

Chapter 14

Inferring Recombination Events in SARS-CoV-2 Variants In Silico



Nihal Najeeb, Aparna B. Murukan, Anagha Renjitha, Malavika Jayaram, Ayisha A. Jabbar, Haripriya Haridasan, Akshara Prijikummar, Sneha Baiju, Adrial Ann Nixon, Ponnambil Anantha Krishnan, Sunu Rodriguez, Somesh Kumar, Sunil K. Polipalli, Keshav K. Singh, Bipin G. Nair, Sudeep D. Ghate, R. Shyama Prasad Rao, Polavarapu Bilhan Kavi Kishor, Arya Aloor, Renuka Suravajhala, Gyaneshwer Chaubey, and Prashanth Suravajhala

Abstract Over the last 34 months, at least 10 severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) distinct variants have evolved. Among these, some were more infectious while others were not. These variants may serve as candidates for identification of the signature sequences linked to infectivity and viral transgres-

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-28012-2_14.

N. Najeeb · A. B. Murukan · A. Renjitha · M. Jayaram · A. A. Jabbar · H. Haridasan · A. Prijikummar · S. Baiju · A. A. Nixon · P. A. Krishnan · S. Rodriguez · B. G. Nair · A. Aloor · R. Suravajhala · P. Suravajhala (✉)
Amrita School of Biotechnology, Amrita Vishwa Vidyapeetham, Clappana, Kerala, India
e-mail: prash@am.amrita.edu

S. Kumar · S. K. Polipalli
Genome Sequencing Lab, Lok Nayak Hospital, Delhi, India

K. K. Singh
Department of Genetics, Heersink School of Medicine, University of Alabama at Birmingham, Kaul Genetics Building, Birmingham, AL, USA

S. D. Ghate · R. S. P. Rao
Center for Bioinformatics, NITTE University, Mangaluru, India

P. B. K. Kishor
Department of Genetics, Osmania University, Hyderabad, India

G. Chaubey
Cytogenetics Laboratory, Department of Zoology, Banaras Hindu University, Varanasi, Uttar Pradesh, India

sions. Based on our previous hijacking and transgression hypothesis, we aimed to investigate whether SARS-CoV-2 sequences associated with infectivity and trespassing of long noncoding RNAs (lncRNAs) provide a possible recombination mechanism to drive the formation of new variants. This work involved a sequence and structure-based approach to screen SARS-CoV-2 variants *in silico*, taking into account effects of glycosylation and links to known lncRNAs. Taken together, the findings suggest that transgressions involving lncRNAs may be linked with changes in SARS-CoV-2–host interactions driven by glycosylation events.

Keywords SARS-CoV-2 · COVID-19 · Spike protein · Variant · Glycosylation · lncRNA

Nihal Najeeb, Aparna B. Murukan, Anagha Renjitha, Malavika Jayaram, Ayisha A. Jabbar, Haripriya Haridasan, Akshara Prijikumar, Sneha Baiju, Adrial Ann Nixon, Ponnambal Anantha Krishnan and Sunu Rodriguez contributed equally with all other contributors.

1 Introduction

Since the emergence of the unique coronavirus 2019 (COVID-19) which first appeared in Wuhan, China, the mechanism of how the severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) spike protein mediates viral binding and how post-translational modifications affect this is beginning to be understood [1, 2]. Glycosylation is a significant post-translational event that can influence protein structure and functional characteristics either directly or indirectly. Structural studies have shown that the spike protein and spike-angiotensin converting enzyme 2 (ACE2) complexes exhibit several glycosylations, which may have a substantial impact on the ability of the virus to infect and induce an immune response in the host [1]. The SARS-CoV-2 spike protein glycans are sometimes referred to as a “glycan shield” because they sterically obscure the underlying polypeptide epitopes from detection from potentially neutralizing antibodies [2–4]. In addition, the spike protein receptor binding domain (RBD) glycans play a critical role in binding proteins involved in COVID-19 pathogenesis, including the ACE2 receptor and transmembrane protease, serine 2 (TMPRSS2), to host glycoproteins [5–6]. Because viral glycoproteins are exposed on the virus surface, they are the primary targets of host antibodies [7]. In turn, all antibodies are glycoproteins and the attached glycans can have a major impact on their function in the immune response [8]. Therefore, understanding how the spike protein is glycosylated has crucial implications for studies on SARS-CoV-2 pathobiology and vaccine development.

