Human miRNAs to Identify Potential Regions of SARS-CoV-2

Nimisha Ghosh,*,⊥ Indrajit Saha,*,⊥ Nikhil Sharma,*,⊥ and Jnanendra Prasad Sarkar*

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: It is two years now but the world is still struggling against COVID-19 due to the havoc created by the SARS-CoV-2 virus and its multiple variants. Considering this perspective, in this work, we have hypothesized a new approach in order to identify potential regions in SARS-CoV-2 similar to the human miRNAs. Thus, they may have similar consequences as caused by the human miRNAs in human body. Therefore, the same way by which human miRNAs are inhibited can be applied for such potential regions of virus as well by administering drugs to the interacting human proteins. In this regard, the multiple sequence alignment technique Clustal Omega is used to align 2656 human miRNAs with the SARS-CoV-2 reference genome to identify the potential regions within the virus reference genome which have high similarities with the human miRNAs. The potential regions in virus genome are identified based on the highest number of nucleotide match, greater than or equal to 5 at a genomic position, for the aligned miRNAs. As a result, 38 potential SARS-CoV-2 regions are identified consisting of 249 human miRNAs. Among these 38 potential regions, some top regions belong to nucleocapsid, RdRp, helicase, and ORF8. To understand the biological significance of these potential regions, the targets of the human miRNAs are considered for KEGG pathways and protein−protein and drug−protein interaction analysis as the human miRNAs are similar to the potential regions of SARS-CoV-2. Significant pathways are found which lead to comorbidities. Subsequently, drugs like emodin, bicalutamide, vorinostat, etc. are identified that may be used for clinical trials.
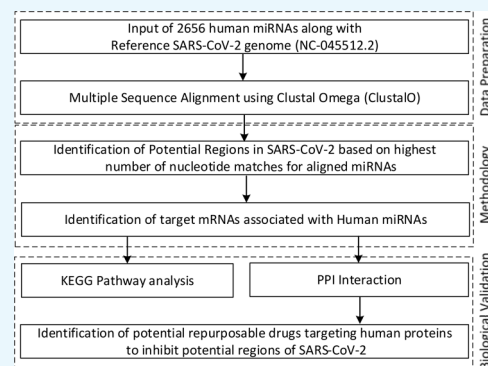
INTRODUCTION

Over the past two years, virus research has played a very important role in the scientific community due to the onset of the COVID-19 pandemic caused by SARS-CoV-2.[1,2] This contagious virus has flummoxed scientists around the world with its rapid mutation process and evolving biological consequences.[3−8] Till April 2022, more than 505 million cases[9] have been reported worldwide, while some countries are witnessing their second and third waves, the probable factor being mutated variants.[10,11] Thus, analysis of the virus and understanding its genetic characteristics are very important in order to combat the same. In this regard, study of the SARS-CoV-2 genome considering human miRNAs can be prove to be useful.

miRNAs are noncoding 20−23 nucleotide long RNA sequences that drives many different gene expressions in the cellular processes of living organisms.[12] The latter can be labeled as an important step in regulating the protein formation through binding to the mRNAs in the form of complementary sequences present majorly in the 3-untranslated region (3′-UTR), subsequently either making them untranslated or degrading them through RNA interference effector complex (RISC) resulting in a reduced number of transcripts. Moreover, involvement of miRNAs in different cellular processes covering development, differentiation, and proliferation are also well established within different studies,[13] whereas an altered expression of these miRNAs also leads to different forms of malignancies such as gastric cancer,[14] lung cancer,[15] and

leukemia.[16] Therefore, the role of miRNAs in the human body is quite complex. miRNAs also plays a vital role during a viral infection by modulating the cytokine response by either increasing the productive responses or lowering the damaging responses.[17] It has also been found that the virus modifies the host cell miRNAs to replicate within the host body.[18] Therefore, miRNAs can act as significant biomarkers and therapeutic targets. In 19, the authors have predicted miRNA sequences targeting the reference SARS-CoV-2 genome, which revealed an miRNA targeting region at 3822 bp ss-RNA of the spike glycoprotein of SARS-CoV-2. Natarelli et al.[20] have investigated some target motifs in the SARS-CoV-2 genome which are suitable for binding to human miRNAs and can be considered as the background for developing miRNA-based drugs against COVID-19. Many potential sets of miRNAs are also proposed by different studies. Elnabi et al.[21] have designed a synthetic miRNA complement to the SARS-CoV-2 virus at 3′-UTR, ORF9, and 5′-UTR. The main focus of the authors was to disrupt the interaction among the eukaryotic translation initiation factors that target mRNAs. Hence, miRNAs from
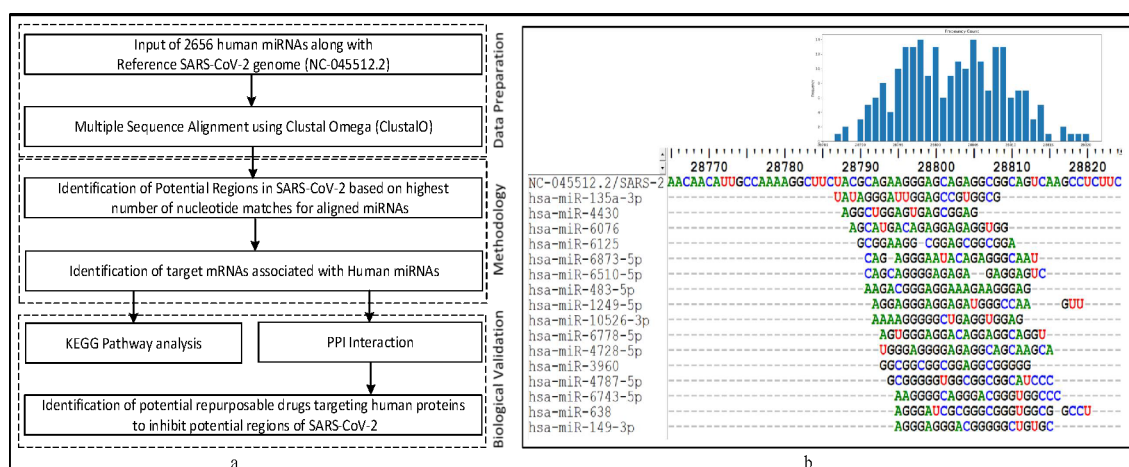
Figure 1. (a) Pipeline of the work. (b) Highest number of nucleotide matches for potential region PR1.

the 3′-UTR and 5′-UTR region can surpass the translation process. Moreover, SARS-CoV-2 can be inhibited through the miRNAs expressed in the host cells. Sardar et al.[22] have discovered a set of six host miRNAs involving miR-101, miR-126, miR-23b, miR-378, and miR-98 with a potential to reduce the effects of Nucleocapsid and Spike glycoprotein. Another study carried out in ref 23 with the help of RNA base-pairing resulted in miR-1307-3p and miR-3613-5p as potential miRNAs to hinder the virus replication. It is also a well established fact that the virus maintains its existence inside a host body by escaping the host immune system. Ying et al.[24] suggest the similarities between SARS-CoV-2 genome and 7 miRNAs (miR 8066, 5197, 3611, 3934-3p, 1307-3p, 3691-3p, 1468-5p) using the virus genome that can target the host genes to escape the host immune surveillance. Moreover, the authors identified that through targeting these miRNAs, SARS-CoV-2 can affect heart and brain development as well as insulin signaling.

Taking cues from the literature, we have hypothesized a new approach in order to identify potential regions in SARS-CoV-2 which are similar to the human miRNAs, thereby leading to similar consequences as caused by the human miRNAs in human body. Therefore, human miRNAs can be considered as an intermediary in order to identify the potential regions of SARS-CoV-2 that may interact with human proteins just like miRNA so that the comorbidity issues can be addressed and subsequently possible repurposable drugs can be identified. Thus, such potential regions of virus can be inhibited in the same way as human miRNAs by providing drugs to the interacting human proteins. In this regard, multiple sequence alignments of 2656 human miRNAs with the SARS-CoV-2 reference genome are performed using ClustalO to identify the potential regions within the virus reference genome which have high similarities with the human miRNAs. The potential regions in virus genome are identified for the aligned miRNAs based on the highest number of nucleotide matches, greater than or equal to 5 at a genomic position. This resulted in identification of 38 potential SARS-CoV-2 regions consisting of 249 human miRNAs. Among these 38 potential regions, some top regions belong to nucleocapsid, RdRp, helicase, and ORF8. Furthermore, the targets of the human miRNAs are considered for KEGG pathways, protein−protein, and drug−protein interaction analysis to understand the biological significance of these potential regions as the human miRNAs are similar to such potential regions of SARS-CoV-2. Consequently, significant

pathways are found which lead to comorbidities. Moreover, drugs like emodin, bicalutamide, vorinostat, etc. are identified which may be used for clinical trial.

■ MATERIAL AND METHODS

In this section, the details of data collection and data preparation are described, followed by a brief discussion on the pipeline of the proposed work.

Data Acquisition. The reference genome (NC_045512.2) of SARS-CoV-2 virus is collected from the National Center for Biotechnology Information (NCBI)[25] followed by the collection of 2656 human miRNAs[26] in fasta format. These human miRNAs are then aligned using Clustal Omega. Note that for the alignment of sequences, the high performance computing (HPC) facility of NITTTR, Kolkata, is used. The HPC cluster has a master node with dual Intel Xeon Gold 6130 Processor having 32 Cores, 2.10 GHz, 22 MB L3 Cache and 128 GB DDR4 RAM, and 2 GPU and 4 CPU computing nodes with a dual Intel Xeon Gold 6152 Processor having 44 Cores, 2.1 GHz, 30 MB L3 Cache, and 192 GB DDR4 RAM each, while GPU nodes have NVIDIA Tesla V100 GPU with 16 GB memory each.

Pipeline of the Work. The pipeline of the work is given in Figure 1a. Initially, 2656 human miRNAs are aligned with respect to the reference genome of SARS-CoV-2 (NC-045512.2) using the Clustal Omega (ClustalO)[27] alignment technique. Clustal Omega is the latest addition to the Clustal multiple sequence alignment family with increased scalability, facilitating thousands of sequences alignment quickly due to HMM probabilistic model while taking care of the evolutionary changes in a set of sequence through capturing position-specific patterns. In Clustal Omega the updated mBed is taken into account, with a complexity of $O(N \log N)$, where mBed refers to embedding layer of the "n" dimension representing each sequence, and "n" is proportional to "$\log N$". Hence, each n-dimensional vector represents each sequence. Each of these sequences can be clustered with the help of K-means and UPGMA methods. Thus, because of the advantage of aligning large sequences quickly by considering the evolutionary patterns, Clustal Omega is used for alignment in this work.

