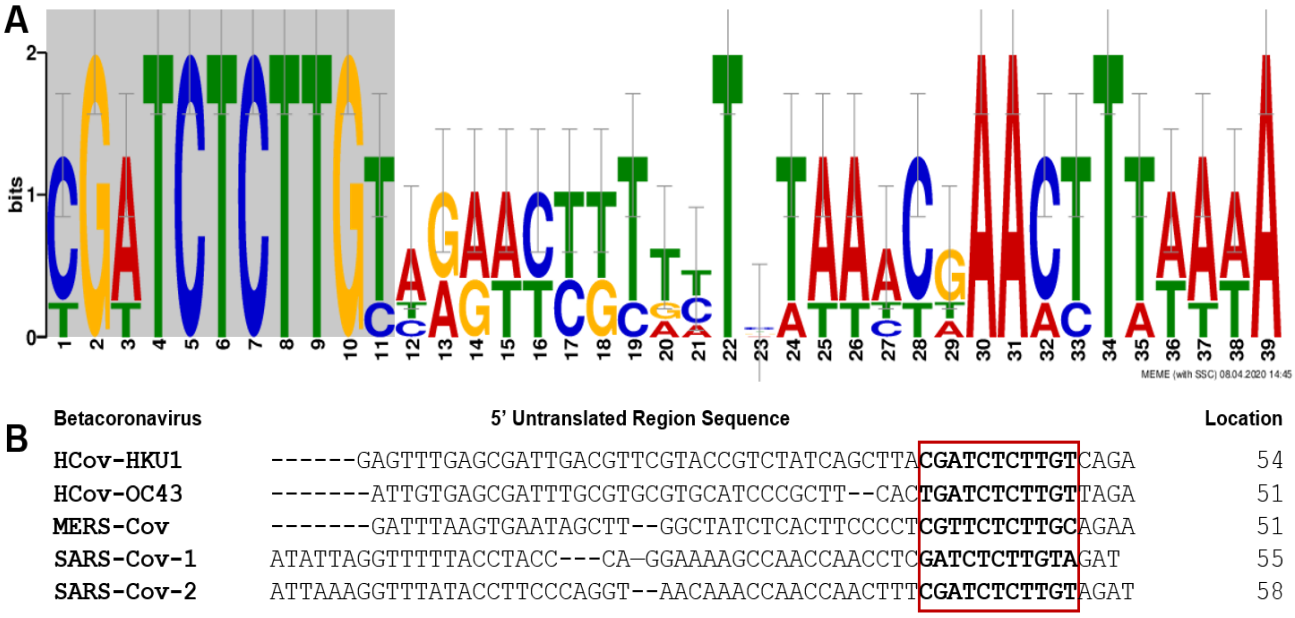


Figure 1. Motif discovery in 5'UTR of betacoronavirus candidates

(A) Motif discovery by ungapped motif search tool (MEME) in betacoronavirus (n=5). Conserved oligopyrimidine motif is highlighted by a gray box (e-value = 1.3e-011). **(B)** Global multiple alignment of 5'UTR of betacoronavirus (n=5) by Clustal Omega. Conserved oligopyrimidine motif is bold and in a red frame.

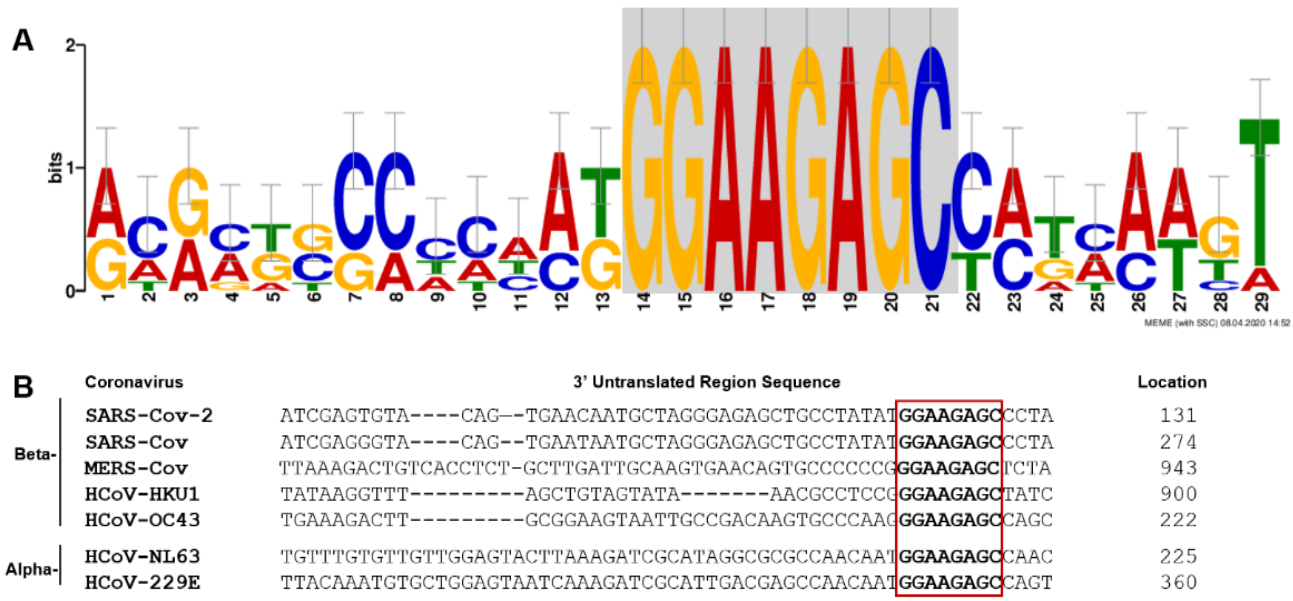


Figure 2. Motif discovery in 3'UTR of alpha- and beta- coronavirus candidates

(A) Motif discovery by ungapped motif search tool (MEME) in betacoronavirus (n=5) and alphacoronavirus (n=2). Conserved octanucleotide motif is highlighted by a gray box (e-value = 3.0e-008). **(B)** Global multiple alignment of in betacoronavirus (n=5) and human-infecting alphacoronavirus (n=2) by Clustal Omega. Conserved octanucleotide sequence is bold and in a red frame.

