DOI 10.1007/s10517-023-05742-0
Cell Technologies in Biology and Medicine, No. 4, February, 2023

SARS-CoV-2 Mutations Lead to a Decrease in the Number of Tissue-Specific MicroRNA-Binding Regions in the Lung

A. P. Zhiyanov[1] and M. Yu. Shkurnikov[1,2]

Translated from *Kletochnye Tekhnologii v Biologii i Meditsine*, No. 4, pp. 216-221, December, 2022
Original article submitted November 2, 2022

RNA interference in vertebrates acts as an antiviral mechanism only in undifferentiated embryonic stem cells and is mediated by microRNAs. In somatic cells, host microRNAs also bind to the genomes of RNA viruses, regulating their translation and replication. It has been shown that viral (+)RNA can evolve under the influence of host cell miRNAs. In more than two years of the pandemic, the SARS-CoV-2 virus has mutated significantly. It is quite possible that some mutations could be retained in the virus genome under the influence of miRNAs produced by alveolar cells. We demonstrated that microRNAs in human lung tissue exert evolutionary pressure on the SARS-CoV-2 genome. Moreover, a significant number of sites of host microRNA binding with the virus genome are located in the NSP3-NSP5 region responsible for autoproteolysis of viral polypeptides.

Key Words: *microRNA; SARS-CoV-2; evolution; NSP3; NSP5*

RNA interference has emerged in plants and invertebrates as an antiviral mechanism that fights viral infections by degrading viral RNA after interacting with small interfering RNAs of an infected cell [1]. In vertebrates, RNA interference acts as an antiviral mechanism in undifferentiated embryonic stem cells only and is realized through miRNAs [2]. In differentiated somatic cells of vertebrates, type I IFNs respond to viral infection [3]. The vertebrate RNA interference system regulates mRNA activity through tissue- and cell-specific miRNA expression. These molecules can specifically bind to mRNA and regulate their translation in cells they are produced by and in neighbouring cells after entering them as part of exosomes [4,5]. Host miRNAs can also bind to the genomes of RNA viruses, regulating their translation and replication and changing the pathogenesis of viral infections [6-8]. There are two main effects of the interaction between the viral RNA genome and miRNA of the host cell: inhibition of virus translation and slowing down of its replication or stabilization of viral RNA and increase in the rate of virus replication. Moreover, the slowdown in viral replication is associated primarily with the interaction of microRNA with the 3'-untranslated region of the virus [9], and the stabilization of viral RNA is associated with interaction with the 5'-untranslated region of the virus [10]. Nevertheless, the influence of microRNA interaction with the protein-coding regions of viruses is still poorly understood.

It has been shown that (+)RNA viruses can evolve under the influence of host cell miRNAs [11]. Many works have been published on predicting the interaction of the SARS-CoV-2 single-stranded (+)RNA virus with human microRNAs [12-15]. However, these studies did not take into account the ability of the SARS-CoV-2 virus to mutate and, accordingly, to change the sequence of binding regions with human microRNA seed regions.

It should be noted that SARS-CoV-2 primarily multiplies in type 2 alveocytes of the lung tissue, which narrows the diversity of microRNAs interacting with it from 2000-3000 to 200-300 types characteristic of this tissue [16].

[1]M. M. Shemyakin and Yu. A. Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia; [2]Faculty of Biology and Biotechnologies, Higher School of Economics (HSE University), Moscow, Russia. ***Address for correspondence:*** mshkurnikov@hse.ru. M. Yu. Shkurnikov

For more than two years of the pandemic, the SARS-CoV-2 virus has significantly mutated, and not all of its mutations have led to changes in the sequence of viral proteins due to the degeneracy of the genetic code [17]. These mutations could stay in the virus genome under the influence of miRNAs produced by alveolar cells.

Within the framework of our study, we tested the hypothesis of the evolutionary pressure of microRNAs characteristic of human lung tissue on the genome of the SARS-CoV-2 virus.

MATERIALS AND METHODS

On September 5, 2022, 12,962,156 RNA sequences of the SARS-CoV-2 virus genome were uploaded from the GISAID sequencing database [18]. The resulting sequences were annotated using Pangolin [19] (program version: 4.1.2, data version: 1.14). Among 2520 sequences related to Moscow and Moscow region, variants of key strains of SARS-CoV-2 were selected that meet the following conditions: the percentage of unrecognized nucleotides in the virus RNA belongs to the interval (0%, 0.01%); the virus RNA length differs from the Wuhan reference strain (GIASID: EPI_ISL_402125) by less than 5%. The 273 sequences of the RNA genome of the virus selected for analysis are dated from March 19, 2020 to August 8, 2022.