Once the alignment is performed for the 2656 human miRNAs with the SARS-CoV-2 reference genome, potential SARS-CoV-2 regions are identified within the virus reference genome which have high similarities with the human miRNAs. Considering the aligned miRNAs, such potential virus regions

Table 1. Top 10 Potential SARS-CoV-2 Regions Based on Highest Nucleotide Match of Aligned Human miRNAs

genomic regions of SARS-CoV-2	genomic coordinates and seqeunce of SARS-CoV-2	human miRNA	highest number of nucleotide match with the aligned miRNA	genomic coverage (%)	coding region
PR1	28787-5′-UACGCAGAAGGGAGCAGAGGCGGCAGUCAAGCCU-3′-28820	hsa-miR-3960, hsa-miR-10526-3p, hsa-miR-483-5p, hsa-miR-4430, hsa-miR-4787-5p, hsa-miR-6076, hsa-miR-149-3p, hsa-miR-4728-5p, hsa-miR-6778-5p, hsa-miR-135a-3p, hsa-miR-6743-5p, hsa-miR-6125, hsa-miR-6873-5p, hsa-miR-638, hsa-miR-6510-5p, hsa-miR-1249-5p	14	99.37	nucleocapsid
PR2	13587-5′-AAAAACUAAAUUGUUGUCGCUUCCA-3′-13610	hsa-miR-548d-5p, hsa-miR-548ay-5p, hsa-miR-548ag, hsa-miR-548ap-5p, hsa-miR-548ae-5p, hsa-miR-548ad-5p, hsa-miR-548o-5p, hsa-miR-548c-5p, hsa-miR-548aq-5p, hsa-miR-548am-5p	10	99.91	RdRp
PR3	17113-5′-AUUGGCCUAGCUCUCUACUACCCUUCUGCUCGCAUAG-3′-17149	hsa-miR-6511b-3p, hsa-miR-6511a-3p, hsa-miR-67500-3p, hsa-miR-99a-3p, hsa-miR-10398-5p, hsa-miR-4749-3p, hsa-miR-3162-3p, hsa-miR-4633-5p	8	99.26	helicase
PR4	28029-5′-UAUAUUAGAGUAGGAGCUAGAAAAAUCAGCACCU′-3′-28061	hsa-miR-8084, hsa-miR-451b, hsa-miR-3914, hsa-miR-378j, hsa-miR-378i, hsa-miR-378f, hsa-miR-378b, hsa-miR-3152-3p	8	84.67	ORF8
PR5	28356-5′-AGAAUGGAGAACGCAGUGGGGCGCGAUCA′-3′-28384	hsa-miR-6850-5p, hsa-miR-6090, hsa-miR-6089, hsa-miR-4665-5p, hsa-miR-4463, hsa-miR-3175, hsa-miR-197-5p	8	98.60	nucleocapsid
PR6	28889-5′-UCUCCCUGCUAGAAUGGCGUGGCAAUGGCGGUGAUGGCUG-3′-28925	hsa-miR-9851-5p, hsa-miR-5681a, hsa-miR-4446-3p, hsa-miR-1207-5p, hsa-miR-7705, hsa-miR-4469, hsa-miR-195-3p, hsa-miR-621	8	97.92	nucleocapsid
PR7	29184-5′-UUGCACAAUUUGCCCCAGCGCCUUCAGCGUUCUUCGGA-3′-29221	hsa-miR-4731-3p, hsa-miR-6848-3p, hsa-miR-1913, hsa-miR-2682-3p, hsa-miR-675-3p, hsa-miR-33a-3p, hsa-miR-5008-3p, hsa-miR-4707-5p, hsa-miR-10396b-3p	8	98.81	nucleocapsid
PR8	3453-5′-UUUACCUUAAACAUGGAGGAGGUGUUGCAGGAGCCUU-3′-3489	hsa-miR-4534, hsa-miR-4511, hsa-miR-877-5p, hsa-miR-4760-5p, hsa-miR-6754-5p, hsa-miR-4533, hsa-miR-7847-3p	7	99.44	NSP3
PR9	6340-5′-UGAUGUACUGAAGUCAGAGUCAGAGGGAAUGGAUAA-3′-6378	hsa-miR-6837-5p, hsa-miR-3198, hsa-miR-4695-5p, hsa-miR-8071, hsa-miR-12118, hsa-miR-6801-5p, hsa-miR-6789-5p	7	99.17	NSP3
PR10	11366-5′-UAUGAUGAUGGUGUGCUAGGAGGAGUGUGGACAC-3′-11396	hsa-miR-3945, hsa-miR-1272, hsa-miR-71110-5p, hsa-miR-6849-5p, hsa-miR-2392, hsa-miR-4721, hsa-miR-12121, hsa-miR-6765-5p	7	99.63	NSP6

Table 2. List of 38 Potential SARS-CoV-2 Regions Based on Aligned miRNAs

genomic regions of SARS-CoV-2	genomic coordinates and sequence of SARS-CoV-2	highest no. of nucleotide matches with the aligned miRNA	genomic coverage (%)	coded protein
PR1	28787-5′-UACGCAGAAGGGAGCAGAGGCGGCAGUCAAGCCU-3′-28820	14	99.37	nucleocapsid
PR2	13587-5′-AAAAACUAAUUGUUGUCGCUUCCA-3′-13610	10	99.91	RdRp
PR3	17113-5′-AUUGGCCUAGCUCUCUACUACCCUUCUGCUCGCAUAG-3′-17149	8	99.26	helicase
PR4	28029-5′-UAUAUUAGAGUAGGAGCUAGAAAAUCAGCACCU′-28061	8	84.67	ORF8
PR5	28356-5′-AGAAUGGAGAACGCAGUGGGGCGCGAUCA′-3′-28384	8	98.60	nucleocapsid
PR6	28889-5′-UCUCCUGCUAGAAUGGCUGGCAAUGGCGGUGAUGCUG-3′-28925	8	97.92	nucleocapsid
PR7	29184-5′-UUGCACAAUUUGCCCCCAGCGCUUCAGCGUUCUUCGGA-3′-29221	8	98.81	nucleocapsid
PR8	3453-5′-UUUACCUUAAACAUGGAGGAGGUGUUGCAGGAGCCUU-3′-3489	7	99.44	NSP3
PR9	6340-5′-UGAUGUACUGAAGUCAGAGGACGCGCAGGGAAUGGAUAA-3′-6378	7	99.17	NSP3
PR10	11366-5′-UAUGAUGAUGGUGCUAGGAGAGUGUGGACAC-3′-11396	7	99.63	NSP6
PR11	29436-5′-AACAGCAAACUGUGACUCUUCUUCCUGCUGCAGAUUU-3′-29472	7	98.21	nucleocapsid
PR12	350-5′-CGUGGCUUUGGAGACUCCGUGGAGGAGGUCUUAUCAGAGGC-3′-390	6	99.06	leader protein
PR13	1386-5′-ACAAUUCAGAAGUAGGACCUGAGCAUAGUCUU-3′-1417	6	99.26	NSP2
PR14	2081-5′-GUUCAGUUGACUUCGCAGUGGCUAACUAACAUCUUUGGCAC-3′-2121	6	97.48	NSP2
PR15	2720-5′-GCACCAACAAAGGUUACUUUUGGUGAU-3′-2746	6	99.75	NSP3
PR16	4266-5′-UAAAUGGUUACACUGUAGAGGAGGCAAAGACAGUGCUUAA-3′-4305	6	98.85	NSP3
PR17	5537-5′-CAGCAGACAACCCUUAAGGGUGUAGAAGCUGUUAUGUA-3′-5574	6	98.88	NSP3
PR18	8165-5′-AUUUCAGCAGCUCGGCAAGGGUUUGUUGAUUC-3′-8196	6	99.44	NSP3
PR19	10803-5′-UAGGACCUCUUUCUGCUCAAACUGGAAUUGCCGUUUUA-3′-10840	6	99.53	3CL-Pro
PR20	13410-5′-GUUGUGAUCAACUCCGCGAACCCAUGCUUCAGUCAGC-3′-13446	6	99.47	NSP10, RdRp
PR21	15090-5′-AAAGAAUAGAGCUCGCACCGUAGCUGGUGUCUCUAUCUG-3′-15128	6	98.32	RdRp
PR22	22009-5′-CAAAAGUUGGAUGGAAAGUGAGUUCAGAGUUUAUUCUA-3′-22046	6	98.25	endoRNase
PR23	22324-5′-UUCUUCAGGUUGGACAGCUGGUGCUGCAG-3′-22352	6	97.79	Spike
PR24	25329-5′-UUGAUGAAGACGACUCUGAGCCAGUGCUCAAAGGAGUCAAAUU-3′-25371	6	99.31	Spike
PR25	166-5′-AGUAACUCGUCUAUCUUCUGCAGGCU-3′-191	5	98.82	5′-UTR
PR26	825-5′-AUAACAACUUCUGUGGCCCUGAUGGCUACCCUCUUGAGUG-3′-864	5	98.58	NSP2
PR27	1659-5′-GUGACUUUAAACUUAAUGAAGAGAUCGCCAUUAUUUUGGCAUCUUUUUCU-3′-1708	5	98.90	NSP2
PR28	2511-5′-CAGAAGUGUUAACAGAGGAAGUUGUCUUG-3′-2539	5	99.47	NSP2
PR29	12785-5′-ACAACAAAGGGAGGUAGGUUUGUACUUG-3′-12812	5	99.10	NSP9
PR30	13385-5′-GGUAUGUGGAAAGGUUAUGGCUGUAGUUGUGAUC-3′-13418	5	99.56	NSP10
PR31	16212-5′-GUACACACCGCAUACAGUCUUACAGGCUGUUGGGGCUU-3′-16249	5	99.58	RdRp, helicase
PR32	17168-5′-AUGCCGCUGUUGAUGCACUAUGUGAGAAGGCAUUAAAAUAUUUGC-3′-17212	5	99.73	helicase
PR33	24690-5′-GUGGAAAGGGCUAUCAUCUUUAUGUCCUUCCCUCAGUCAGCACCUCAU-3′-24736	5	99.75	Spike
PR34	26793-5′-AUGUGGCUCAGCUACUUCAUUGCUUC-3′-26818	5	85.58	membrane
PR35	27500-5′-CUUCUGGAACAUACGAGGGCAAUUCACCAUUUCAU-3′-27534	5	98.96	ORF7a
PR36	28600-5′-UUUCUACUACCUAGGAACUGGGCCAGAAGCUGGACUUC-3′-28637	5	98.63	nucleocapsid
PR37	28915-5′-CGGUGAUGCUGCUCUUGCUUUGCUGCUGCUUGACAG-3′-28950	5	85.65	nucleocapsid
PR38	29534-5′-ACUCAUGCAGACCACACAAGGCAGAUGGGCUAU-3′-29566	5	97.20	nucleocapsid, ORF10

are identified based on the highest number of nucleotide matches. To avoid small length regions, nucleotides greater than or equal to 5 at a particular genomic position were studied. This means a minimum of 5 miRNAs are aligned. Moreover, to study the possible consequences of the identified potential regions, at most 10 target mRNAs with the highest scores associated with each human miRNA for each of the 38 potential regions are identified using the miRDB database.[28] Subsequently, these target mRNAs are considered to study the KEGG pathways using the EnrichR tool.[29−31,32] Furthermore, these targets are provided as inputs to the STRING database[33,34] to identify the protein−protein interactions. It should be noted that the STRING database returns human protein−protein interactions for those inputs and may include additional human proteins apart from the ones that are provided as inputs as well as exclude

some in the process. Finally, at most 10 human proteins based on the highest degree as derived from the protein−protein interactions are provided as input to the EnrichR tool to identify the potential repurposable drugs targeting the human proteins associated with the 38 potential SARS-CoV-2 regions as the miRNAs are aligned with them.