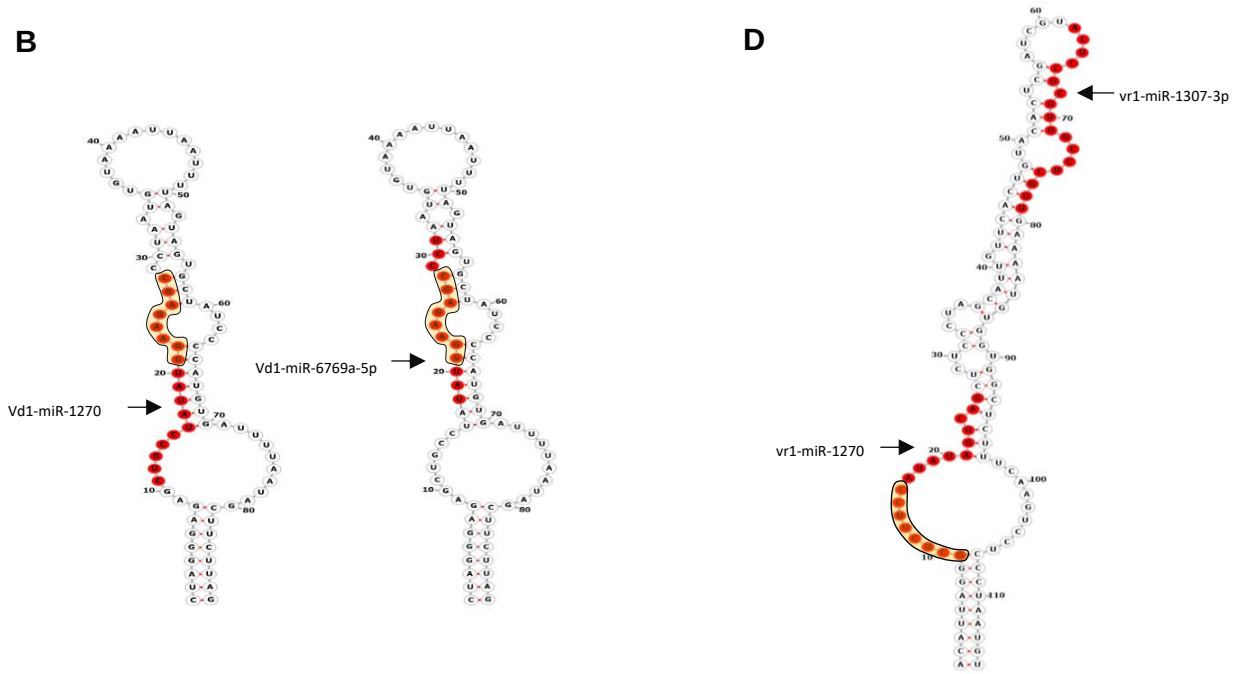


Figure 3. Alignment with mammalian mature miRNAs and hairpin structure of SARS-CoV-2 viral precursor miRNAs (v-md1, v-mr1)

(A) Alignment of v-md1 with mature miRNA of mammalia. Sequence in red is predicted to be viral mature miRNA and octanucleotide sequence highlighted in yellow box. Human mature miRNA sequence is bold **(B)** Predicted hairpin structure of v-md1 by RNAfold (MFE = -17.35 kcal/mol)**.** Sequence in red is a predicted viral mature miRNA conserved in human. Sequence highlighted in yellow is octanucleotide

(C) Alignment of v-mr1 with mature miRNA of mammalian. Sequence in red is predicted to be viral mature miRNA and octanucleotide sequence highlighted in yellow box. Human mature miRNA sequence is bold. **(D)** Predicted hairpin structure of v-md1 by RNAfold (MFE = -32.50 kcal/mol). Sequence in red is a predicted viral mature miRNA conserved in human. Sequence highlighted in yellow is octanucleotide. MFE: Minimum Free Energy.

