SARS-CoV-2 mutations altering regulatory properties: deciphering host's and virus's perspectives

Abul Bashar Mir Md. Khademul Islam^{1,*} and Md. Abdullah-Al-Kamran Khan²

¹Department of Genetic Engineering & Biotechnology, University of Dhaka, Dhaka, Bangladesh.

²Department of Mathematics and Natural Sciences, BRAC University, Dhaka, Bangladesh.

*** Correspondence:**

Dr. Abul Bashar Mir Md. Khademul Islam
Associate Professor
Department of Genetic Engineering and Biotechnology
University of Dhaka
Dhaka 1000, Bangladesh.
Email: khademul@du.ac.bd

Keywords: SARS-CoV-2, COVID-19, miRNA-microRNA, Viral SNPs, Mutations, IRES

28 **Abstract**

29 Since the first recorded case of the SARS-CoV-2, it has acquired several mutations in its
30 genome while spreading throughout the globe. However, apart from some changes in protein
31 coding, functional importance of these mutations in disease pathophysiology are still largely
32 unknown. In this study, we investigated the significance of these mutations both from the
33 host's and virus's perspective by analyzing the host miRNA binding and virus's internal
34 ribosome entry site (IRES), respectively. Strikingly, we observed that due to the acquired
35 mutations, host miRNAs bind differently compared to the reference; where few of the
36 miRNAs lost and few gained the binding affinity for targeting the viral genome. Moreover,
37 functional enrichment analysis suggests that targets of both of these gained and lost miRNAs
38 might be involved in various host immune signaling pathways. Also, we sought to shed some
39 insights on the impacts of mutations on the IRES structure of SARS-CoV-2. Remarkably, we
40 detected that three particular mutations in the IRES can disrupt its secondary structure which
41 can further make the virus less functional. These results could be valuable in exploring the
42 functional importance of the mutations of SARS-CoV-2 and could provide novel insights into
43 the differences observed different parts of the world.

44

45

46

47

48

49

50

51

52

53

54

55

56 **Introduction**

57 The recent COVID-19 pandemic has posed a serious threat to the public health sector around
58 the whole globe as it has already spread to 213 countries/territories. Approximately, 7.8
59 million people got infected with SARS-CoV-2 and half a million suffered death from this till
60 the date of writing this article [1]. The woes of this pandemic are still on the rise as the
61 detailed molecular aspects of the pathogenesis are still intangible.

62 Since the reporting of the first genome sequence of SARS-CoV-2, a huge number of genomes
63 are being sequenced everyday. Comparing to the reference, genome sequences from different
64 parts of the globe have suggested that this virus might have been acquiring many sequence
65 alterations while spreading. But the functional significance of these variations are largely
66 unknown and are not well-correlated with the disease pathobiology or infectivity.

67 MicroRNAs are small non-coding RNA (ncRNAs) which can play important roles during a
68 viral infection. miRNAs can provide benefit for the host by exerting antiviral mechanisms
69 through the innate and adaptive immune systems [2]; on the other hand, they can also
70 facilitate the viral propagation by negatively modulating the host immune responses [3]. It
71 was also reported that mutations in the miRNA binding sites in the viral genome can expedite
72 the viral proliferation during an infection [2].

73 Ribosomal frame-shifting using the internal ribosome entry site (IRES) has previously been
74 reported as important for coronaviruses as they use this site for synthesizing their proteins
75 [4]. Moreover, alterations in the IRES can result in lethal effects on the SARS-CoV
76 propagation and life-cycle [5]. Kelly et al. suggested that SARS-CoV-2 possess similar IRES
77 structure like SARS-CoV [6].

78 In this present study, we sought to find out the comparative effects of the SARS-CoV-2
79 genome mutations on the basis of the host miRNA targeting profiles and IRES alterations.
80 We illuminated the putative alterations of the host-miRNA binding profiles that might have
81 arose due to the acquired mutations in different world-wide isolates. Also, the probable
82 structural alterations due to the variations in SARS-CoV-2 IRES were investigated.