■ RESULTS

The results of this work are executed according to the pipeline as shown in Figure 1a. This study focuses on the identification of potential regions of SARS-CoV-2 which are highly similar to the human miRNAs and thereby can have similar consequences as caused by human miRNAs in a human body. Therefore, it can be hypothesized that such potential regions of SARS-CoV-2 can be inhibited in the same way as human miRNAs. In this regard,
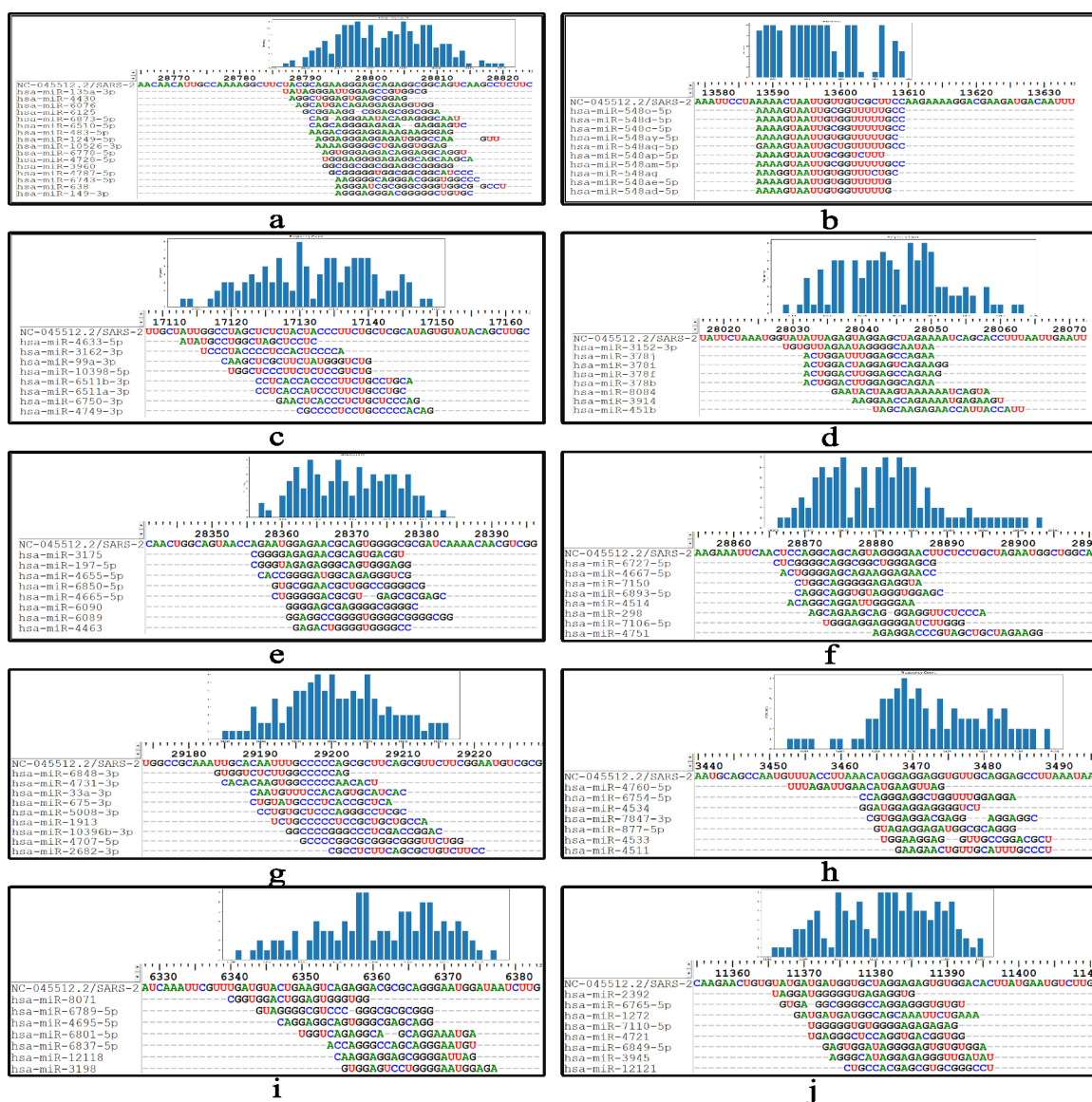
Figure 2. Top 10 (a−j) potential SARS-CoV-2 regions based on highest number of nucleotide matches.

2656 human miRNAs are aligned with the SARS-CoV-2 reference genome using ClustalO. These aligned sequences are provided in the Supporting Information. Subsequently, from these aligned human miRNAs, the potential regions in virus genome are identified based on the highest number of nucleotide match which should be greater than or equal to 5 at a genomic position. As a result, we have obtained 38 potential SARS-CoV-2 regions. Among these 38, on the basis of the highest number of nucleotide matches, the top 10 potential regions are reported in Table 1 while the list of all 38 regions is reported in Table 2. As can be seen from Table 1, the top 10 regions belong to nucleocapsid, RdRp, helicase, ORF8, NSP3, and NSP6 with the highest number of nucleotide matches of 14, 10, 8, and 7. The SARS-CoV-2 potential region PR1 with the highest number of nucleotide matches of 14 lies between the coordinates 28787 and 28820 (both inclusive), belongs to nucleocapsid, and is reported in Figure 1b and Figure 2a as well. Furthermore, the genomic coverage (%) of the potential regions is also reported in Tables 1 and 2. Genomic coverage refers to the presence of the potential regions in 10407 SARS-CoV-2 sequences; for example, the genomic coverage of PR1 is 99.37%.

The details of all 38 potential regions are given in Table S1. As can be seen from Figure 2a, the highest number of nucleotide matches, 14, is at two coordinates, 28798 and 28805. The rest of the figures in Figure 2 show the nine potential regions of SARS-CoV-2 with the highest number of nucleotide matches, while the rest are reported in Figure S1. This way, we are able to discover those potential virus regions with high similarity to human miRNAs; thus, the same methods of inhibition of human miRNAs can be applied on similar potential virus regions as well.

Furthermore, to understand the biological consequences of the miRNAs, at most 10 target human mRNAs based on highest scores are identified for each human miRNA corresponding to a SARS-CoV-2 potential region using miRDB.[35] For example, for potential region PR1, which corresponds to 16 human miRNAs, i.e., hsa-miR-3960, hsa-miR-10526-3p, hsa-miR-483-5p, hsa-miR-4430, hsa-miR-4787-5p, hsa-miR-6076, hsa-miR-149-3p, hsa-miR-4728-5p, hsa-miR-6778-5p, hsa-miR-135a-3p, hsa-miR-6743-5p, hsa-miR-6125, hsa-miR-6873-5p, hsa-miR-638, hsa-miR-6510-5p, and hsa-miR-1249-5p, 160 target human mRNAs are identified using miRDB. These 160 mRNAs are then provided as inputs to the EnrichR[32] tool for the KEGG

Table 3. List of Top 10 Key Human Proteins Associated with Each Potential Region of SARS-CoV-2 along with the Top 5 KEGG Pathways Based on the FDR-Corrected p-Values

genome regions in SARS-CoV-2	key proteins in PPI network	KEGG pathways	FDR corrected p-value
PR1	PIK3CA, TP53BP1, CHD3, ARID1A, SNW1	hepatocellular carcinoma	2.81×10^{-01}
	E2F3, SMARCD1, ARID2, AR, RELN	Type II diabetes mellitus	4.16×10^{-01}
		axon guidance	6.73×10^{-01}
		maturity onset diabetes of the young	6.81×10^{-01}
		ErbB signaling pathway	6.90×10^{-01}
PR2	PSMB1, PSMB6, PSMC4, PSMA2, PSMA4	Th17 cell differentiation	1.95×10^{-01}
	PSMA8, PPP3R1, UBE2E3	spinocerebellar ataxia	1.95×10^{-01}
		axon guidance	2.06×10^{-01}
		prion disease	2.06×10^{-01}
		circadian rhythm	2.06×10^{-01}
PR3	MAPK1, HDAC2, KMT2D, SESTD1, ZBTB10	thyroid hormone signaling pathway	1.24×10^{-01}
	DACT1, ERCC6L, BAZ2A, RRM2, S1PR1	sphingolipid signaling pathway	4.34×10^{-01}
		Type II diabetes mellitus	4.34×10^{-01}
		notch signaling pathway	4.34×10^{-01}
		MicroRNAs in cancer	4.34×10^{-01}
PR4	NOTCH1, PAX5, ATL1, CD40LG, CERKL	PPAR signaling pathway	4.92×10^{-01}
	DLG5, GTF2I, NCAPG, NKX3–1, RAB10	Fc gamma R-mediated phagocytosis	4.92×10^{-01}
		AMPK signaling pathway	4.92×10^{-01}
		breast cancer	4.92×10^{-01}
		cell adhesion molecules	4.92×10^{-01}
PR5	ACTB, AGO2, PRKACA, CSNK1A1, MECP2	gastric acid secretion	3.45×10^{-02}
	YWHAB, KLRD1, ADCY9, TUBB6, RUNX1	*Vibrio cholerae* infection	7.44×10^{-02}
		oxytocin signaling pathway	1.49×10^{-01}
		tight junction	1.49×10^{-01}
		gap junction	1.49×10^{-01}
PR6	BRCA1, CREBBP, CCNH, RPA3, HIF1A	homologous recombination	3.07×10^{-01}
	GTF2A1, UBQLN2, C1orf162, CBX5, RBBP5	basal transcription factors	3.07×10^{-01}
		nucleotide excision repair	3.07×10^{-01}
		Fanconi anemia pathway	3.07×10^{-01}
		notch signaling pathway	3.07×10^{-01}

genome regions in SARS-CoV-2	key proteins in PPI network	KEGG pathways	FDR corrected p-value
PR20	XPO1, MAPK1, CAND1, DCUN1D3, KLHL20	retinol metabolism	1.09×10^{-01}
	ADH1A, ADH1B, AXIN1, DDX3X, GGH	metabolism of xenobiotics by cytochrome P450	1.09×10^{-01}
		drug metabolism	1.83×10^{-01}
		tyrosine metabolism	1.83×10^{-01}
		fatty acid degradation	1.83×10^{-01}
PR21	PIK3R3, SIAH1, UBE2I, IGFBP5, INHBB	Fc gamma R-mediated phagocytosis	2.07×10^{-01}
	LRRC55, NOA1, OLIG1, EPHA3, PLCXD3	Ras signaling pathway	2.07×10^{-01}
		shigellosis	2.07×10^{-01}
		autophagy	2.07×10^{-01}
		ubiquitin-mediated proteolysis	2.07×10^{-01}
PR22	HDAC2, FBXL3, CBX5, RFX5, YOD1	chronic myeloid leukemia	1.09×10^{-01}
	TGS1, SPOPL, SH2B3, SCN1B, RUNX1	Th17 cell differentiation	1.46×10^{-01}
		cell adhesion molecules	1.72×10^{-01}
		human papillomavirus infection	1.72×10^{-01}
		PI3K-Akt signaling pathway	1.72×10^{-01}
PR23	IGDCC3, NLK, SDK2, WNT10B	selenocompound metabolism	5.30×10^{-01}
		terpenoid backbone biosynthesis	5.30×10^{-01}
		butanoate metabolism	5.30×10^{-01}
		Wnt signaling pathway	5.30×10^{-01}
		vascular smooth muscle contraction	5.50×10^{-01}
PR24	UBB, HIP1, DAB2, GNG13, BMP2	hepatocellular carcinoma	5.30×10^{-01}
	CBLB, HTR2C, INTS2, POLI, PRKCE	fatty acid elongation	2.52×10^{-01}
		Glycosaminoglycan biosynthesis	3.21×10^{-01}
		Cushing syndrome	3.21×10^{-01}
		basal cell carcinoma	3.27×10^{-01}
PR25	CREBBP, ARID1B, HDAC3, KDM6A, SMAD2	inflammatory mediator regulation of TRP channels	4.74×10^{-01}
	SMURF2, AGO3, CCDC135, CLTCL1, NDFIP1	hedgehog signaling pathway	2.64×10^{-02}
		TGF-beta signaling pathway	5.96×10^{-02}
		Hippo signaling pathway	1.86×10^{-01}
		adherens junction	2.31×10^{-01}