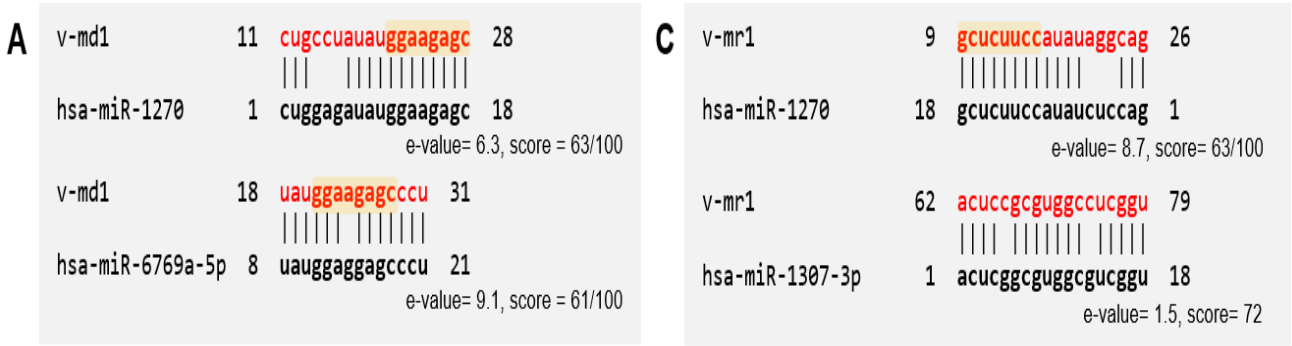
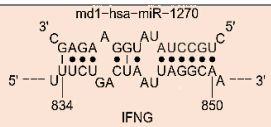
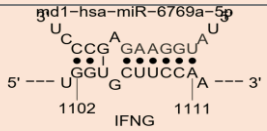
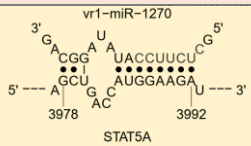
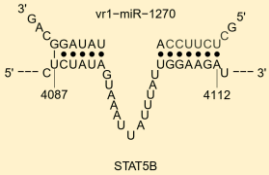
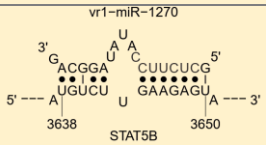
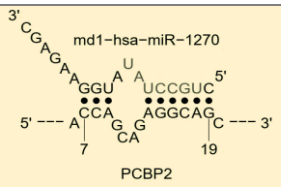
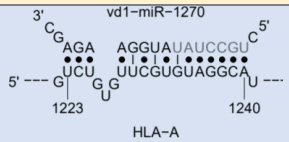
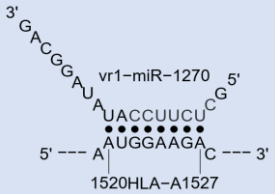
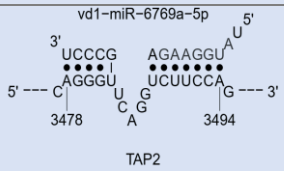
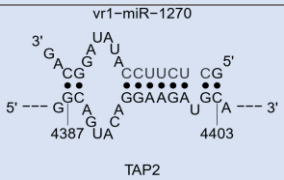
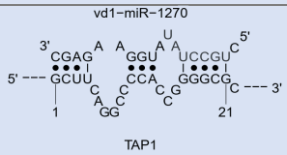
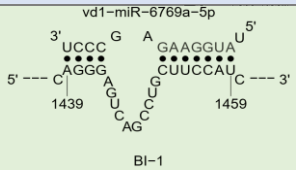
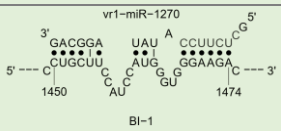
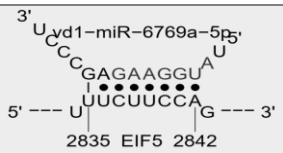
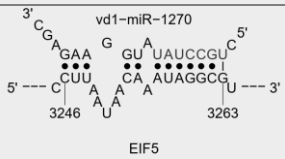
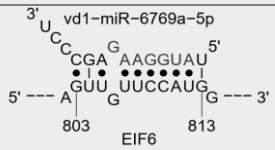
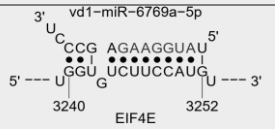


Table 3. RNA hybridization of predicted SARS-CoV-2 encoded miRNAs with human targets by SFold

No.	Category	Gene	Viral miRNA	Hybridization confirmation	Seed type	ΔG total	ΔG hybrid
1	Interferon	**IFN-γ**	vd1-miR-1270		7mer-A1	-8.841	-17.600
			vd1-miR-1270		6mer	-8.163	-21.100
			vd1-miR-6769a-5p		Offset-6mer	-8.713	-17.400
2	Interferon/ cytokine regulation	**STAT5A**	vr1-miR-1270		Offset-6mer	-7.172	-19.600
		STAT5B	vr1-miR-1270		Offset-6mer	-9.861	-16.600
			vr1-miR-1270		6mer	-12.389	-19.200

		PCBP2	vd1-miR-1270		offset-6mer	-16.300	-9.587
3	Antigen representation	**HLA-A**	vd1-miR-1270		6mer	-6.262	-23.200
			vr1-miR-1270		Offset-6mer	-13.463	-15.900
		TAP2	vd1-miR-6769a-5p		offset-6mer	-3.156	-21.800
			Vr1-miR-1270		offset-6mer	-3.148	-18.000
		TAP1	vr1-miR-1270		6mer	-6.275	-18.300
4	ER stress response	**BI-1**	vd1-miR-6769a-5p		7mer-m8	-2.040	-18.300
			vr1-miR-1270		offset-6mer	-1.836	-18.800

5	Translation initiation	**EIF5**	vd1-miR-6769a-5p		offset-6mer	-11.832	-15.600
			vd1-miR-1270		Offset-6mer	-8.864	-16.400
		EIF6	vd1-miR-6769a-5p		offset-6mer	-12.362	-16.100
		EIF4E	vd1-miR-1270		6mer	-12.019	-19.100
			vd1-miR-6769a-5p		7mer-m8	-5.317	-19.300