Lung miRNA sequencing data were downloaded from the LUAD collection of the GDC portal (https://portal.gdc.cancer.gov/). By the attached clinical summary, 46 samples were selected from the sequencing data of 510 samples, classified as "adjacent healthy tissue". Normalization of sequencing libraries followed by removal of low-expressed fragments was performed using the edgeR-TMM algorithm [20]. Of the 316 types of microRNAs obtained, the most represented were selected, accounting for 95% of all microRNAs in lung tissue. The mean number of reads per million (CPM-scale) taken from previously selected 46 samples was considered as miRNA expression.

For each of the selected miRNAs and variants of SARS-CoV-2, regions of viral RNA were identified that are reverse complementary to the region from 2 to 7 nucleotides of the 5'-end of the mature microRNA (the so-called seed region). Following the widely accepted classification [21], such binding regions are called 6mer. The interaction of the corresponding miRNA with them leads to the suppression of translation and degradation of the target mRNA. To determine whether the 6mer belongs to the regions encoding proteins of the SARS-CoV-2 virus, the binding positions were pairwise aligned relative to the reference sequence of the Wuhan strain using the MAFFT algorithm [22]. After alignment, the coordinates of the beginning of the binding regions were brought to the coordinates on the RNA of the Wuhan strain.

The significance of differences was compared using the Mann–Whitney U test. Statistical analysis of the results was carried out in the R environment.

RESULTS

Among more than 129 million SARS-CoV-2 sequences, we selected 273 sequences reflecting the key strains and their variants circulating in Moscow and the Moscow region from the beginning of the pandemic until September 2022 (Fig. 1). The selected variants were chronologically divided into two groups: the "early strains" group included the Alpha (5 variants) and Delta (52 variants) strains circulating in the Moscow region from the beginning of the epidemic until January 2022; the "Omicron strains" group included 216 variants of the Omicron strain. It turned out that the obtained groups have a similar quality of RNA reading: the dominance hypothesis for the distributions of the shares of unrecognized nucleotides between the two groups was rejected (p_U=0.5).

Analysis of sequencing results of the lung tissue samples made it possible to identify more than 300 microRNA species. At the same time, only 32 accounted for more than 95% of all microRNA molecules in this tissue. Highly represented microRNAs and their proportion in the total pool of all microRNA molecules in the lung tissue are shown in Figure 2. In addition, for highly expressed microRNAs, regions of viral RNA were identified that are reverse complementary to their seed region.

To test the hypothesis about the evolutionary pressure of miRNAs characteristic of human lung tissue on the genome of the SARS-CoV-2 virus, the

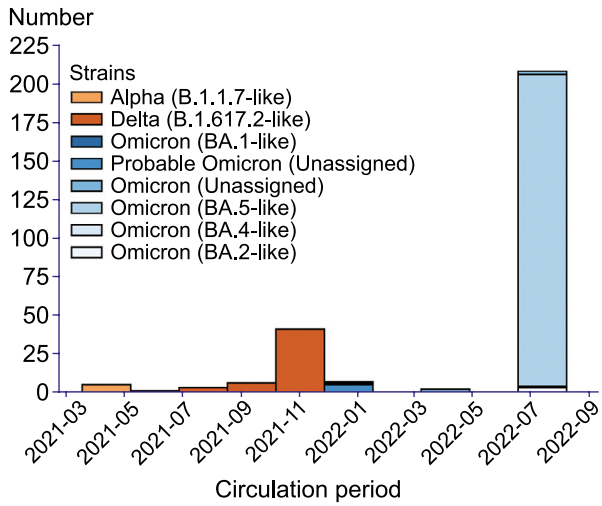

Fig. 1. Strains of the SARS-CoV-2 virus included in the study.