Table 3. continued

genome regions in SARS-CoV-2	key proteins in PPI network	KEGG pathways	FDR corrected p-value
PR7	CUL3, ASB13, LSM14A, KLHL42, UBE2B	oxytocin signaling pathway	5.83×10^{-01}
	CADM2, TIA1, COPS7B, HNRNPR, PRKCA	aldosterone synthesis and secretion	5.83×10^{-01}
		vascular smooth muscle contraction	5.83×10^{-01}
		spinocerebellar ataxia	5.83×10^{-01}
		notch signaling pathway	5.83×10^{-01}
PR8	TRIP12, YWHAE, FGB, RAB1A, KIF1A	mRNA surveillance pathway	2.86×10^{-01}
	GSPT1, PABPC3, SMG6, SYP, TOM1L2	gap junction	5.59×10^{-01}
		inflammatory mediator regulation of TRP channels	5.59×10^{-01}
		endocytosis	5.59×10^{-01}
		serotonergic synapse	5.59×10^{-01}
PR9	SNW1, GATA4, HDAC2, CDK8, CPPED1	endocytosis	1.26×10^{-01}
	PLEKHS1, PFKP, AGAP4, OLFML1, SOX10	thyroid hormone signaling pathway	3.01×10^{-01}
		notch signaling pathway	4.19×10^{-01}
		nonhomologous end-joining	5.73×10^{-01}
		TNF signaling pathway	5.73×10^{-01}
PR10	GNG13, PRKACA, CCR1, PDYN, KCNB1	Herpes simplex virus 1 infection	1.58×10^{-02}
	KBTBD2, TULP4, SYNGAP1, RNFT2, RNF144A	human cytomegalovirus infection	4.24×10^{-01}
		cocaine addiction	4.24×10^{-01}
		signaling pathways regulating pluripotency of stem cells	4.24×10^{-01}
		retrograde endocannabinoid signaling	4.24×10^{-01}
PR11	TNPO1, PSMD1, DKC1, GRIN3A, TAOK1	prion disease	6.40×10^{-02}
	STK4, RFC5, PPP3R1, OPRM1, NOTUM	MAPK signaling pathway	6.40×10^{-02}
		spinocerebellar ataxia	9.78×10^{-02}
		Wnt signaling pathway	1.25×10^{-01}
		glutamatergic synapse	2.46×10^{-01}
PR12	BCL3, HOXA7, LPCAT3, SH3GL1, AAK1	glycosaminoglycan degradation	5.16×10^{-01}
	AZIN1, BHLHE22, DUSP1, GART, HOXB7	one carbon pool by folate	5.16×10^{-01}
		terpenoid backbone biosynthesis	5.16×10^{-01}

genome regions in SARS-CoV-2	key proteins in PPI network	KEGG pathways	FDR corrected p-value
PR26	CHST12, EXOSC6, ANAPC2, CERK, EPHA7	hepatitis C	2.29×10^{-01}
	ERC1, GNE, HS6ST1, KBTBD13, LSM2	glycosaminoglycan biosynthesis	4.32×10^{-01}
		RNA degradation	4.32×10^{-01}
		NF-kappa B signaling pathway	4.32×10^{-01}
		ubiquitin mediated proteolysis	4.32×10^{-01}
PR27	CREBBP, CCNA2, SSB, OGT, SNRPC	protein export	4.32×10^{-01}
	UBE2D1, YBX1, CDK17, GFPT1, PLIN1	cell cycle	1.09×10^{-01}
		insulin resistance	3.60×10^{-01}
		hepatitis B	3.60×10^{-01}
		protein processing in endoplasmic reticulum	3.60×10^{-01}
PR28	DOCK5, MED12L, ARHGAP12, COPA, DYNC1L12	maturity onset diabetes of the young	3.60×10^{-01}
	FSD1L, MPZL3, PUM1, PURA, SEC62	salmonella infection	1.48×10^{-01}
		pantothenate and CoA biosynthesis	4.21×10^{-01}
		protein export	4.21×10^{-01}
		protein processing in endoplasmic reticulum	4.21×10^{-01}
PR29	MECP2, STRBP, ELAVL2, GNL3L, MAPK1	sphingolipid signaling pathway	4.47×10^{-01}
	ADAMTS5, RAP1A, PAPOLG, MTM1, MMP15	neurotrophin signaling pathway	2.89×10^{-02}
		renal cell carcinoma	4.56×10^{-02}
		ErbB signaling pathway	5.19×10^{-02}
		focal adhesion	5.19×10^{-02}
PR30	FOXP3, FOXP4, HDLBP, IREB2, LATS2	Ras signaling pathway	7.00×10^{-02}
	LSAMP, NEGR1, NFAT5, PAPOLG, RORA	cell adhesion molecules	1.27×10^{-01}
		axon guidance	1.27×10^{-01}
		inflammatory bowel disease	1.27×10^{-01}
		Th17 cell differentiation	2.43×10^{-01}
PR31	NR3C1, ARID1A, CWC25, JPH3, KLHL11	selenocompound metabolism	2.75×10^{-01}
	NRIP1, RBFOX1, RNF114, TAOK1, USP25	colorectal cancer	7.15×10^{-02}
		hepatocellular carcinoma	2.13×10^{-01}

Table 3. continued

genome regions in SARS-CoV-2	key proteins in PPI network	KEGG pathways	FDR corrected p-value
PR13	ELAVL2, NOVA1, INSR, NRXN1, NR1D2	DNA replication	5.16×10^{-01}
		ferroptosis	5.16×10^{-01}
		nonalcoholic fatty liver disease	1.09×10^{-01}
	NCOR2, ANKRD34A, SPEN, SOX2, SNX1	arginine and proline metabolism	3.18×10^{-01}
		Fanconi anemia pathway	3.18×10^{-01}
		protein processing in endoplasmic reticulum	3.18×10^{-01}
		diabetic cardiomyopathy	3.18×10^{-01}
PR14	AFF4, CCNT1, GCC2, GXYLT1, PDSS8	other types of O-glycan biosynthesis	1.51×10^{-01}
		small cell lung cancer	2.77×10^{-01}
	PDZRN4, POGLUT1, RGPD6	glycosaminoglycan degradation	3.90×10^{-01}
		circadian rhythm	3.90×10^{-01}
		cholesterol metabolism	3.90×10^{-01}
PR15	TNRC6B, CREBBP, CBFB, CPEB3, UBE2D1	renal cell carcinoma	5.96×10^{-02}
	CPEB4, ETS1, PUM2, PHC3, MTMR1	FoxO signaling pathway	1.73×10^{-01}
		Vibrio cholerae infection	1.73×10^{-01}
		human papillomavirus infection	1.73×10^{-01}
		notch signaling pathway	1.73×10^{-01}
PR16	ESR1, EGR1, MAPK1, ABLIM3, AFAP1	AGE-RAGE signaling pathway in diabetic complications	2.02×10^{-01}
	CPEB3, HS3ST5, RAD21, RC3H1, SEC24C	thyroid hormone signaling pathway	2.02×10^{-01}
		Prion disease	2.02×10^{-01}
		type II diabetes mellitus	2.02×10^{-01}
		retrograde endocannabinoid signaling	2.02×10^{-01}
PR17	NFIB, TMCC1, DNAJC5, FSTL5, KLF15	estrogen signaling pathway	4.55×10^{-01}
	MC2R, NEU3, NFIA, PLAG1, POMC	sphingolipid metabolism	4.55×10^{-01}
		cortisol synthesis and secretion	4.55×10^{-01}
		adipocytokine signaling pathway	4.55×10^{-01}
		human immunodeficiency virus 1 infection	4.55×10^{-01}
PR18	ZFX, FOXO1, RBBP6, PAK1, PACSIN2	longevity regulating pathway	3.91×10^{-01}
	VGLL3, TXLNG, TFCP2L1, SOX6, SEMA3A	insulin resistance	3.91×10^{-01}