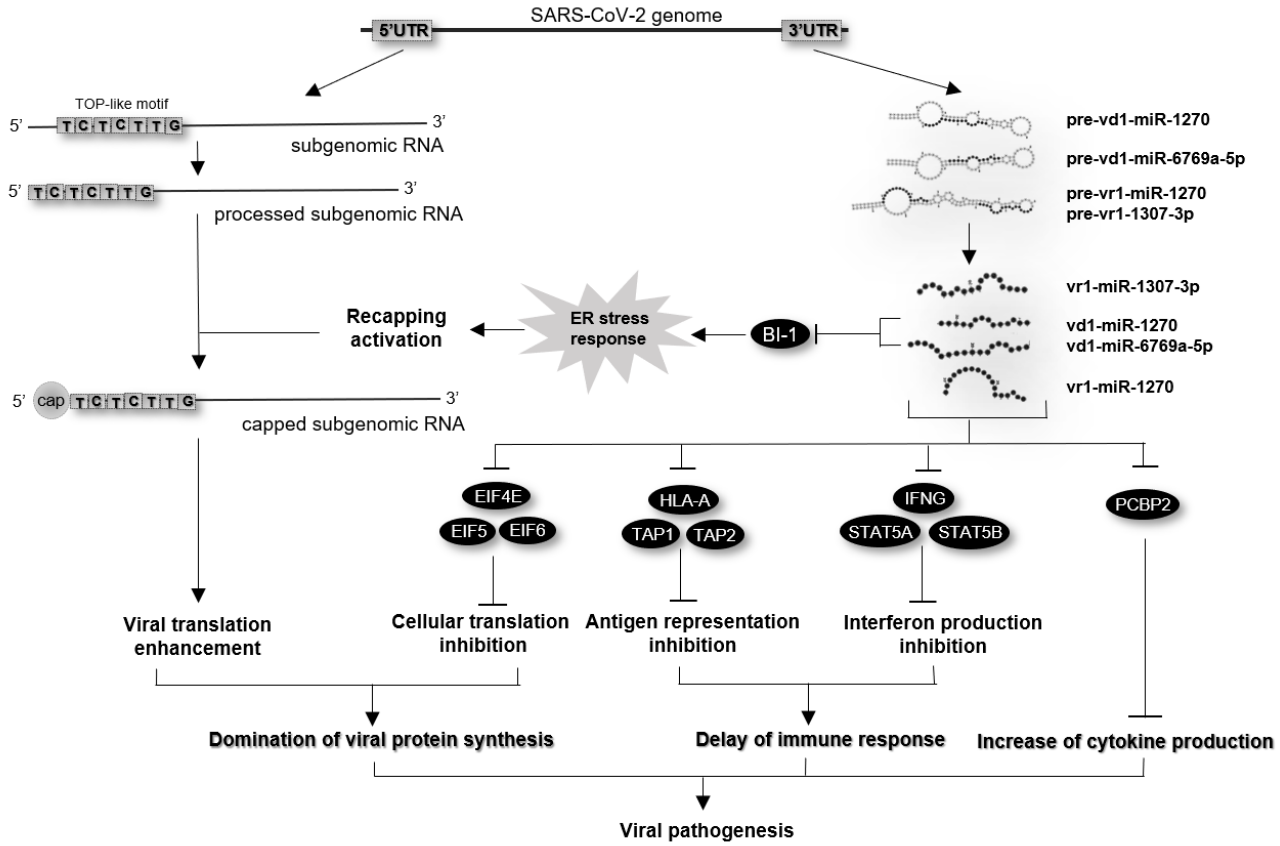


Figure 4. Model of cellular expression and immune evasion by SARS-CoV-2 virus' non-coding elements.

3'UTR of SARS-CoV-2 predictably encodes for four mature miRNAs. Three of them (Vd1-miR-1270, vr1-miR-1270 and vd1-miR6769a-5p) significantly target interferon gamma and interferon-inducing genes (STAT5A, STAT5B), which inhibits production of interferons for anti-viral response. Additionally, these viral miRNAs inhibit translation of HLA-A, TAP1, and TAP2 mRNA transcripts to interfere antigen representation process, thereby preventing from infected-cell regconition. These two effects contribute to the delay of immune response in general. In another scenario, vd1-miR-1270 could inhibit PCBP2, a negative regulator of cytokine production in order to induce cytokine production, consequently may contribute to the phenomenon of cytokine storm. Viral miRNAs also target eukaryotic translation initiation factors (EIF4E, EIF5, EIF6); therefore, inhibiting host translation. Vd1-miR-1270 and vd1-miR6769a-5p potentially target BI-1, a negative regulator of ER stress response, thereby inducing ER stress response. ER stress subsequently activate recapping process which targets TOP mRNAs. As 5'UTR of SARS-CoV-2's genome contains a TOP-like motif, it is likely to produce subgenomic RNAs containing TOP-like motif at 5'UTR terminal after processing. This make these viral RNAs similar to host TOP mRNAs. As a result, TOP-like motif viral RNAs might undergo recapping pathway as host TOP mRNAs to activate translatability and promote viral translation. Both cellular expression inhibition and viral translation enhancement effect lead to domination of viral protein synthesis. In summary, domination of viral protein synthesis, delay of immune response, and

increase of cytokine production which are mediated by TOP-like-motif and viral miRNAs simultaneously contribute to viral pathogenesis.

IV. Discussion

In order to identify important molecular features in UTRs of SARS-CoV-2's genome, we have utilised classical methods of identifying conserved motifs in these regions. The virus has just been classified into two type S and L recently and does not have an adequate time to change significantly within genome of these two types[29]. Therefore, we compared the UTR regions of these two types and different human-infecting coronaviruses to find significantly real conserved motifs. In doing so, we identified a novel conserved motif (*5'*-CGAUCUCUUGT-*3'*) in 5'UTR and a known conserved motif (*5'*-GGAAGAGC-*3'*) in 3'UTR of SARS-CoV-2's genome. The motif in the 5'UTR is also conserved in its location in all betacoronavirus candidates' genomes. In the next step, we carried out bioinformatic and literature analyses to predict biological roles of these motifs in activities and pathogenesis of the virus. We suggest a model of cellular expression and immune evasion by SARS-CoV-2's non-coding elements (Figure 4).