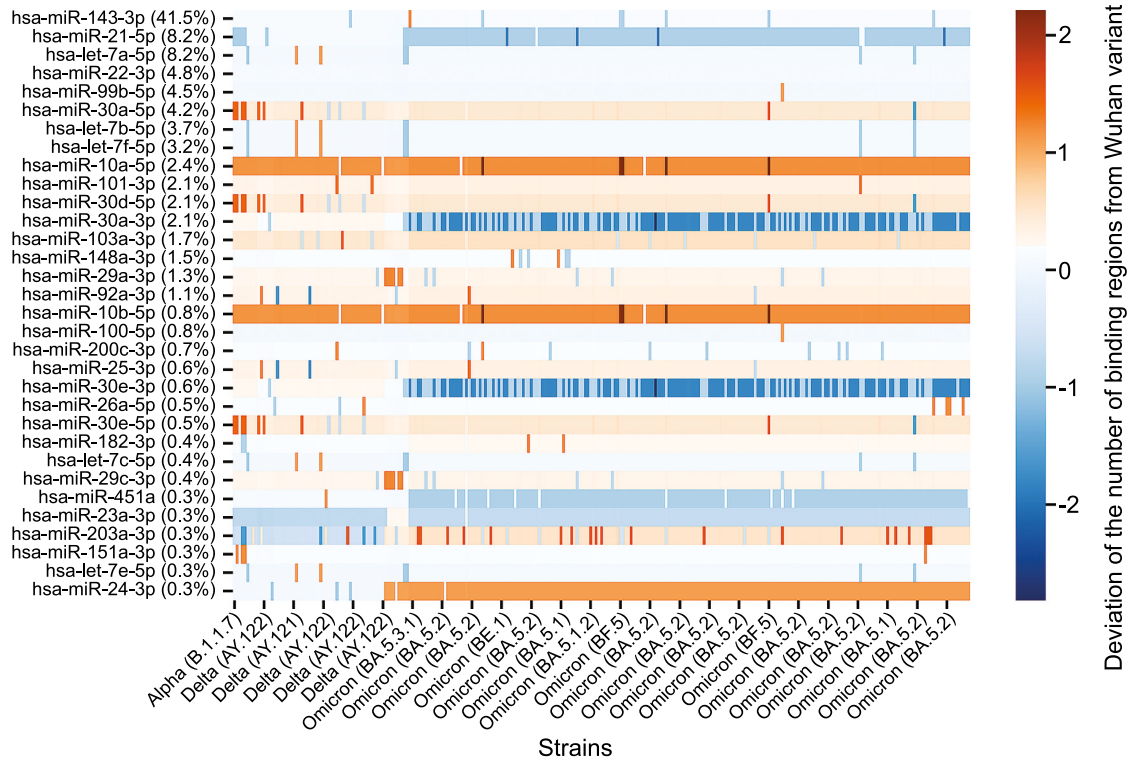

Fig. 2. Deviation of the number of binding regions for each highly expressed miRNA in each of the analyzed virus variants from the Wuhan strain per 30,331 nucleotides (Wuhan strain length). MicroRNAs are ordered by their expression in lung tissue. Next to the name of microRNAs, their share of the total number of microRNA molecules in the tissue is indicated.

change in the number of binding regions between comparison groups was assessed. First, the number of binding regions for each SARS-CoV-2 variant with highly expressed miRNAs was averaged, taking into account their presence in the lung tissue (Fig. 3). To eliminate the influence of viral RNA length, the resulting weighted average was normalized to the RNA length and adjusted to the length of the Wuhan strain (30,331 nucleotides). In the "early strains" group, the weighted mean is statistically significantly lower than in the "Omicron strains" group (p_U=0.0002). Thus, later mutations in SARS-CoV-2 RNA lead to a possible loss of microRNA regulatory activity.

Analysis of the contribution of individual miRNAs to this decrease (Fig. 2) showed that it is primarily associated with miR-21-5p, miR-30a/e-3p, and miR-451a. It should be noted that the effect of miR-24-3p increased in the Omicron strains group.

However, the question of which SARS-CoV-2 coding regions lose their binding positions remains open. Alignment of the analyzed virus strains to the Wuhan strain made it possible to compare the distribution of