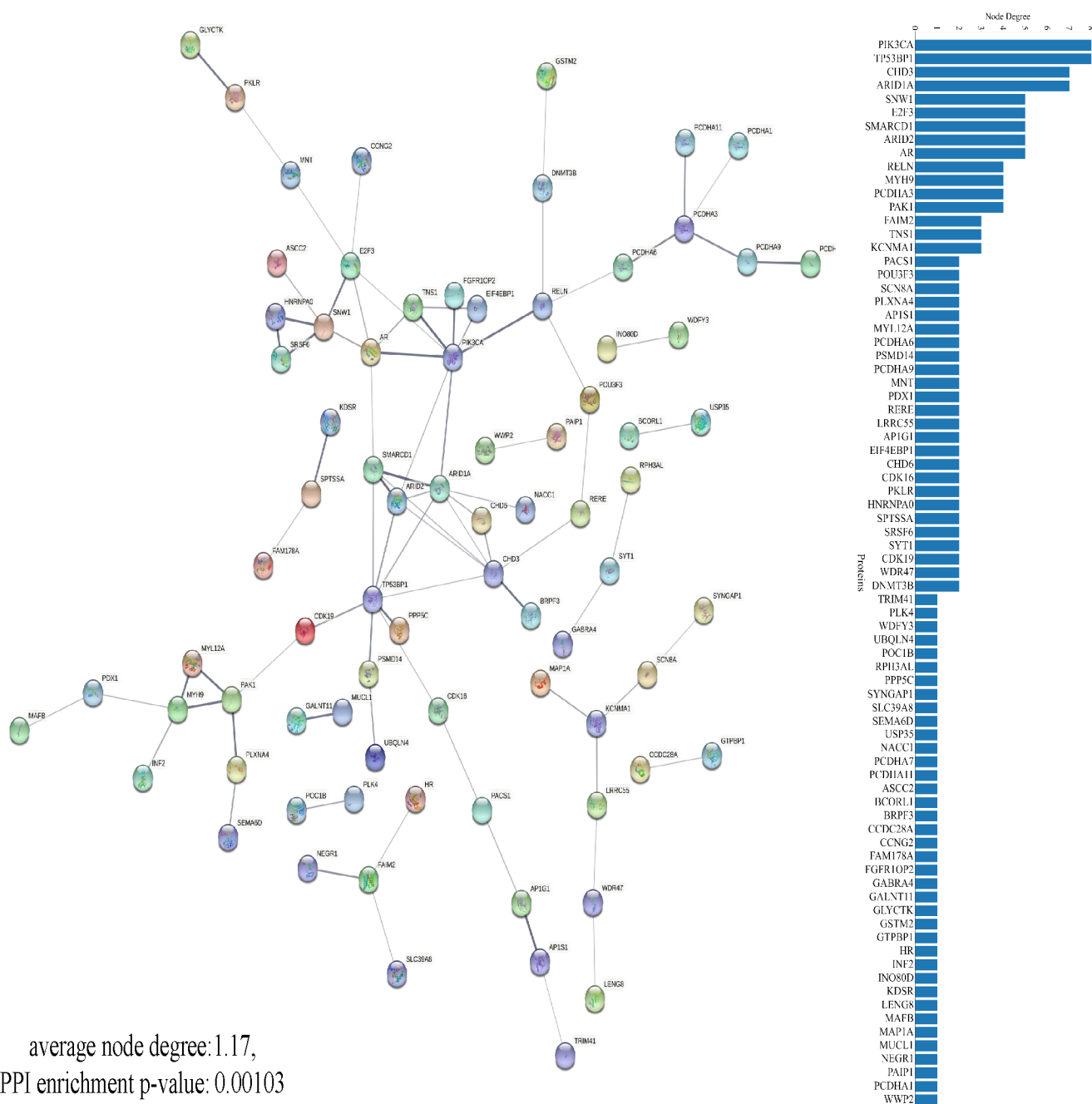| genome regions in SARS-CoV-2	key proteins in PPI network	KEGG pathways	FDR corrected p-value
		mitophagy	2.13×10^{-01}
		pancreatic cancer	2.13×10^{-01}
PR32	RHOA, GNG13, PRKCA, PLXNA4, TBC1D2B	FoxO signaling pathway	4.21×10^{-01}
	CCR1, CSDE1, EFNB3, CBFA2T3, PNOC	axon guidance	8.22×10^{-02}
		morphine addiction	8.22×10^{-02}
		human cytomegalovirus infection	8.22×10^{-02}
		sphingolipid signaling pathway	8.22×10^{-02}
PR33	PCDHA10, PCDHA4, PCDHA7, ARL8B, CBFA2T3	endocytosis	8.22×10^{-02}
	CPN1, CREBBP, GOLPH3, IQGAP1, NFYA	spinocerebellar ataxia	4.79×10^{-01}
		long-term potentiation	4.79×10^{-01}
		adherens junction	4.79×10^{-01}
		regulation of actin cytoskeleton	4.79×10^{-01}
PR34	FBXL16, COMMD9, TULP4, FBXL12, ACAA1	melanogenesis	4.79×10^{-01}
	HPCAL4, TDG, SDC2, MIS12, HUS1	tryptophan metabolism	1.72×10^{-01}
		fatty acid degradation	1.72×10^{-01}
		thyroid hormone signaling pathway	4.12×10^{-01}
		selenocompound metabolism	4.12×10^{-01}
PR35	AKAP1, CTGF, ADCYAP1, AKAP8, KCNB1	proximal tubule bicarbonate reclamation	4.12×10^{-01}
	PDGFA, PITPNM2, PTEN, SLC1A2, SSH2	other types of O-glycan biosynthesis	2.44×10^{-01}
		focal adhesion	2.44×10^{-01}
		melanoma	2.44×10^{-01}
		glioma	2.44×10^{-01}
PR36	RPL22L1, SLC1A2, E2F2, EIF5A2, HECTD2	regulation of actin cytoskeleton	2.44×10^{-01}
	NEFL, PPP2R2A, RBM5, SF3A3, TEAD1	Chagas disease	4.68×10^{-01}
		sphingolipid signaling pathway	4.68×10^{-01}
		phenylalanine metabolism	4.68×10^{-01}
		protein export	4.68×10^{-01}
PR37	HADHB, DYNC1LI2, CCL20, ABCA1, CD40	proximal tubule bicarbonate reclamation	4.68×10^{-01}
	RAD21, MAPRE1, CROT, ADCY1, CDK6	cell cycle	2.68×10^{-01}

Table 3. continued

genome regions in SARS-CoV-2	key proteins in PPI network	KEGG pathways	FDR corrected p-value	genome regions in SARS-CoV-2	key proteins in PPI network	KEGG pathways	FDR corrected p-value
PR19	CLOCK, ESR1, MBD2, ARID4B, EPHA3 HMGA2, IGF2BP3, LRP6, RORA, SEMA3A	neurotrophin signaling pathway	3.91×10^{-01}	PR38	DGKH, GRIA2, MAP3K1, PLCB4, RGS4 ADAMTS2, ATP1B1, GPM6A, PCDH19, STARD13	cytokine–cytokine receptor interaction	2.68×10^{-01}
		AMPK signaling pathway	3.91×10^{-01}			transcriptional misregulation in cancer	2.68×10^{-01}
		thyroid hormone signaling pathway	3.91×10^{-01}			chemokine signaling pathway	2.68×10^{-01}
		circadian rhythm	2.66×10^{-02}			Epstein–Barr virus infection	2.68×10^{-01}
		breast cancer	2.05×10^{-01}			adrenergic signaling in cardiomyocytes	3.43×10^{-02}
		Wnt signaling pathway	2.05×10^{-01}			thyroid hormone synthesis	3.43×10^{-02}
		axon guidance	2.05×10^{-01}			Insulin secretion	3.43×10^{-02}
		endocrine and other factor-regulated calcium reabsorption	3.11×10^{-01}			aldosterone synthesis and secretion	3.75×10^{-02}

pathway. Furthermore, these 160 targets are provided as inputs to the STRING database as well to identify the protein−protein interaction (PPI) network. Out of the 160 targets, the results for at most 10 key proteins as identified from the PPI network and their corresponding top 5 KEGG pathways based on FDR corrected p-values are reported in Table 3, while the detailed analysis is provided in Table S2. The results for the top 5 GO-Enrichment analysis corresponding to each of the 38 potential regions are reported in Table S3 as well. Figure 3 shows the PPI network for potential region PR1, where PIK3CA and TP53BP1 are the proteins with the highest node degree of 8. The node degree in each case represents the interactions in between proteins which may be affected if the specific genes is regulated by the miRNAs. Therefore, a higher node degree gene may affect other related genes as well, eventually leading to different diseases. The network has an average PPI enrichment p-value of 0.00103 with an average node degree of 1.17. In the final phase of this study, repurposable drugs for at most 10 human proteins based on the highest degree as derived from the PPI network are identified using the EnrichR tool. As can be seen from Figure 3 for region PR1, these top 10 human proteins are PIK3CA, TP53BP1, CHD3, ARID1A, SNW1, E2F3, SMARCD1, ARID2, AR, and RELN. Figure 4 shows the docking of the nucleocapsid protein of the potential region PR1 with some of the key human proteins like AR, ARID1A, E2F3, PIK3CA, SMARCD1, and TP53BP1 as presented in Table 3. Their respective docking scores are −138.57, −151.02, −177.80, −233.11, −98.74, and −105.10. Furthermore, based on their p-values, the top 2 drugs targeting the key human proteins as shown in Table 3 are reported in Table 4 where it can be seen that for the top target human proteins corresponding to potential region PR1, the identified drugs are trichostatin A and emodin. The docking of these two drugs with key human proteins like ARID1A, E2F3, SMARCD1, TP53BP1, AR, and PIK3CA are shown in Figures 5 and 6 respectively. Table S4 reports human miRNAs aligned with each of the potential regions along with the corresponding repurposable drugs.

■ DISCUSSION

KEGG Pathway Analysis. To analyze the potential illnesses for the interacting human mRNAs associated with the human miRNAs which are aligned with the 38 potential SARS-CoV-2 regions, KEGG pathway analysis is used by considering KEGG Human EnrichR Tool to reveal the pathways leading to comorbidities. In the tool, all of the target human mRNAs as returned by miRDB are provided as inputs to identify the KEGG pathways. Subsequently, at most 10 key targets as derived from the PPI network are identified. On the basis of the p-value, the top 5 pathways for these key proteins are selected and reported in Table 3. It is to be noted that since the identified potential regions of SARS-CoV-2 as provided in 2 have high similarity with the human miRNAs, the same characteristics of human miRNAs can be exhibited by SARS-CoV-2 once it enters the human body. As shown in Table 3, for example, for potential region PR1 the most significant pathways corresponding to the top 10 target human proteins like PIK3CA, TP53BP1, CHD3, ARID1A, SNW1, E2F3, SMARCD1, ARID2, AR, and RELN that are involved in various comorbidities like hepatocellular carcinoma (FDR corrected p-value 2.81×10^{-01}), Type-II diabetes mellitus (FDR corrected p-value 4.16×10^{-01}), axon guidance (FDR corrected p-value 6.73×10^{-01}), maturity onset diabetes of the young (FDR corrected p-value 6.81×10^{-01}), and ErbB signaling pathway (FDR corrected p-value 6.90×10^{-01}).

Figure 3. Protein−protein interaction network of human target proteins associated with the 16 human miRNAs aligned with PR1.

Protein−Protein Interaction Network. The STRING database is used to study the protein−protein interaction networks for the obtained target human proteins corresponding to each potential SARS-CoV-2 region. As an example, the PPI network of potential region PR1 is depicted in Figure 3, while the rest are given in Figure S2. In the network of Figure 3, PIK3CA and TP53BP1 are the highest interacting nodes with a degree of 8. It should be noted that PIK3CA and TP53BP1 along with E2F3 are responsible for different types of cancer related diseases in human. On the other hand, RELN, PCDHA3, PCDHA2, PCDHA9, POU3F3, PCDHA10, PCDHA7, and PCDHA6 are responsible for nerve degeneration, chemical- and drug-induced liver injury, inflammation, necrosis, weight loss, hypertension, and edema. In summary, the PPI network analysis

suggests that related proteins share common functions although they may not physically interact with one another.