The novel conserved motif that we discovered in 5' UTR of SARS-CoV-2 genome consists of seven nucleotides in length and mainly pyrimidines (U/C), which makes SARS-CoV-2 5'UTR motif similar to 5' terminal oligopyrimidine (TOP) motif. TOP is a sequence of 4–14 pyrimidines following a cysteine and locates in 5' end of many eukaryotic mRNAs known as TOP mRNAs. However, the difference between our motif and TOP is that the novel motif locates at position 47-53 of the 5'UTR, and TOP

motif is adjacent to 5'terminal of eukaryotic mRNAs[30]. The latest study has revealed the processing of new coronavirus's UTRs to produce subgenomic RNAs with shorter UTRs[31]. Hence, it is possible that the novel TOP-like motif becomes 5'terminal of subgenomic RNAs after undergoing UTR processing. We also want to note that while most eukaryotic TOP follows to a cysteine at 5' terminal, our motif locates downstream of and non-pyrimidine nucleotide distance from a well-conserved cysteine. Because of location and sequence similarities between two motifs, SARS-CoV-2's RNAs may undergo cellular processes that TOP mRNAs of Eukaryotes naturally undergo. Recent evidence indicates that TOP is a signal for recapping process which restores translatability and subsequently enhances expression of uncapped TOP mRNAs[13]. Acquisition of 5' cap structure is extremely critical for the stability and translation initiation of mRNAs in Eukaryotes as well as viruses[32]. Coronaviruses were also cap-dependent and evolves their own enzymes for processing cap[5] as Eukaryotes, in specific, SARS-CoV-2 genome has enough encoding non-structural proteins for capping process[33]. Hence, we propose a hypothesis that SARS-CoV-2 exploits host recapping pathway via TOP-like motif residing at 5' terminal of processed subgenomic RNAs in order to alternatively enhance viral translation.

Recapping pathway is likely activated during stress response because inhibition of cytoplasmic recapping activity reduces ability of cells to recover from stress[34]. Coronaviruses have been shown to induce stress responses, especially endoplasmic reticulum (ER) stress response in host cells[35-37]. Due to ER stress induction, recapping pathway can be activated to enhance the translation of cellular TOP mRNAs, and

SARS-CoV-2 RNAs containing TOP-like motif at 5' terminal can take an advantage from activation of recapping pathway to increase their translation. Interestingly, our analysis identifies viral miRNAs that might support 5' cap acquisition of viral RNAs by suppressing a well-known negative regulator BI-1 of ER stress response[28]. In particular, both vd1-miR-6769a-5p and vr1-miR-1270 predictably significantly target BI-1 mRNA at 3'UTR seed and with low RNA hybrid energy. Therefore, SARS-CoV-2 is likely to exploit ER stress response-induced recapping activation to gain 5' cap of host cells. Since acquisition of 5' cap promote viral protein synthesis during infection, this activity will contribute to the viral pathogenesis. Ribose 2'-O-methylation in viral caps allows virus to escape cellular sensor of foreign cap by mimicking host caps, thereby not triggering type I interferon production for anti-viral response[38]. Depletion of 2'O methyltransferase increased sensitivity to immune response and caused attenuation of SARS-CoV virus in both *in vitro* and *in vivo*[39], raising a promising target for vaccine and drug development. Because previously interfering viral capping is effective for SARS-CoV attenuation, disrupting TOP-like motif might attenuate SARS-CoV-2, too.

Octanucleotide 5'GGAAGAGC 3' has been indicated to be conserved in the 3' UTRs of coronaviruses for a long time; however, role of this conserved motif remains unclear [5]. Mutational analysis on the octanucleotide reveals that it is not essential for viral RNA synthesis but deletion of octanucleotide-containing hypervarible region causes a dramatic attenuation of virulence in the mouse[40]. Our results recommend a biological function of the octanucleotide as a seed region in viral predicted microRNAs

which target multiple genes including interferons, interferon-inducing genes, antigen-representation genes, translation factor genes, and ER stress regulatory gene.

Viruses develop antagonistic mechanisms to evade host innate immune in general[41]. Lack of interferon response in the early stage of coronavirus infection and delayed immune response[42] can be resulted from viral innate immune evasion. Here, we suggest a viral miRNA-mediated antagonistic mechanism that potentially contributes to the phenomenon of delayed innate immune response. This consequently may lead to long incubation period and asymptomatic phenomenon in SARS-CoV-2 infection. After pathogen recognition, normally INF-β and IFN-γ, which are key pro-inflammatory cytokines, are secreted to stimulate antiviral effect of innate immune system[43] and consequently inflammatory symptoms within several days since viral infection. However, incubation period of SARS-CoV-2 infection (2-14 days, mean 5.1 days) is longer than normal[44]. This could be due to the delayed innate immune caused by SARS-CoV-2-encoded miRNAs (vd1-miR-6769a-5p, vd1-miR-1270, and vr1-miR-1270). After SARS-CoV-2 infection, viral miRNAs might be produced to inhibit expression of interferon gamma via targeting IFN-γ inducing genes such as IFN-β, STAT5B, STAT5A mRNA and directly targeting IFN-γ mRNA. Inhibition of IFN-γ expression can disrupt stimulation of inflammatory cytokine production and further antiviral response and viral clearance. The miRNA-interaction pathway also helps to explain why a previous investigation did not find IFN system (IFN production and IFN signaling pathway) was suppressed in SARS-CoV infection when measuring mRNA levels[45,46].