83 **Materials and Methods**

84 **Mining of SNPs and their world-wide frequencies**

We extracted the SNPs of SARS-CoV-2 which have world-wide frequency ≥ 10 for the miRNA binding analysis; while for analyzing the effects on IRES structure, we took all SNPs found in the IRES region of SARS-CoV-2. We obtained these information from “The 2019 Novel Coronavirus Resource” database along with the SNP annotations as on 30th April, 2020 which included information from 8614 “High quality” (As set by GISAID database: Complete and high coverage sequences only with $< 1\%$ Ns, $< 0.05\%$ unique amino acid mutations, and no insertion or deletion unless confirmed by submitter) SARS-CoV-2 genomes [7].

Viral RNA-host miRNA interaction analysis

We have only taken fragmented viral sequences around our targeted viral reference sequence and SNPs (20 nts upstream and 20 nts downstream of a SNP) for the interaction analysis. Then we used three different tools, namely- RNAhybrid [8], miRanda [9], IntaRNA [10] for analyzing the miRNA-RNA interaction. We considered those as high confidence interaction when it has been predicted by every tools that we used and have values- (i) for RNAhybrid: $MFE \geq -35$ KJ/mol and $p\text{-value} < .001$, (ii) for miRanda: energy ≥ -15 KJ/mol, (iii) for IntaRNA: $\Delta\Delta G \leq -15$ KJ/mol. We compiled the commonly predicted miRNAs from these three tools using the respective abovementioned cut-offs to reduce the false positives predictions.

Retrieval of country-wise COVID-19 statistics

COVID-19 related statistics (Total deaths, Deaths/1 Million population, Case fatality rate) from different countries were extracted from the Worldometer website [1] on 24th April, 2020. We presented the country-wise COVID-19 fatality statistics along with the miRNA binding profiles around the targeted SNPs of the associated country.

Target genes functional enrichment analysis

We performed the functional enrichment of the experimentally validated targets (obtained from mirTarBase database [11]) of miRNAs that can bind around the SNPs of SARS-CoV-2 using Gtools v1.8.4 [12] utilizing KEGG and Bioplane pathway modules. Resulting p-values were adjusted for multiple testing using the Benjamin and Hochberg's method of False Discovery Rate (FDR).

Prediction of the secondary structures of the IRES

The coordinates of the experimentally validated IRES of SARS-CoV-2 was extracted from previously conducted study [6] and adjusted accordingly. While predicting the secondary structures of the IRES incorporating the reference and the mutated bases, we took 50 additional nucleotides both upstream and downstream from the coordinates of the IRES. We used RNAfold tool [13] for the prediction of the IRES secondary structures.

Results and Discussion

We have obtained 377 high confidence (mutation frequency ≥ 10) SNPs from the database and used the regions around those for miRNA binding analysis (Data not shown). Not all SNPs could alter the miRNA binding affinity. Ten miRNAs were observed to bind both the targeted reference and mutated regions of the SARS-CoV-2 genome (Supplementary file 1). Five miRNAs were found to bind only the regions of the reference genome which were not found significant for the mutated sequences, where 5 different other miRNAs can bind to the regions around the observed mutations but not found in the reference sequences (Figure 1). We have considered the miRNAs as “Gained” ones which can bind the mutated sequences; and miRNAs which were not found to bind the mutated sequences but can bind the reference are termed as “Lost” miRNAs. Most of the miRNAs binding were observed for the targeted regions of the N and ORF1ab gene sequence. Also, miRNAs were also found to bind the targeted regions of ORF3a, ORF7a, and 3’UTR (Figure 1).

We observed that the countries in which these mutations were much prevalent (Supplementary file 2), a less number of miRNAs can bind to the SARS-CoV-2 genome/transcripts due to the acquired mutations, whereas these miRNAs can bind to the associated reference sequences (Figure 1). It was also detected that countries having high fatalities from COVID-19 have more mutations, and maybe these mutations can resist the binding of host miRNAs and giving those viruses with the mutations a competitive edge over the reference virus. Still, more experimental evidences should be searched in order to find out the definitive roles of these mutations in escaping from the host antiviral miRNAs.