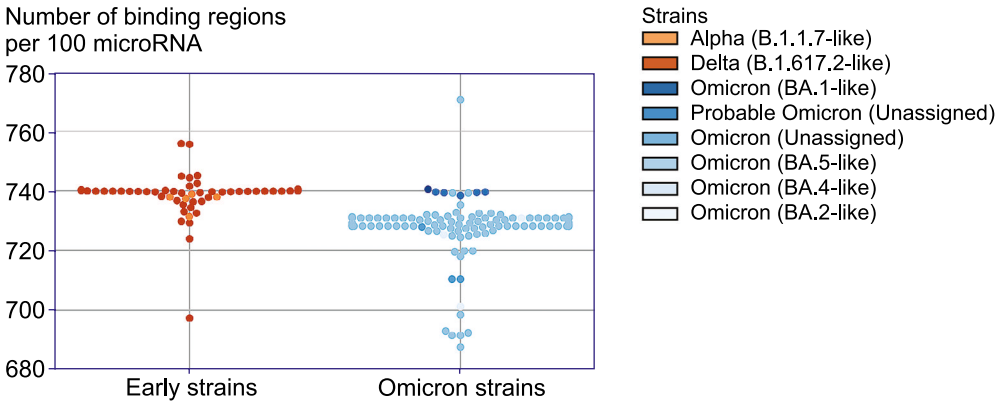

Fig. 3. Weighted mean (for miRNAs, considering their expression) of the number of binding regions for the groups "early strains" and "Omicron strains" per 30,331 nucleotides (Wuhan strain length).

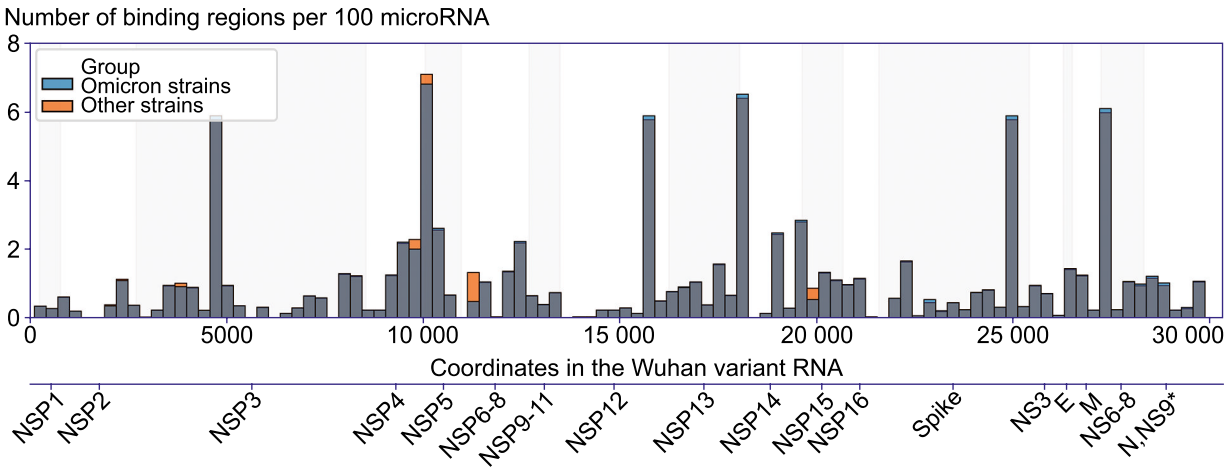

Fig. 4. Weighted mean (for miRNAs, considering their expression) of the number of binding regions for the groups "early strains" and "Omicron strains".

positions of the binding regions, considering miRNA expression. Each binding region associated with some microRNA contributed to the distribution equally to the expression of that microRNA. To compare the distributions of the binding regions between the "Omicron strains" and "early strains" groups, the distributions were averaged over all strains belonging to the corresponding group. This analysis made it possible to establish that the binding regions are unevenly distributed over the coding regions of RNA (Fig. 4).

More detailed analysis of significant changes in the regulatory contribution of miRNAs for each of the coding regions is presented in Table 1. In the 5p- and 3p-untranslated regions of the virus RNA, there are practically no sites for binding to microRNAs in the lung tissue. The vast majority of microRNA binding sites were located in four protein-coding regions of the virus: NSP3, NSP4, NSP12, and NSP14. All of them belong to ORF1ab. The *ORF1ab* gene encodes several nonstructural proteins responsible for further expression of structural and accessory proteins and virus replication [23]. The virus carries out its translation in the first place. As a result, a polypeptide is synthesized, which, after autoproteolysis, is cut into 16 individual proteins (NSP1-NSP16) [24]. Autoproteolytic activity is due to papain-like proteolytic domains in the multidomain protein NSP3 [24], which, according to our data, accounts for the largest microRNA-binding regions of the host cell. This may be due to the evolutionary adaptation of the virus to the microRNA environment and its use for epigenetic regulation of its genome.