As a result, virus could bypass host antiviral activity by delaying immune response, thereby allowing patients being asymptomatic in the early stage of infection. This proposed mechanism is consistent with the following pre-clinical features: the decrease of interferon-γ is observed in serum of SARS patient[47], and low interferon concentration links to severity of SARS-CoV-2 progression[48]. This can be a reason why interferon-γ is showed to be effective to treat SARS-CoV in *vi vitro*[49]. In fact, decrease of STAT5A and STAT5B to bellow critical threshold induces production of pro-inflammatory cytokines. This is a reason why virus only could evade immune system for a certain period and when host cells bypass viral antagonistic mechanisms, immune reactions will be stimulated to fight against viruses, resulting in clinical symptoms[26].

Besides interfering INF signaling, SARS-CoV-2 could involve mechanism of preventing the antigen representation of MHC class I which triggers activity of T cell against intracellular virus and pathogen. Our prediction demonstrates that expression of HLA-A[50], TAP1 and TAP2 (HLA-A transporter)[27] can be decreased by vr1-miR-1270, vd1-miR-1270 and vd1-miR-6769a-5p respectively, which limits antigen representation on the surface of antigen-representing cells and latter limiting activity of T cells. Hence, by miRNA-mediated antagonizing antigen representation, SARS-CoV-2 can escape host immune attack when adaptive immune starts in the middle stage of viral infection.

Severity of SARS-CoV-2 infection is linked to cytokine storm in post stage. Here, we found that PCPB2 was a common target of SARS-CoV-2 and H5N1 viral microRNAs. H5N1 virus is proved to express microRNA-like small RNA, miR-HA-3p in order to suppress expression of PCPB2. The inhibition of PCPB2 expression mediates

enhanced cytokine production and increased risk of mortality[10]. Therefore, similar to miR-HA-3p, vr1-miR-1270 encoded by SARS-CoV-2 can be a virulence factor for cytokine storm in SARS-CoV-2 infection[51].

Finally, we suggest that RNA viruses suppress host translation factors to reduce host protein synthesis and activate an alternative pathway that favors viral RNA translation[52]. The results predicted that translation initiation factors (EIF4E, EIF5, and EIF6) were targeted by vd1-miR-6769a-5p and vd1-miR-1270. Thus, SARS-CoV-2 possibly antagonizes normal translation pathways to induce non-canonical translation mechanisms[53] that allows to translate viral RNAs, thereby enhancing viral protein synthesis.

V. Conclusion

In summary, in this study we clearly identified two conserved oligonucleotide motifs in UTRs of SARS-CoV-2's genome and provided sound explanations for pathogenic role of the non-coding elements and virus-host interactions. The function of novel TOP-like motif in 5'UTR and the octanucleotide motif in 3'UTR of SARS-CoV-2 are proposed to relate to recapping-mediated cap acquisition and viral miRNA mechanism respectively. The novel TOP-like motif in 5'UTR can be an option for vaccine and drug development for SARS-CoV-2 infection. SARS-CoV-2 might develop multiple strategies to modulate many cellular processes (interferon production, antigen representation, ER stress response, and translation initiation) using virus-encoded miRNAs (vr1-miR-1270, vd1-miR-6769a-5p and vd1-miR-1270) in order to maintain viral infection in different infection stages. These miRNA interactions can explain a

serious discrepancy in previous experiments when only mRNA levels were measured. Interestingly, the proposed interaction pathways link together the effect of both 5'UTR and 3'UTR on the viral pathogenesis. We firmly believe that these results are worth to investigate experimentally to produce more detailed insights into viral pathogenesis and novel targets for anti-viral drug development.

Author contribution: K.D.M.N: investigation, analyze, data treatment, and writing ; N.H.M.K ; T.A.K.; N.T.L: analyze, data treatment, and writing; L.D.D.: conceptualization, review the investigation and methodology, supervision, reviewing, and editing. All authors read and approved the final manuscript.

Data availability: All data and software in this study are available online as referred in this article.

Declarations: Ethical approval - Not applicable.

References

1. Valencia DN. Brief Review on COVID-19: The 2020 Pandemic Caused by SARS-CoV-2. *Cureus* (2020) doi:10.7759/cureus.7386.
2. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020 Mar 13; **367**(6483):1260-1263. doi: 10.1126/science.abb2507.
3. Johansson MA, Quandelacy TM, Kada S, et al. SARS-CoV-2 Transmission From People Without COVID-19 Symptoms. *JAMA Netw Open*. 2021;4(1):e2035057. doi:10.1001/jamanetworkopen.2020.3505
4. Boyton RJ, Altmann DM. The immunology of asymptomatic SARS-CoV-2 infection: what are the key questions? *Nat Rev Immunol.* 2021 Dec; **21**(12):762-768. doi: 10.1038/s41577-021-00631-x.
5. Yang, D. & Leibowitz, J. L. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res.* **206**, 120–133 (2015).
6. Lan, T.C.T., Allan, M.F., Malsick, L.E. et al. Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. *Nat Commun* **13**, 1128 (2022). https://doi.org/10.1038/s41467-022-28603-2.
7. Goebel SJ, Miller TB, Bennett CJ, Bernard KA, Masters PS. A hypervariable region within the 3' cis-acting element of the murine coronavirus genome is nonessential for RNA synthesis but affects pathogenesis. *J Virol.* 2007 Feb;**81**(3):1274-87. doi: 10.1128/JVI.00803-06.

8. Cox JE, Sullivan CS. Balance and Stealth: The Role of Noncoding RNAs in the Regulation of Virus Gene Expression. *Annu Rev Virol.* 2014 Nov;**1(1)**:89-109. doi: 10.1146/annurev-virology-031413-085439.