Moreover, the functional enrichment analysis using the targets of these found miRNAs suggested that due to the loss of miRNA binding can impact the host adaptive and innate immune signaling pathways, hypoxia response etc. (Figure 2A-B); whereas targets of the gained miRNAs for the mutations are also involved in pathways like- interleukin signaling,

autophagy etc. (Figure 2A-B). So, the acquired mutations can lead to the alterations in miRNAs which can in turn result in the altered host immune responses during the viral infection.

We scrutinized 19 mutations in the IRES region of SARS-CoV-2 which was previously defined by Kelly et al. [6]. We aimed to illustrate the impact of the mutations in the secondary structure of the IRES of SARS-CoV-2 and predicted the secondary structures using RNAfold tool [13]. Interestingly, we observed that among the 19 mutations, 3 can significantly alter the secondary structure of the IRES of SARS-CoV-2 (Figure 2C). These 3 mutations in the IRES region are 13487: C>T, 13506: C>T, and 13507: G>A (Supplementary file 3). Mutations in these positions can destroy the viral translation process and render the virus less virulent. As, the studies regarding the IRES alterations are still in infancy, so more targeted researches should be conducted to elucidate the functional importance of these variations.

In our current analysis, we focused on the functional significance of the mutations reported for SARS-CoV-2 on the basis of host miRNA binding and the putative alterations in the IRES structure. Our preliminary results can suggest that these mutations can alter the binding pattern of host miRNAs which can ultimately result in alternative responses from these binding. Also, we revealed that several mutations in the IRES can disrupt its secondary structure which might suggest why the virus is affecting less in some countries compared to the others. Taking these observations in count, further studies should be conducted to completely understand the functional effects of the acquired mutations in the SARS-CoV-2 genome, and integration of other clinical information with these results can provide more insights on the dynamics of the SARS-CoV-2 infection.

Conflict of Interest

The authors declare no conflict of interest.

Author's contribution

ABMMKI conceived the project, designed the workflow and performed the analyses. Both authors wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We would also like to thank all the authors who have kindly deposited and shared genome data on GISAID (<https://www.gisaid.org/>).

Funding

This project was not associated with any internal or external source of funding.

Data Availability Statement

Publicly available data were utilized. Analyses generated data are deposited as supplementary files.

References

1. Worldometer, *Coronavirus Cases*. 2020. p. 1-22.
2. Trobaugh, D.W. and W.B. Klimstra, *MicroRNA Regulation of RNA Virus Replication and Pathogenesis*. Trends in molecular medicine, 2017. **23**(1): p. 80-93.
3. Głobińska, A., M. Pawełczyk, and M.L. Kowalski, *MicroRNAs and the immune response to respiratory virus infections*. Expert review of clinical immunology, 2014. **10**(7): p. 963-971.
4. Dinman, J.D., *Mechanisms and implications of programmed translational frameshifting*. Wiley interdisciplinary reviews. RNA, 2012. **3**(5): p. 661-673.
5. Plant, E.P., et al., *Altering SARS coronavirus frameshift efficiency affects genomic and subgenomic RNA production*. Viruses, 2013. **5**(1): p. 279-294.
6. Kelly, J.A. and J.D. Dinman, *Structural and functional conservation of the programmed -1 ribosomal frameshift signal of SARS-CoV-2*. bioRxiv, 2020: p. 2020.03.13.991083.
7. Zhao, W.M., et al., *The 2019 novel coronavirus resource*. Yi Chuan, 2020. **42**(2): p. 212-221.
8. Krüger, J. and M. Rehmsmeier, *RNAhybrid: microRNA target prediction easy, fast and flexible*. Nucleic Acids Research, 2006. **34**(suppl_2): p. W451-W454.
9. Betel, D., et al., *The microRNA.org resource: targets and expression*. Nucleic Acids Research, 2008. **36**(suppl_1): p. D149-D153.
10. Mann, M., P.R. Wright, and R. Backofen, *IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions*. Nucleic acids research, 2017. **45**(W1): p. W435-W439.
11. Huang, H.-Y., et al., *miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database*. Nucleic Acids Research, 2019. **48**(D1): p. D148-D154.
12. Perez-Llamas, C. and N. Lopez-Bigas, *Gitools: analysis and visualisation of genomic data using interactive heat-maps*. PloS one, 2011. **6**(5).
13. Gruber, A.R., et al., *The Vienna RNA websuite*. Nucleic acids research, 2008. **36**(Web Server issue): p. W70-W74.