Repurposable Drugs. So far, no potent drug has been discovered to combat COVID-19. Instead of discovering new drugs, which is both time-consuming and expensive, drug repurposing can be considered as an alternative option. In this regard, human proteins as identified from protein−protein interactions can be considered as targets for drug repurposing. The U.S. Food and Drug Administration (FDA) approved drugs that interact with the human proteins are identified using EnrichR's DSigDB. Table 4 reports the top 2 drugs that target the human proteins corresponding to each SARS-CoV-2 potential region. It should be noted that many of the identified drugs are related to cancer. For example, emodin, resveratrol, bicalutamide, vinblastine, paclitaxel, tanespimycin, raloxifene,
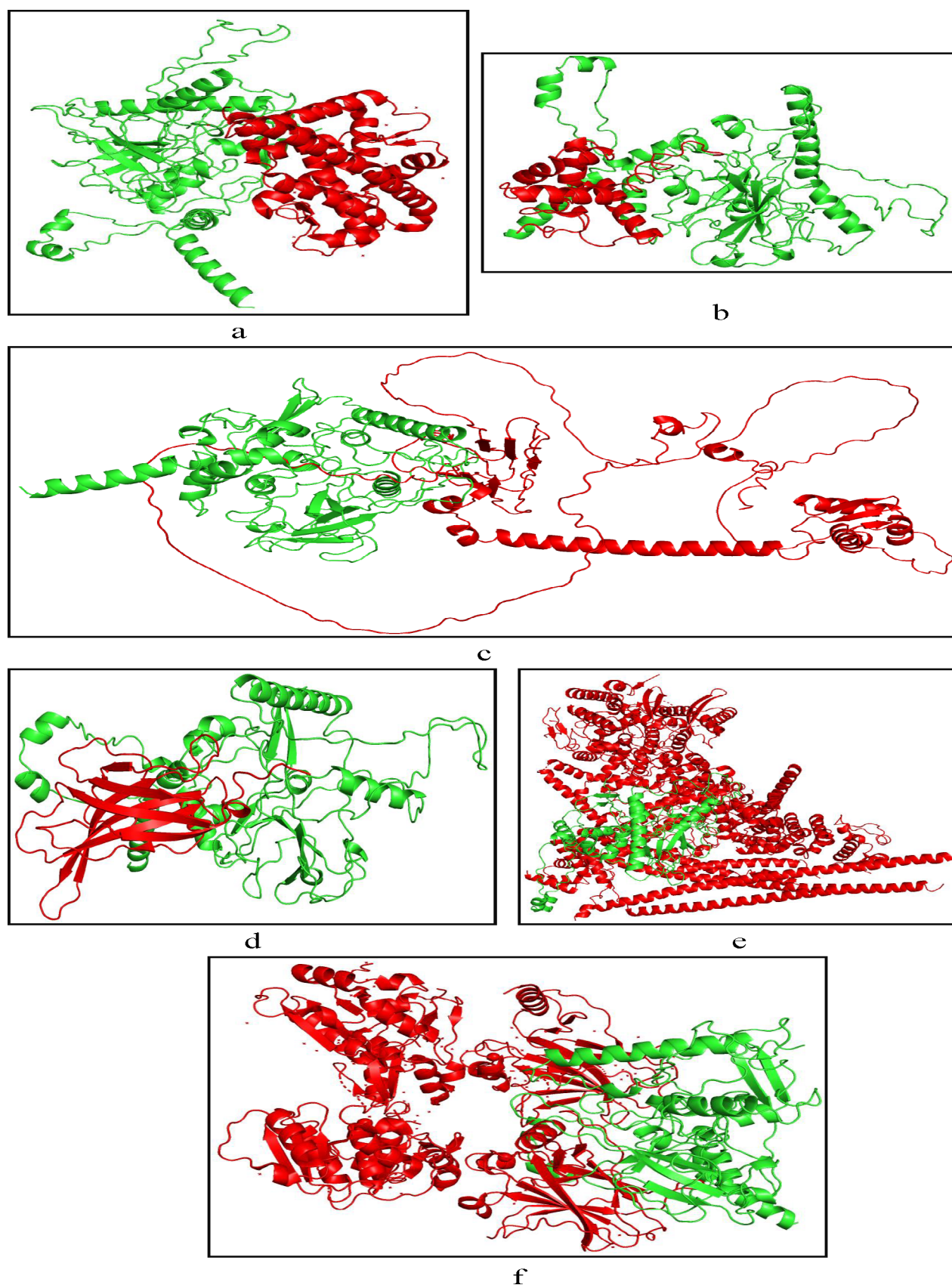
Figure 4. Docking of nucleocapsid protein (shown in green) of potential region (PR1) with key human proteins (red) such as (a) AR, (b) ARID1A, (c) E2F3, (d) PIK3CA, (e) SMARCD1, and (f) TP53BP1 proteins

luteolin, fulvestrant, and doxorubicin are used for treating blood cancer, lymphoma, metastases, lung cancer, solid tumors, and

breast, prostate, and ovarian cancers. On the other hand, other drugs like resveratrol, captopril, digoxin, and lanatoside C are

Table 4. Top 2 Drugs Targeting the Human Key Proteins Associated with Each Potential SARS-CoV-2 Region along with Their p-Values and Treatment Purpose

genomic regions in SARS-CoV-2	target human proteins	drug	p-value	drug bank ID	treatment
PR1	SMARCD1, SNW1, E2F3, TP53BP1, ARID1A	trichostatin A	1.15×10^{-04}	DB04297	antifungal antibiotic
	AR, PIK3CA	emodin	1.40×10^{-04}	DB07715	breast and ovarian cancer
PR2	TCERG1, PPP3R1, MEX3D	resveratrol	1.46×10^{-04}	DB02709	high cholesterol, cancer, heart disease
	FAM63B, IGF2BP3, MEX3D, RGL1, C14ORF169	primaquine	1.49×10^{-04}	DB01087	malaria
PR3	KMT2D, HDAC2, ERCC6L, RRM2, ZBTB10, S1PR1, MAPK1, BAZ2A, DACT1	estradiol	8.46×10^{-06}	DB00783	estrogen
	RRM2, MAPK1	bicalutamide	3.09×10^{-03}	DB01128	prostate cancer
PR4	CD40LG, NCAPG, GTF2I	vinblastine	1.63×10^{-04}	DB00570	breast, testicular cancer, neuroblastoma, Hodgkin's and non-Hodgkins lymphoma, mycosis fungoides, histiocytosis, and Kaposi's sarcoma
	CD40LG, NCAPG, GTF2I	paclitaxel	1.57×10^{-03}	DB01229	breast, ovarian and non-small cell lung cancer
PR5	CSNK1A1, YWHAB	mesalazine	6.62×10^{-04}	DB00244	ulcerative colitis
	ADCY9, YWHAB, AGO2	diclofenac	1.04×10^{-03}	DB00586	aches, pains
PR6	CREBBP, CBX5, BRCA1, HIF1A	vorinostat	3.81×10^{-05}	DB02546	cutaneous T-cell lymphoma
	CREBBP, BRCA1	colchicine	1.65×10^{-04}	DB01394	inflammation and pain
PR7	CUL3, ASB13, HNRNPR, PRKCA	emetine	4.18×10^{-05}	DB13393	antiprotozoal and induce vomiting
	UBE2B, CUL3, LSM14A	clopamide	5.58×10^{-04}	DB13792	gastroparesis in patients with diabetes
PR8	TIA1, HNRNPR, LSM14A	ethosuximide	1.03×10^{-04}	DB00593	Petit Mal seizures
	SYP	abacavir	8.97×10^{-03}	DB01048	human immunodeficiency virus (HIV) infection
PR9	HDAC2	trichostatin a	5.99×10^{-03}	DB04297	antifungal antibiotic
		tanespimycin	2.72×10^{-02}	DB05134	cancer, solid tumors, chronic myelogenous leukemia
PR10	CCR1, PRKACA	tamibarotene	3.02×10^{-02}	DB04942	recurrent APL
	PRKACA	amsacrine	3.49×10^{-02}	DB00276	remission of tumor
PR11	RFC5, DKC1, PSMD1, TNPO1, STK4	captopril	2.99×10^{-05}	DB01197	renovascular hypertension, congestive heart failure, left ventricular dysfunction, and nephropathy
	SPEN;NR1D2	dronabinol	8.71×10^{-04}	DB00470	nausea, vomiting, anorexia and weight loss
PR12	DUSP1, GART, AZIN1	diclofenac	1.04×10^{-03}	DB00586	osteoarthritis and rheumatoid arthritis
	DUSP1, BCL3	pyrvinium	2.02×10^{-03}	DB06816	pinworm infestations
PR13	SPEN, NOVA1, ELAVL2	sulfaguanidine	2.54×10^{-03}	DB13726	bacillary dysentery and other enteric infections
	SPEN, INSR, NR1D2	captopril	7.48×10^{-03}	DB01197	high blood pressure and heart failure
PR14	CCNT1, RGPD6, PDS5B, AFF4	digoxin	4.50×10^{-06}	DB00390	irregular heartbeats including atrial fibrillation
		proscillaridin	8.28×10^{-06}	DB13307	heart failure and cardiac arrhythmi
PR15	MTMR1, CREBBP, PUM2	trichostatin a	5.89×10^{-05}	DB04297	antifungal antibiotic
		vorinostat	6.07×10^{-04}	DB02546	T-cell lymphoma
PR16	EGR1, MAPK1, ESR1	artenimol	2.62×10^{-07}	DB11638	*Plasmodium falciparum* infection
		clioquinol	7.81×10^{-06}	DB04815	fungal infections
PR17	POMC, DNAJC5	imipramine	2.01×10^{-04}	DB00458	antidepressant
	NFIB, TMCC1	ambroxol	3.27×10^{-03}	DB06742	airway secretion clearance therapy
PR18	SEMA3A, RBBP6, TFCP2L1, ZFX, FOXO1, VGLL3	trichostatin a	3.58×10^{-03}	DB04297	antifungal antibiotic
	SEMA3A, ZFX, FOXO1	raloxifene	4.02×10^{-03}	DB00481	osteoporosis and ivasive breast cancer
PR19	XPO1, ADH1B, ADH1A, AXIN1, MAPK1, GGH	phenobarbital	5.51×10^{-06}	DB01174	seizures
	DDX3X, MAPK1, KLHL20	hesperidin	1.42×10^{-05}	DB04703	hemorrhoids, varicose veins, and poor circulation
PR20	TCERG1, PPP3R1, MEX3D	resveratrol	1.46×10^{-04}	DB02709	high cholesterol, cancer, heart disease
	FAM63B, IGF2BP3, MEX3D, RGL1, C14ORF169	primaquine	1.49×10^{-04}	DB01087	relapse of vivax malaria
PR21	IGFBP5, NOA1, SIAH1, INHBB	menadione	1.67×10^{-03}	DB00170	hypoprothrombinemia
	UBE2I, INHBB	anisomycin	9.74×10^{-03}	DB07374	antibiotic
PR22	HDAC2, RUNX1	decitabine	2.23×10^{-03}	DB01262	myelodysplastic syndrome
	CBX5, SH2B3	luteolin	2.71×10^{-03}	DB15584	hypertension, inflammatory disorders, and cancer
PR23	WNT10B, IGDCC3, NLK, SDK2	trichostatin a	1.03×10^{-03}	DB04297	antifungal antibiotic
	WNT10B	permethrin	4.99×10^{-03}	DB04930	scabies
PR24	HTR2C, POLI	clozapine	9.07×10^{-05}	DB00363	schizophrenia
	HTR2C	mirtazapine	5.49×10^{-03}	DB00370	major depression