9. Tycowski KT, Guo YE, Lee N, et al. Viral noncoding RNAs: more surprises. *Genes Dev.* 2015 Mar 15;**29(6)**:567-84. doi: 10.1101/gad.259077.115.

10. Hussain M, Asgari S. MicroRNA-like viral small RNA from Dengue virus 2 autoregulates its replication in mosquito cells. Proc Natl Acad Sci U S A. 2014 Feb 18;**111(7)**:2746-51. doi: 10.1073/pnas.1320123111.

11. Li X, Fu Z, Liang H, et al. H5N1 influenza virus-specific miRNA-like small RNA increases cytokine production and mouse mortality via targeting poly(rC)-binding protein 2. *Cell Res.* 2018 Feb;**28(2)**:157-171. doi: 10.1038/cr.2018.3.

12. Islam MS, Khan MA, Murad MW, Karim M, Islam ABMMK. In silico analysis revealed Zika virus miRNAs associated with viral pathogenesis through alteration of host genes involved in immune response and neurological functions. *J Med Virol.* 2019 Sep;**91(9)**:1584-1594. doi: 10.1002/jmv.25505..

13. Withers JB, Mondol V, Pawlica P, etal.. Idiosyncrasies of Viral Noncoding RNAs Provide Insights into Host Cell Biology. *Annu Rev Virol.* 2019 Sep 29;**6(1)**:297-317. doi: 10.1146/annurev-virology-092818-015811.

14. Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009 Jul;**37**(Web Server issue):W202-8. doi: 10.1093/nar/gkp335.

15. Sullivan CS, Grundhoff A. Identification of viral microRNAs. **Methods Enzymol.** 2007;**427**:3-23. doi: 10.1016/S0076-6879(07)27001-6.

16. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. **Nucleic Acids Res.** 2008 Jul 1;**36**(Web Server issue):W70-4. doi: 10.1093/nar/gkn188.

17. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 2008 Jan;**36**(Database issue):D154-8. doi: 10.1093/nar/gkm952.

18. Menachery VD, Schäfer A, Burnum-Johnson KE, et al. MERS-CoV and H5N1 influenza virus antagonize antigen presentation by altering the epigenetic landscape. *Proc Natl Acad Sci U S A.* 2018 Jan 30;**115(5)**:E1012-E1021. doi: 10.1073/pnas.1706928115.

19. Nakagawa K, Lokugamage KG, Makino S. Viral and Cellular mRNA Translation in Coronavirus-

Infected Cells. *Adv Virus Res.* 2016;**96**:165-192. doi: 10.1016/bs.aivir.2016.08.001.

20. Ding Y, Chan CY, Lawrence CE. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.* 2004 Jul 1;**32**(Web Server issue):W135-41. doi: 10.1093/nar/gkh449.

21.Del Valle Morales D, Trotman JB, Bundschuh R, Schoenberg DR. Inhibition of cytoplasmic cap methylation identifies 5' TOP mRNAs as recapping targets and reveals recapping sites downstream of native 5' ends. *Nucleic Acids Res.* 2020 Apr 17;**48(7)**:3806-3815. doi: 10.1093/nar/gkaa046..

22. Tengs T, Jonassen CM. Distribution and Evolutionary History of the Mobile Genetic Element s2m in Coronaviruses. *Diseases*. 2016 Jul 28;**4(3)**:27. doi: 10.3390/diseases4030027.

23. Gottwein E, Mukherjee N, Sachse C, et al. A viral microRNA functions as an orthologue of cellular miR-155. *Nature*. 2007 Dec 13;**450(7172)**:1096-9. doi: 10.1038/nature05992.

24. Skalsky RL, Samols MA, Plaisance KB, et al. Kaposi's sarcoma-associated herpesvirus encodes an ortholog of miR-155. *J Virol*. 2007 Dec;**81(23)**:12836-45. doi: 10.1128/JVI.01804-07.

25. Manzano M, Shamulailatpam P, Raja AN, Gottwein E. Kaposi's sarcoma-associated herpesvirus encodes a mimic of cellular miR-23. *J Virol*. 2013 Nov;**87(21)**:11821-30. doi: 10.1128/JVI.01692-13.

26. Long D, Lee R, Williams P, Chan CY, Ambros V, Ding Y. Potent effect of target structure on microRNA function. *Nat Struct Mol Biol*. 2007 Apr;**14(4)**:287-94. doi: 10.1038/nsmb1226.

27. Hedl M, Sun R, Huang C, Abraham C. STAT3 and STAT5 Signaling Thresholds Determine Distinct Regulation for Innate Receptor-Induced Inflammatory Cytokines, and STAT3/STAT5 Disease Variants Modulate These Outcomes. *J Immunol*. 2019 Dec 15;**203(12)**:3325-3338. doi: 10.4049/jimmunol.1900031.

28. Corradi V, Singh G, Tieleman DP. The human transporter associated with antigen processing: molecular models to describe peptide binding competent states. *J Biol Chem*. 2012 Aug 10;**287(33)**:28099-111. doi: 10.1074/jbc.M112.381251.

29. Lisbona F, Rojas-Rivera D, Thielen P, et al. BAX inhibitor-1 is a negative regulator of the ER stress sensor IRE1alpha. *Mol Cell.* 2009 Mar 27;**33(6)**:679-91. doi: 10.1016/j.molcel.2009.02.017.

30. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z, Cui J, Lu J. On the origin and continuing evolution of SARS-CoV-2. *Natl Sci Rev.* 2020 Jun;**7(6)**:1012-1023. doi: 10.1093/nsr/nwaa036.