Glycosylation is the enzyme-catalyzed addition of a sugar molecule/oligosaccharide to a macromolecule such as a protein. Nitrogen (N)-linked glycosylation takes place co-translationally on asparagine residues at a specific sequence on the nascent protein known as a sequon, which consists of asparagine-X-threonine/

serine/cysteine (AsnXThr/Ser/Cys), where X cannot be proline (Pro) [9]. Oxygen (O)-linked glycosylation occurs post-translationally on the side chain of Ser or Thr residues during transport of the nascent proteins through the Golgi compartment of cells [9]. *N*-glycans contain a common pentacore which consists of two *N*-acetylglucosamine (GlcNAc) and three mannose residues which can be extended by various monosaccharide units via the action of various glycosyl-transferases, and these units can be modified by glycosidases. Since the process depends on multiple factors such as cell type and metabolic state, the resultant glycan structures are often heterogeneous in nature. If the pentacore is extended only with mannose, the structure is known as high mannose. *N*-glycans of the second type are complex sugars where the two antennae of the pentacore are extended by different sugars including GlcNAc, galactose, fucose, and sialic acid residues. If one antenna is extended with mannose and other with various monosaccharides the sugar structure is known as hybrid. Based on linkages and composition, the structure can be further divided to many subtypes creating high complexity [10].

Host glycoproteins on cells such as those of the immune system play a major role in the pathogenic and immunogenic activity during infections. SARS-CoV-2 infection induces changes in the pattern of host antibody glycosylations with significant variations in the levels of IgG galactosylation and fucosylation [11]. The increase in fucosylation can lead to increased production of proinflammatory cytokines which can lead to the damaging cytokine storm effect in patients [12–17]. In addition to these effects, spike protein glycan variations can modify binding of viruses to host receptors and alter the severity of the pathogenesis and immune responses [10, 18]. Importantly, the composition of *N*-glycosylation modifications on viruses and host cell receptors has been reported to have a significant impact on virus-receptor identification, binding, and cellular penetration [19, 20]. For example, seven glycosylation sites on the SARS-CoV-1 spike protein from the 2002–2004 epidemic were shown to be necessary for dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC/L-SIGN)-mediated infection [21]. The extracellular domain of the human ACE2 receptor contains seven *N*-glycosylations (Asn53, Asn90, Asn103, Asn322, Asn432, Asn546, and Asn690) and several *O*-glycosylations, which are likely to impact viral entry into host cells [22, 23]. The glycosylations at N90 and N322 appear to be important in binding to SARS-CoV-2 spike protein RBD [21, 24]. Also, molecular dynamic simulations have shown that the glycan linked to the ACE2 Asn90 position interferes with virus binding, explaining reports of heightened susceptibility to infection when glycosylation at this site is removed [25]. With hyper sialylation and oligomannose-type modification of ACE2 glycans, the binding affinity between ACE2 and SARS-CoV-2 spike decreases modestly [26]. Acting in concert with the ACE2 receptor, the TMPRSS2 protease involved in SARS-CoV-2 entry into host cells is glycosylated at amino acids Asn213 and Asn249 [6]. However, the structural impact of this has not been investigated extensively. In terms of glycosylated structures, SARS-CoV-2 is reported to bind specifically to heparan sulfate and sialic acid residues on host cells [27, 28]. As many immune cells such as macrophages, dendritic cells, T cells and B cells, and

immune system proteins are glycosylated, it is likely that SARS-CoV-2 may have interactions with these in the ensuing pathogenic and immunogenic processes.

Although most studies on the host response to viral infections have focused on genes that encode proteins, it is now emerging that noncoding RNA molecules are also involved [29]. Early studies in this field found that changes in the expression of long noncoding RNAs (lncRNAs) can alter the innate immune response during viral infections [30, 31]. A more recent study found changes in the expression of multiple lncRNAs during SARS-CoV-2 infection of human bronchial epithelial cells [32]. Another investigation found that dysregulated lncRNAs in SARS-CoV-2 infection are involved in multiple aspects of the infection process including viral proliferation, the host immune response, and disease outcome [33]. There is also evidence that rearrangements or polymorphisms in lncRNAs may drive disease-causing mutations, as shown in cancer research [34, 35].