Table 4. continued

genomic regions in SARS-CoV-2	target human proteins	drug	p-value	drug bank ID	treatment
PR25	SMAD2, CREBBP, SMURF2, AGO3, KDM6A	irinotecan	1.99×10^{-04}	DB00762	metastatic carcinoma of the colon or rectum
	CREBBP, HDAC3, SMURF2	aspirin	2.27×10^{-03}	DB00945	pain, fever, inflammation, migraines, and cardiovascular
PR26	CERK, CHST12	parthenolide	3.49×10^{-03}	DB13063	analgesic, anti-inflammatory and antipyretic
	CERK, CHST12, GNE	trichostatin a	7.92×10^{-03}	DB04297	antifungal antibiotic
PR27	CCNA2, CREBBP	clofibrate	2.83×10^{-04}	DB00636	hypertriglyceridemia and high cholesterol
	GFPT1, OGT	desipramine	5.19×10^{-04}	DB01151	antidepressant
PR28	PURA, COPA, DYNC1LI2, PUM1, SEC62	captopril	2.99×10^{-05}	DB01197	renovascular hypertension, congestive heart failure, left ventricular dysfunction, and nephropathy
	PURA, PUM1, SEC62, ARHGAP12	staurosporine	1.97×10^{-04}	DB02010	antitumor
PR29	MECP2, MAPK1	zebularine	2.51×10^{-04}	DB03068	antitumor
	RAP1A, MMP15, MAPK1	fulvestrant	4.08×10^{-04}	DB00947	breast cancer
PR30	NFAT5, FOXP3	cyclosporin a	3.92×10^{-03}	DB00091	immunosuppressant
	RORA	gemfibrozil	7.48×10^{-03}	DB01241	reduction of serum triglyceride
PR31	TAOK1, NR3C1	vorinostat	9.13×10^{-04}	DB02546	cutaneous manifestations, recurrent cutaneous T- cell lymphoma
	CWC25, NR3C1, ARID1A	lanatoside c	1.21×10^{-03}	DB13467	congestive heart failure, cardiac arrhythmia
PR32	PRKCA, TBC1D2B	menadione	3.27×10^{-03}	DB00170	hypoprothrombinemia
	EFNB3, PRKCA, RHOA	doxorubicin	5.17×10^{-03}	DB00997	cancers and Kaposi's sarcoma
PR33	CREBBP, GOLPH3, NFYA, PCDHA4, PCDHA10, PCDHA7	alsterpaullone	6.23×10^{-06}	DB04014	antitumor
	NFYA, PCDHA4, PCDHA10, PCDHA7	azacitidine	1.41×10^{-04}	DB00928	antineoplastic
PR34	MIS12, HUS1, FBXL12	anisomycin	2.91×10^{-04}	DB07374	antibiotic
	SDC2	salbutamol	1.09×10^{-02}	DB01001	asthma, bronchitis, COPD, bronchospasms
PR35	KCNB1, AKAP8, PTEN, SLC1A2, PDGFA, AKAP1, CTGF	trichostatin a	4.26×10^{-04}	DB04297	antifungal, antibiotic
	PTEN, CTGF	lorazepam	1.80×10^{-03}	DB00186	panic disorders, severe anxiety, and seizures
PR36	RBM5	metoclopramide	1.05×10^{-02}	DB01233	gastresophageal reflux, nausea and vomiting
	E2F2, PPP2R2A	diclofenac	1.83×10^{-02}	DB00586	aches, pains
PR37	ABCA1, CCL20	rimexolone	5.66×10^{-05}	DB00896	inflammation of the eye
	ABCA1, CD40, CROT	diphenylpyraline	6.69×10^{-05}	DB01146	allergic rhinitis, hay fever, and allergic skin disorders
PR38	RGS4, GRIA2, GPM6A, ATP1B1	cytarabine	2.15×10^{-04}	DB00987	leukemia
	RGS4, ATP1B1	trifluoperazine	2.57×10^{-04}	DB00831	depression, anxiety, agitation

used to treat heart and blood vessels, which can help in preventing the thickening of vessel linings. Moreover, proscillaridin, which targets CCNT1, RGPD6, PDS5B, and AFF4 genes, is used to treat heart failure and cardiac arrhythmia. Furthermore, diclofenac, colchicine, and aspirin which target human proteins like ADCY9, YWHAB, AGO2, CREBBP, BRCA1, HDAC3, and SMURF2 are used for the treatment of fever, body pain, and inflammations. It is to be noted that mostly drugs like doxycycline, remedesvir, and ribavirin, etc., are prescribed for COVID-19-afflicted patients and target the virus proteins. When it comes to human proteins, drugs are prescribed for TMPRSS2 and ACE2. Apart from these two human proteins, other target human proteins are mostly unexplored, which we have identified in this work. Since all these target proteins are associated with the human miRNAs and the potential SARS-CoV-2 regions in turn are similar to those miRNAs, the identified drugs for the target human proteins may be used for clinical trials to combat SARS-CoV-2.

It is worth mentioning here that the work presented in Li et al.[36] considers the whole human genomic sequence to identify similar regions with SARS-CoV-2. In this regard, they have identified five such short sequences known as Human Identical Sequences (HIS). These five sequences have starting and ending coordinates at 7570−7595, 12494−12517, 6766−6789, 29860−29886, and 8610−8633 in the reference sequence of SARS-CoV-2 (NCBI Reference Sequence: NC_045512.2). Thereafter, they have concluded that these HIS of SARS-CoV-2 activate expressions of both adjacent and distant genes among which hyaluronan synthase 2 (HAS2) resulted in the accumulation of hyaluronan, which was closely correlated with the severity of COVID-19. On the contrary, in our work, we have considered a new approach in order to identify potential regions in SARS-CoV-2 similar to the human miRNAs (not the whole human genome); thus, the same way by which human miRNAs are inhibited can be applied for such potential regions of virus as well by administering drugs to the interacting human proteins. Therefore, our work and that of Li et al.[36] produce different results as they are conducted on different backgrounds.

■ CONCLUSION

In this work, a new approach has been hypothesized for identifying potential regions in SARS-CoV-2 which are similar to the human miRNAs, thereby exhibiting similar consequences as caused by the human miRNAs in human body. Thus, the same method of inhibition of human miRNAs can be applied for such potential regions of SARS-CoV-2 as well by targeting the interacting human proteins. To achieve this, 2656 human miRNAs are aligned with respect to SARS-CoV-2 reference
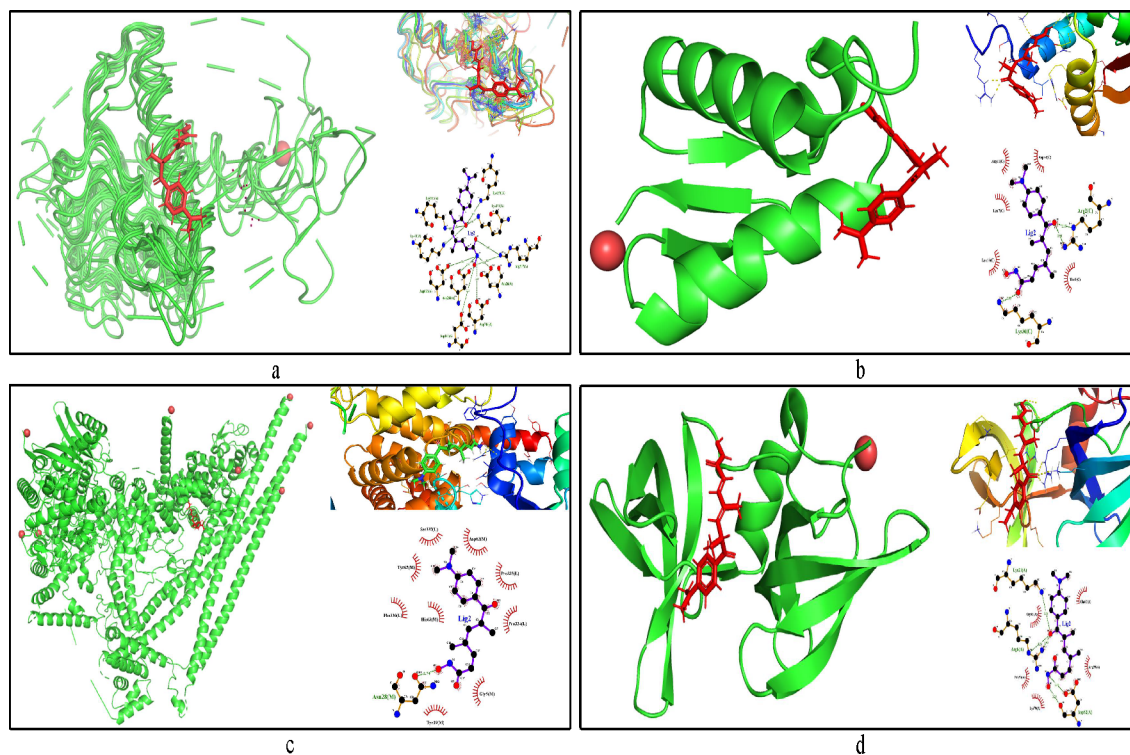
Figure 5. Docking of target human proteins such as (a) ARID1A, (b) E2F3, (c) SMARCD1, and (d) TP53BP1 with trichostatin A for potential region 1 (PR1).
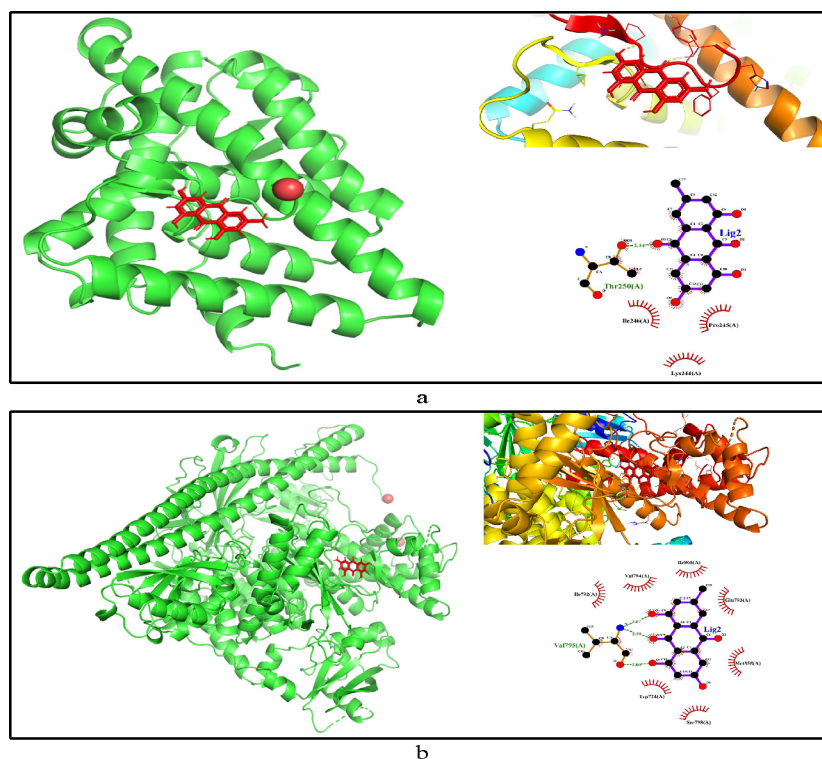

Figure 6. Docking of target human proteins such as (a) AR and (b) PIK3CA with emodin for potential region 1 (PR1).

genome using ClustalO to find the potential regions within the reference genome having high similarities with the human miRNAs. For the aligned miRNAs, the potential regions in SARS-CoV-2 are identified based on the highest number of nucleotide matches which should be greater than or equal to 5 at

a genomic position. As a result, 38 potential SARS-CoV-2 regions are identified, consisting of 249 human miRNAs. Among these 38 potential regions, some top regions belong to nucleocapsid, RdRp, helicase, and ORF8. To understand the biological significance of these potential regions, the targets of

the human miRNAs are considered for KEGG pathways and protein−protein and drug−protein interaction analysis as the human miRNAs are similar to the potential regions of SARS-CoV-2. As a consequence, significant pathways are found which lead to comorbidities like cancer, diabetes, hepatitis C etc. Moreover, repurposable drugs like emodin, bicalutamide, vorinostat, etc. are identified which can be used for clinical trials targeting the human proteins associated with the 38 potential SARS-CoV-2 regions as the human miRNAs are aligned with them. As the findings in our work are *in silico*, as a future scope of study we are trying to collaborate with hospitals having research laboratories to verify the same.