31. Lahr RM, Fonseca BD, Ciotti GE, Al-Ashtal HA, Jia JJ, Niklaus MR, Blagden SP, Alain T, Berman AJ. La-related protein 1 (LARP1) binds the mRNA cap, blocking eIF4F assembly on TOP mRNAs. *Elife.* 2017 Apr 7;**6:e24146**. doi: 10.7554/eLife.24146

32. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. The Architecture of SARS-CoV-2 Transcriptome. *Cell.* 2020 May 14;**181(4)**:914-921.e10. doi: 10.1016/j.cell.2020.04.011.

33. Ramanathan A, Robb GB, Chan SH. mRNA capping: biological functions and applications. *Nucleic Acids Res.* 2016 Sep 19;**44(16)**:7511-26. doi: 10.1093/nar/gkw551.

34. Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome. https://www.ncbi.nlm.nih.gov/nuccore/1798174254.

35. Mukherjee C, Patil DP, Kennedy BA, Bakthavachalu B, Bundschuh R, Schoenberg DR. Identification of cytoplasmic capping targets reveals a role for cap homeostasis in translation and mRNA stability. *Cell Rep.* 2012 Sep 27;**2(3)**:674-84. doi: 10.1016/j.celrep.2012.07.011.

36. Fung TS, Huang M, Liu DX. Coronavirus-induced ER stress response and its involvement in regulation of coronavirus-host interactions. *Virus Res.* 2014 Dec 19;**194**:110-23. doi: 10.1016/j.virusres.2014.09.016.

37. Jindal S, Malkovsky M. Stress responses to viral infection. *Trends Microbiol.* 1994 Mar;**2(3)**:89-91. doi: 10.1016/0966-842x(94)90540-1.

38. Chan CP, Siu KL, Chin KT, Yuen KY, Zheng B, Jin DY. Modulation of the unfolded protein response by the severe acute respiratory syndrome coronavirus spike protein. *J Virol.* 2006 Sep;**80(18)**:9279-87. doi: 10.1128/JVI.00659-06.

39. Züst R, Cervantes-Barragan L, Habjan M, et al. Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nat Immunol.* 2011 Feb;**12(2)**:137-43. doi: 10.1038/ni.1979.

40. Menachery VD, Yount BL Jr, Josset L, et al. Attenuation and restoration of severe acute respiratory syndrome coronavirus mutant lacking 2'-o-methyltransferase activity. *J Virol.* 2014 Apr;**88(8)**:4251-64. doi: 10.1128/JVI.03571-13.

41. Goebel SJ, Miller TB, Bennett CJ, Bernard KA, Masters PS. A hypervariable region within the 3' cis-acting element of the murine coronavirus genome is nonessential for RNA synthesis but affects pathogenesis. *J Virol.* 2007 Feb;**81(3)**:1274-87. doi: 10.1128/JVI.00803-06.

42. Mishra R, Kumar A, Ingle H, Kumar H. The Interplay Between Viral-Derived miRNAs and Host Immunity During Infection. *Front Immunol.* 2020 Jan 23;**10**:3079. doi: 10.3389/fimmu.2019.03079.

43. Kikkert M. Innate Immune Evasion by Human Respiratory RNA Viruses. *J Innate Immun.* 2020;**12(1)**:4-20. doi: 10.1159/000503030.

44. Kang S, Brown HM, Hwang S. Direct Antiviral Mechanisms of Interferon-Gamma. *Immune Netw.* 2018 Oct 17;**18(5)**:e33. doi: 10.4110/in.2018.18.e33.

45. Lauer SA, Grantz KH, Bi Q, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med.* 2020 May 5;**172(9)**:577-582. doi: 10.7326/M20-0504.

46. Cheung CY, Poon LL, Ng IH, et al. Cytokine responses in severe acute respiratory syndrome coronavirus-infected macrophages in vitro: possible relevance to pathogenesis. *J Virol.* 2005 Jun;**79(12)**:7819-26. doi: 10.1128/JVI.79.12.7819-7826.2005.

47. Okabayashi T, Kariwa H, Yokota S, et al. Cytokine regulation in SARS coronavirus infection compared to other respiratory virus infections. *J Med Virol.* 2006 Apr;**78(4)**:417-24. doi: 10.1002/jmv.20556.

48. Zhang Y, Li J, Zhan Y, et al. Analysis of serum cytokines in patients with severe acute respiratory syndrome. *Infect Immun.* 2004 Aug;72(8):4410-5. doi: 10.1128/IAI.72.8.4410-4415.2004.

49. Pedersen SF, Ho YC. SARS-CoV-2: a storm is raging. *J Clin Invest.* 2020 May 1;130(5):2202-2205. doi: 10.1172/JCI137647.

50. Sainz B Jr, Mossel EC, Peters CJ, Garry RF. Interferon-beta and interferon-gamma synergistically inhibit the replication of severe acute respiratory syndrome-associated

coronavirus (SARS-CoV). *Virology.* 2004 Nov 10;**329(1)**:11-7. doi: 10.1016/j.virol.2004.08.011.

51. Chouaib S, Bensussan A, Termijtelen AM, et al. Allogeneic T cell activation triggering by MHC class I antigens. *J Immunol.* 1988 Jul 15;**141(2)**:423-9.

52. Mehta P, McAuley DF, Brown M, et al. HLH Across Speciality Collaboration, UK. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet.* 2020 Mar 28;**395(10229)**:1033-1034. doi: 10.1016/S0140-6736(20)30628-0.

53. Walsh D, Mathews MB, Mohr I. Tinkering with translation: protein synthesis in virus-infected cells. *Cold Spring Harb Perspect Biol.* 2013 Jan 1;**5(1)**:a012351. doi: 10.1101/cshperspect.a012351.