Moreover, in the region encoding NSP15, NS8, and NSP6, the number of microRNA-binding regions is significantly reduced in the Omicron strains group ($p<0.01$). It was previously shown that the decrease

TABLE 1. Significant Differences in the Number of Binding Regions in the SARS-CoV-2 Genome with Lung Tissue-Associated microRNAs

Protein	Length, nucleotides	Number of binding regions per 100 miRNA molecules		
		early strains	Omicron strains	FDR
NSP15	1038	33.72	31.14	0.0000
NS8	366	13.79	14.07	0.0000
NSP6	870	18.60	11.38	0.0000
N	1260	19.54	19.60	0.0000
NSP2	1914	19.15	18.52	0.0000
NSP13	1803	40.21	38.63	0.0000
NSP3	5835	106.36	106.58	0.0037
NSP5	918	32.73	32.70	0.0198
Total virus genome	30331	723.88	711.80	0.0000

Note. FDR (false discovery rate): the significance of differences after adjusting for multiple comparisons.

in the number of such regions leads to a decrease in the virulence of the eastern equine encephalitis virus (EEEV) [9]. Moreover, the NS8 protein (ORF8) can suppress the maturation of class I MHC molecules and their translocation to the surface of the infected cell [25,26].

It can be concluded that the SARS-CoV-2 virus has practically no binding regions in 5p- and 3p-untranslated regions with miRNAs characteristic of the lung tissue. Nevertheless, the virus has many microRNA binding sites in the NSP3-NSP5 region responsible for the autoproteolysis of viral polypeptides and virion formation. In variants of strain Omicron, there was a significant decrease in the sites of binding to miRNAs of the host cells, which could contribute to a decrease in the virulence of this strain.

The study was supported by a grant from the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2021-1049).