Ethics Approval and Consent to Participate. The ethical approval or individual consent was not applicable.

Availability of Data and Materials. All files which include the data set (raw and aligned sequences), codes, supplementary PDFs, and analysis files for each region are available at http://www.nitttrkol.ac.in/indrajit/projects/COVID-Human-miRNA-SARS-CoV-2-Drug.

Consent for Publication. Not applicable.

ASSOCIATED CONTENT

SI **Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c01907.

List of 38 potential SARS-CoV-2 regions based on aligned human miRNAs along with their highest nucleotide match (Table S1), list of human mRNAs associated with the human miRNAs aligned with the 38 potential SARS-CoV-2 regions along with their KEGG pathways (Table S2), top 5 GO:Biological Process, GO:Molecular Function and GO:Cellular Component of the target human mRNAs for the 38 potential SARS-CoV-2 regions (Table S3), top 2 drugs associated with each of the 41 potential SARS-CoV-2 regions (Table S4), alignment of the human miRNAs with the 38 potential regions of SARS-CoV-2 (Figure S1), protein−protein interactions of the 38 potential regions of SARS-CoV-2 (Figure S2) (PDF)

AUTHOR INFORMATION

Corresponding Authors

Nimisha Ghosh − *Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, 02-097 Warsaw, Poland; Department of Computer Science and Information Technology, Institute of Technical Education and Research, Siksha 'O' Anusandhan (Deemed to be) University, Bhubaneswar 751030 Odisha, India;* ⊙ orcid.org/0000-0002-0697-6368; Email: ghosh.nimisha@gmail.com

Indrajit Saha − *Department of Computer Science and Engineering, National Institute of Technical Teachers' Training and Research, Kolkata 700106 West Bengal, India;* Email: indrajit@nitttrkol.ac.in

Nikhil Sharma − *Department of Electronics and Communication Engineering, Jaypee Institute of Information Technology, Noida 201309 Uttar Pradesh, India;* Email: sharmanikhiljiit@gmail.com

Jnanendra Prasad Sarkar − *Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032 West Bengal, India;* Email: jpsarkar.mcse@gmail.com

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.2c01907

Author Contributions

[⊥]N.G., I.S., and N.S. contributed equally.

ACKNOWLEDGMENTS

REFERENCES

(1) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, 382.

(2) Kim, D.; Lee, J.-Y.; Yang, J.-S.; Kim, J. W.; Kim, V. N.; Chang, H. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **2020**, 181.

(3) Brookman, S.; Cook, J.; Zucherman, M.; Broughton, S.; Harman, K.; Gupta, A. Effect of the new SARS-CoV-2 variant B. 1.1. 7 on children and young people. *Lancet Child Adolesc. Health* **2021**, 5, e9.

(4) van Dorp, L.; Richard, D.; Tan, C.; Shaw, L. P.; Acman, M.; Balloux, F. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat. Commun.* **2020**, 11, 5986.

(5) Coppée, F.; Lechien, J.; Declèves, A.-E.; Tafforeau, L.; Saussez, S. Severe acute respiratory syndrome coronavirus 2: virus mutations in specific European populations. *New Microbes New Infect* **2020**, 36, 100696.

(6) Tang, J.; Toovey, O.; Harvey, K.; Hui, D. S. C. Introduction of the South African SARS-CoV-2 variant 501Y.V2 into the UK. *J. Infect.* **2021**, 82, e8.

(7) Chen, J.; Wang, R.; Wang, M.; Wei, G.-W. Mutations Strengthened SARS-CoV-2 Infectivity. *J. Mol. Biol.* **2020**, 432.

(8) Saha, I.; Ghosh, N.; Pradhan, A.; Sharma, N.; Maity, D.; Mitra, K. Whole genome analysis of more than 10 000 SARS-CoV-2 virus unveils global genetic diversity and target region of NSP6. *Briefings Bioinf* **2021**, 22, 1106−1121.

(9) Worldometers, https://www.worldometers.info/coronavirus/.

(10) Salyer, S.; Maeda, J.; Sembuche, S.; Kebede, Y.; Tshangela, A.; Moussif, M.; Ihekweazu, C.; Mayet, N.; Abate, E.; Ouma, A. O.; Nkengasong, J. Articles The first and second waves of the COVID-19 pandemic in Africa: a cross-sectional study. *Lancet* **2021**, 397, 1265.

(11) Ko, K.; Nagashima, S.; E, B.; Ouoba, S.; Akita, T.; Sugiyama, A.; Ohisa, M.; Sakaguchi, T.; Tahara, H.; Ohge, H.; et al. Molecular characterization and the mutation pattern of SARS-CoV-2 during first and second wave outbreaks in Hiroshima, Japan. *PloS one* **2021**, 16, No. e0246383.

(12) Carthew, R.; Sontheimer, E. Origins and mechanisms of miRNAs and siRNAs. *Cell* **2009**, 136, 642−55.

(13) Jafri, M.; Qahtani, M.-A.; Shay, J. Role of miRNAs in human cancer metastasis: Implications for therapeutic intervention. *Semin. Cancer Biol.* **2017**, 44, 117.

(14) Kim, S.; Bae, W.; Ahn, J.; Heo, J.-H.; Kim, K.-M.; Choi, K. W.; Sung, C. O.; Lee, D. MicroRNA signatures associated with lymph node metastasis in intramucosal gastric cancer. *Mod. Pathol.* **2020**, 34.

(15) Tang, S.; Li, S.; Liu, T.; He, Y.; Hu, H.; Yunhe; Zhu; Tang, S.; Zhou, H. MicroRNAs: Emerging oncogenic and tumor-suppressive regulators, biomarkers and therapeutic targets in lung cancer. *Cancer Lett.* **2021**, 502.

(16) Javandoost, E.; Majd, E.-F.; Rostamian, H.; K-Koosheh, M.; Hamid; Mirzaei, R. Role of microRNAs in Chronic Lymphocytic

Leukemia Pathogenesis. *Curr. Med. Chem.* **2019**, DOI: 10.2174/0929867326666190911114842.

(17) Bernier, A.; Sagan, S. The Diverse Roles of microRNAs at the Host−Virus Interface. *Viruses* **2018**, *10*, 440.

(18) Zhao, D. MicroRNAs: Mediators and Therapeutic Targets to Airway Hyper Reactivity After Respiratory Syncytial Virus Infection. *Frontiers in Microbiology* **2018**, DOI: 10.3389/fmicb.2018.01236.

(19) Haddad, H.; Al-Zyoud, W. miRNA target prediction might explain the reduced transmission of SARS-CoV-2 in Jordan, Middle East. *Non-coding RNA Res.* **2020**, *5*, 135−143.

(20) Natarelli, L.; Parca, L.; Mazza, T.; Weber, C.; Virgili, F.; Fratantonio, D. MicroRNAs and Long Non-Coding RNAs as Potential Candidates to Target Specific Motifs of SARS-CoV-2. *Non-coding RNA* **2021**, *7*, 14.

(21) Elnabi, S.; Elhiti, M.; El-Sheekh, M. A new approach for COVID-19 treatment by micro-RNA. *Med. Hypotheses* **2020**, *143*, 110203.

(22) Sardar, R.; Satish, D.; Birla, S.; Gupta, D. Integrative analyses of SARS-CoV-2 genomes from different geographical locations reveal unique features potentially consequential to host-virus interaction, pathogenesis and clues for novel therapies. *Heliyon* **2020**, *6*, No. e04658.

(23) Chen, L.; Zhong, L. Genomics functional analysis and drug screening of SARS-CoV-2. *Genes Dis* **2020**, *7*, 542.

(24) Ying, Y.-T.; Min, C.; De, W.-C. Annotation of miRNAs in COVID-19 coronavirus. *J. Electron. Sci. Technol.* **2020**, *19*, 100060.

(25) NCBI. https://www.ncbi.nlm.nih.gov/nuccore/1798174254.

(26) Mirbase, http://www.mirbase.org/.

(27) Sievers, F.; Higgins, D. Clustal Omega. *Curr. Protoc. Bioinf.* **2014**, *48*, 3.13.1−3.13.16.

(28) Chen, Y.; Wang, X. MiRDB: An online database for prediction of functional microRNA targets. *Nucleic Acids Res.* **2019**, *48*, D127.

(29) Chen, E.; Tan, C.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G. V.; Clark, N. R.; Ma'ayan, A. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinf* **2013**, *14*, 128.

(30) Kuleshov, M.; Jones, M.; Rouillard, A.; Fernandez, N. F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S. L.; Jagodnik, K. M.; Lachmann, A.; McDermott, M. G.; Monteiro, C. D.; Gundersen, G. W.; Ma'ayan, A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **2016**, *44*, W90−W97.

(31) Xie, Z.; Bailey, A.; Kuleshov, M. V.; Clarke, D. J. B.; Evangelista, J. E.; Jenkins, S. L.; Lachmann, A.; Wojciechowicz, M. L.; Kropiwnicki, E.; Jagodnik, K. M.; Jeon, M.; Ma'ayan, A. Gene Set Knowledge Discovery with Enrichr. *Curr. Protoc.* **2021**, DOI: 10.1002/cpz1.90.

(32) Enrichr. https://maayanlab.cloud/Enrichr/.

(33) Stringdb. https://string-db.org/.

(34) Szklarczyk, D.; Gable, A. L.; Nastou, K. C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N. T.; Legeay, M.; Fang, T.; Bork, P.; Jensen, L. J.; v. Mering, C. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **2020**, *49*, D605.

(35) Mirdb. http://mirdb.org/.

(36) Li, W.; Yang, S.; Xu, P.; Zhang, D.; Tong, Y.; Chen, L.; Jia, B.; Li, A.; Lian, C.; Ru, D.; et al. SARS-CoV-2 RNA elements share human sequence identity and upregulate hyaluronan via NamiRNA-enhancer network. *EBioMedicine* **2022**, *76*, 103861.