Figure legends

Figure 1: Country-wise distribution of the analyzed SNPs and the predicted miRNA bind profile around the regions of these SNPs. Country-wise deaths/1M population, CFR, and total deaths are also provided in color coded heatmaps.

Figure 2: Functional enrichment analysis using the targets of host miRNA which can bind differentially around the mutated regions of SARS-CoV-2 utilizing **A.** KEGG pathway and, **B.** Bioplanet pathway modules. **C.** Secondary structures of the IRES of SARS-CoV-2 for the reference and mutated sequences. Significance of enrichment in terms of adjusted p-value (< 0.05) is represented in color coded P-value scale for all heatmaps. Color towards red indicates higher significance and color towards yellow indicates less significance, while grey means non-significant. Only selected significant enriched terms are shown.

List of supplementary files

Supplementary file 1: SNPs, their frequencies and the miRNAs which can bind around these SNPs.

Supplementary file 2: Frequencies of the SNPs along with their associated countries' fatality statistics.

Supplementary file 3: List of mutations found in the IRES of SARS-CoV-2 and their frequencies.

Lost	Gained	
h3a-miR-7109-5p		Adherens junction cell adhesion
h3a-miR-9718		Myc repressed pathway
h3a-miR-4722-5p		p53 signaling pathway
h3a-miR-6775-5p		Signal transduction
h3a-miR-4741		Focal adhesion
h3a-miR-6758a-3p		FRA pathway
h3a-miR-3194-5p		Intronless pre-mRNA processing
h3a-miR-3154		AP-1 transcription factor network
h3a-miR-1183		Immune system
h3a-miR-3916		Interleukin-2 signaling pathway
		Transmembrane transport of small molecules
		TGF-beta regulation of extracellular matrix
		Adaptive immune system
		FSH regulation of apoptosis
		Insulin signaling pathway
		Downstream signaling events Of B cell receptor (BCR)
		EGF/EGFR signaling pathway
		Protein metabolism
		Cell cycle
		Lipid and lipoprotein metabolism
		Endocytosis
		Post-translational protein modification
		TNF-alpha signaling pathway
		p53 activity regulation
		Phagosome
		Messenger RNA processing
		MicroRNA regulation of DNA damage response
		Innate immune system
		mTOR signaling pathway
		PDGFR signaling pathway
		Actin cytoskeleton regulation
		SLC-mediated transmembrane transport
		Signaling by the B cell receptor (BCR)
		Apoptosis
		Bone remodeling
		Clathrin derived vesicle budding
		Post-transcriptional silencing by small RNAs
		Neuronal system
		MHC class II antigen presentation
		Activation of NF-kappaB in B cells
		Small interfering RNA (siRNA) biogenesis
		Tumor necrosis factor (TNF) pathway
		Signaling by TGF-beta receptor complex
		ERBB1 downstream pathway
		Epidermal growth factor receptor (EGFR) pathway
		MAPK signaling pathway
		Fas signaling pathway
		Interferon signaling

Lost	Gained	
hsa-miR-7109-5p		TGF-beta signaling pathway
hsa-miR-9718		NOF signaling via TRKA from the plasma membrane
hsa-miR-4722-5p		PI3K/AKT activation
hsa-miR-6775-5p		Leptin signaling pathway
hsa-miR-4741		Signaling by NGF
hsa-miR-6768a-3p		Axon guidance
hsa-miR-3194-5p		T cell receptor signaling pathway
hsa-miR-3154		Signaling by FGFR in disease
hsa-miR-1183		Interleukin-4 signaling pathway
hsa-miR-3916		Senescence and autophagy
		Interleukin-5 signaling pathway
		Alpha-6 beta-4 integrin signaling pathway
		Signaling events regulated by Ret tyrosine kinase
		Angiopoietin receptor Tie2-mediated signaling
		MAPK cascade role in angiogenesis
		T cell receptor/Ras pathway
		RhoA signaling pathway
		p38 gamma/delta MAPK pathway
		Integrin signaling pathway
		Zinc influx into cells by the SLC39 gene family
		Zinc transporters
		Fatty acid metabolism
		Signaling events mediated by HDAC class I
		Toll-like receptor signaling pathway regulation
		p38 MK2 pathway
		Interleukin-4 regulation of apoptosis
		Adipogenesis
		PI3K events in ERBB2 signaling
		CD28 co-stimulation
		Jak-STAT signaling pathway
		E-cadherin nascent AJ-like junctions pathway
		Heart development
		Interleukin-6 signaling pathway
		Signaling by SCF-KIT
		Interleukin-10 anti-inflammatory signaling pathway
		Oxidative stress-induced gene expression via Nrf2
		T cell signal transduction
		TPQ signaling pathway
		Nitric oxide metabolism
		Protein processing in the endoplasmic reticulum
		ERK1/ERK2 MAPK pathway
		Signaling by Hippo
		Myc active pathway
		Signaling by constitutively active EGFR
		Lysosome vesicle biogenesis
		CXCR4 signaling pathway
		Gap junction pathway
		ERBB signaling pathway
		HIV life cycle early phase