REFERENCES

1. Ding SW, Voinnet O. Antiviral immunity directed by small RNAs. Cell. 2007;130(3):413-426. doi: 10.1016/j.cell.2007.07.039
2. Li Y, Lu J, Han Y, Fan X, Ding SW. RNA interference functions as an antiviral immunity mechanism in mammals. Science. 2013;342:231-234. doi: 10.1126/science.1241911
3. Siegal FP, Kadowaki N, Shodell M, Fitzgerald-Bocarsly PA, Shah K, Ho S, Antonenko S, Liu YJ. The nature of the principal type 1 interferon-producing cells in human blood. Science. 1999;284:1835-1837. doi: 10.1126/science.284.5421.1835
4. Turchinovich A, Tonevitsky AG, Cho WC, Burwinkel B. Check and mate to exosomal extracellular miRNA: new lesson from a new approach. Front. Mol. Biosci. 2015;2:11. doi: 10.3389/fmolb.2015.00011
5. Makarova J, Turchinovich A, Shkurnikov M, Tonevitsky A. Extracellular miRNAs and cell-cell communication: problems and prospects. Trends Biochem. Sci. 2021;46(8):640-651. doi: 10.1016/j.tibs.2021.01.007
6. Huang J, Wang F, Argyris E, Chen K, Liang Z, Tian H, Huang W, Squires K, Verlinghieri G, Zhang H. Cellular microRNAs contribute to HIV-1 latency in resting primary $CD4^+$ T lymphocytes. Nat. Med. 2007;13(10):1241-1247. doi: 10.1038/nm1639
7. Ingle H, Kumar S, Raut AA, Mishra A, Kulkarni DD, Kameyama T, Takaoka A, Akira S, Kumar H. The microRNA miR-485 targets host and influenza virus transcripts to regulate antiviral immunity and restrict viral replication. Sci. Signal. 2015;8(406):ra126. doi: 10.1126/scisignal.aab3183
8. Yekta S, Shih IH, Bartel DP. MicroRNA-directed cleavage of HOXB8 mRNA. Science. 2004;304:594-596. doi: 10.1126/science.1097434
9. Trobaugh DW, Gardner CL, Sun C, Haddow AD, Wang E, Chapnik E, Mildner A, Weaver SC, Ryman KD, Klimstra WB. RNA viruses can hijack vertebrate microRNAs to suppress innate immunity. Nature. 2014;506:245-248. doi: 10.1038/nature12869
10. Shimakami T, Yamane D, Jangra RK, Kempf BJ, Spaniel C, Barton DJ, Lemon SM. Stabilization of hepatitis C virus RNA by an Ago2-miR-122 complex. Proc. Natl Acad. Sci. USA. 2012;109(3):941-946. doi: 10.1073/pnas.1112263109
11. Scheel TK, Luna JM, Liniger M, Nishiuchi E, Rozen-Gagnon K, Shlomai A, Auray G, Gerber M, Fak J, Keller I, Bruggmann R, Darnell RB, Ruggli N, Rice CM. A broad RNA virus survey reveals both miRNA dependence and functional sequestration. Cell Host Microbe. 2016;19(3):409-423. doi: 10.1016/j.chom.2016.02.007
12. Saçar Demirci MD, Adan A. Computational analysis of microRNA-mediated interactions in SARS-CoV-2 infection. PeerJ. 2020;8:e9369. doi: 10.7717/peerj.9369
13. Khan MA, Sany MRU, Islam MS, Islam ABMMK. Epigenetic regulator miRNA pattern differences among SARS-CoV, SARS-CoV-2, and SARS-CoV-2 world-wide isolates delineated the mystery behind the epic pathogenicity and distinct clinical characteristics of pandemic COVID-19. Front. Genet. 2020;11:765. doi: 10.3389/fgene.2020.00765
14. Lukiw WJ. microRNA heterogeneity, innate-immune defense and the efficacy of SARS-CoV-2 infection — a commentary. Noncoding RNA. 2021;7(2):37. doi: 10.3390/ncrna7020037
15. Nersisyan S, Engibaryan N, Gorbonos A, Kirdey K, Makhonin A, Tonevitsky A. Potential role of cellular miRNAs in coronavirus-host interplay. PeerJ. 2020;8:e9994. doi: 10.7717/peerj.9994
16. Nersisyan S, Gorbonos A, Makhonin A, Zhiyanov A, Shkurnikov M, Tonevitsky A. isomiRTar: a comprehensive portal of pan-cancer 5'-isomiR targeting. PeerJ. 2022;10:e14205. doi: 10.7717/peerj.14205
17. Nersisyan S, Zhiyanov A, Shkurnikov M, Tonevitsky A. T-CoV: a comprehensive portal of HLA-peptide interactions affected by SARS-CoV-2 mutations. Nucleic Acids Res. 2022;50(D1):D883-D887. doi: 10.1093/nar/gkab701
18. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, Akite N, Ho J, Lee RT, Yeo W, Curation Team GC, Maurer-Stroh S. GISAID's Role in Pandemic Response. China CDC Wkly. 2021;3(49):1049-1051. doi: 10.46234/ccdcw2021.255
19. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, Colquhoun R, Ruis C, Abu-Dahab K, Taylor B, Yeats C, du Plessis L, Maloney D, Medd N, Attwood SW, Aanensen DM, Holmes EC, Pybus OG, Rambaut A. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol. 2021;7(2):veab064. doi: 10.1093/ve/veab064
20. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-140. doi: 10.1093/bioinformatics/btp616
21. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell. 2007;27(1):91-105. doi: 10.1016/j.molcel.2007.06.017
22. 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

23. Badua CLDC, Baldo KAT, Medina PMB. Genomic and proteomic mutation landscapes of SARS-CoV-2. J. Med. Virol. 2021;93(3):1702-1721. doi: 10.1002/jmv.26548
24. Hartenian E, Nandakumar D, Lari A, Ly M, Tucker JM, Glaunsinger BA. The molecular virology of coronaviruses. J. Biol. Chem. 2020;295(37):12910-12934. doi: 10.1074/jbc.REV120.013930
25. Zhang Y, Chen Y, Li Y, Huang F, Luo B, Yuan Y, Xia B, Ma X, Yang T, Yu F, Liu J, Liu B, Song Z, Chen J, Yan S, Wu L, Pan T, Zhang X, Li R, Huang W, He X, Xiao F, Zhang J, Zhang H. The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-I. Proc. Natl Acad. Sci. USA. 2021;118(23):e2024202118. doi: 10.1073/pnas.2024202118
26. Matsuoka K, Imahashi N, Ohno M, Ode H, Nakata Y, Kubota M, Sugimoto A, Imahashi M, Yokomaku Y, Iwatani Y. SARS-CoV-2 accessory protein ORF8 is secreted extracellularly as a glycoprotein homodimer. J. Biol. Chem. 2022;298(3):101724. doi: 10.1016/j.jbc.2022.101724