Lost	Gained	
hsa-miR-7109-5p		Metabolic pathways
hsa-miR-4718		Focal adhesion
hsa-miR-4722-5p		Adherens junction
hsa-miR-6775-5p		Endocrine resistance
hsa-miR-4741		FoxO signaling pathway
hsa-miR-6768a-3p		HEF-1 signaling pathway
hsa-miR-3104-5p		Wnt signaling pathway
hsa-miR-3154		Endocytosis
hsa-miR-1183		p53 signaling pathway
hsa-miR-3316		Phagosome
		Cell cycle
		TNF signaling pathway
		Hippo signaling pathway
		Biosynthesis of unsaturated fatty acids
		Apoptosis
		Pyrimidine metabolism
		NOD-like receptor signaling pathway
		SNARE interactions in vesicular transport
		AMPK signaling pathway
		NF-kappa B signaling pathway
		IGF-1-like receptor signaling pathway
		Chenokine signaling pathway
		Hippo signaling pathway - multiple species
		Non-alcoholic fatty liver disease (NAFLD)
		Biosynthesis of amino acids
		Apoptosis - multiple species
		Lysosome
		Purine metabolism
		Alanine, aspartate and glutamate metabolism
		Glycine, serine and threonine metabolism
		Glyoxylate and dicarboxylate metabolism
		Sulfur metabolism
		Cysteine and methionine metabolism
		Complement and coagulation cascades
		Regulation of actin cytoskeleton

P-value for enrichment

0.0 0.05 1.0

Significant Non-Significant

Lost	Gained	
hsa-miR-7109-5p		Tight junction
hsa-miR-9718		mTOR signaling pathway
hsa-miR-4722-5p		Ras signaling pathway
hsa-miR-6775-5p		Rap1 signaling pathway
hsa-miR-4741		ABC transporters
hsa-miR-6769a-3p		Type II diabetes mellitus
hsa-miR-3194-5p		Oxytocin signaling pathway
hsa-miR-3154		RNA degradation
hsa-miR-1183		RNA transport
hsa-miR-3016		Fatty acid degradation
		T cell receptor signaling pathway
		Th17 cell differentiation
		Jak-STAT signaling pathway
		Osteoclast differentiation
		Th1 and Th2 cell differentiation
		PI3K-Akt signaling pathway
		MAPK signaling pathway
		Protein processing in endoplasmic reticulum
		mRNA surveillance pathway
		Neurotrophin signaling pathway
		Prolactin signaling pathway
		Insulin signaling pathway
		Longevity regulating pathway - multiple species
		EGFR tyrosine kinase inhibitor resistance
		MicroRNAs in cancer
		N-Glycan biosynthesis
		Chronic myeloid leukemia
		Peroxisome
		Fc gamma R-mediated phagocytosis
		Insulin resistance
		Autophagy - animal
		Cell adhesion molecules (CAMs)
		Phospholipase D signaling pathway
		Ether lipid metabolism

The diagram illustrates the complex secondary structure of the 16S rRNA molecule. It features several prominent stem-loops and helices, with nucleotides represented by their base letters (A, U, G, C). The structure is highly branched and folded, characteristic of large ribosomal RNA molecules.

IRES 10