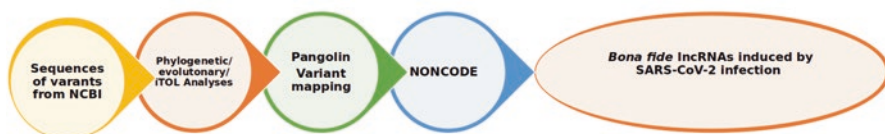
In this study, we have carried out *in silico* analyses to determine (1) if any new lncRNAs are known in transgression pathways induced by SARS-CoV-2 infection, (2) whether or not lncRNAs encoded or transgressed by the virus could provide clues into the mechanisms of how the SARS-CoV-2 variants have emerged, and (3) if changes in RBD N-glycosylation status are associated with the altered binding affinity of different SARS-CoV-2 variants. For the latter, we used *in silico* docking complex analyses to calculate the effect of binding energies between host glycan and spike protein variants. This involved comparison of the binding energies of the Omicron (7WPB) and Delta variants (7TEW) with that of the Wuhan strain (6LZG), with respect to the three commonly found host glycan structures A2F, 6G1, and Man 5.

2 Methods

2.1 Datasets

SARS-CoV-2 and selected nucleotide sequences were retrieved from an NCBI database search [36]. We filtered the several thousands of results using Boolean expressions AND, OR, and NOT and retrieved RefSeq accession numbers of relevant annotated sequences. Following this, we performed NCBI BLAST search in which SARS-CoV-2 reference sequences were compared to the SARS-CoV-2 genome using different databases (nucleotide collection [nr/nt], sequence read archives, refseq representative genome, Protein Data Bank, refseq genome database, whole-genome shotgun contigs, refseq select RNA sequences, expressed sequence tag). The results obtained were tabulated with information on similarity and dissimilarity between the query and subject (Fig. 14.1a). We chose accessions based on characteristics such as e-value, mismatches, and % identity, and selected hits were used for downstream analysis (Table 14.1 and Supplementary Table 14. ST1). We then used Protein Data Bank (PDB) to screen candidate SARS-CoV-2

A)



B)

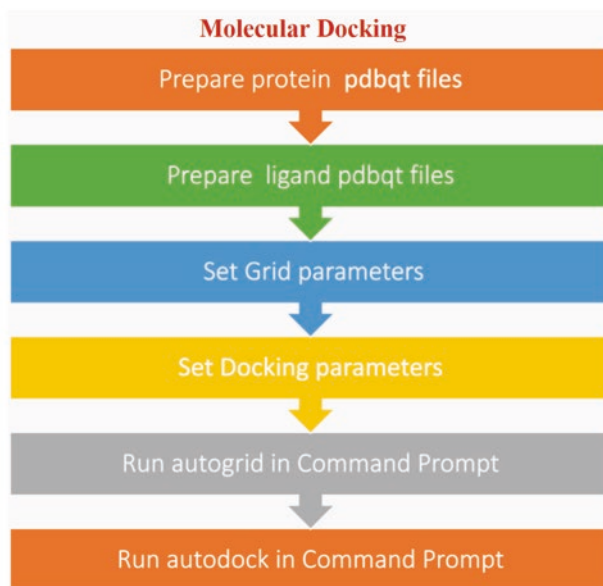


Fig. 14.1 Pictorial methodology of the tools used for the analysis. **(a)** Sequences of the SARS-CoV-2 variants were downloaded from NCBI and subjected to phylogenetic/evolutionary analyses before reconfirming their lineages with Pangolin. As a final check to understand the matching lncRNAs, we used the NONCODE.org database to analyze these by BLAST. **(b)** Overview of the molecular docking analysis

spike protein sequences from the Delta (pdb id: 7TEW) and Omicron (pdb id: 7WPB) variants against the spike sequences of the Alpha, Beta, and Gamma variants as reference (pdb id: 6LZG) (Fig. 14.1b).

2.2 Structural Interpretation and Docking

MolView was used to visualize small molecules of 2D and 3D structures. We inserted the Simplified Molecular Input Line Entry System (SMILE) of the molecule to obtain 2D and 3D structures and downloaded these in spatial data file (SDF)

Table 14.1 List of spike proteins associated with lncRNAs

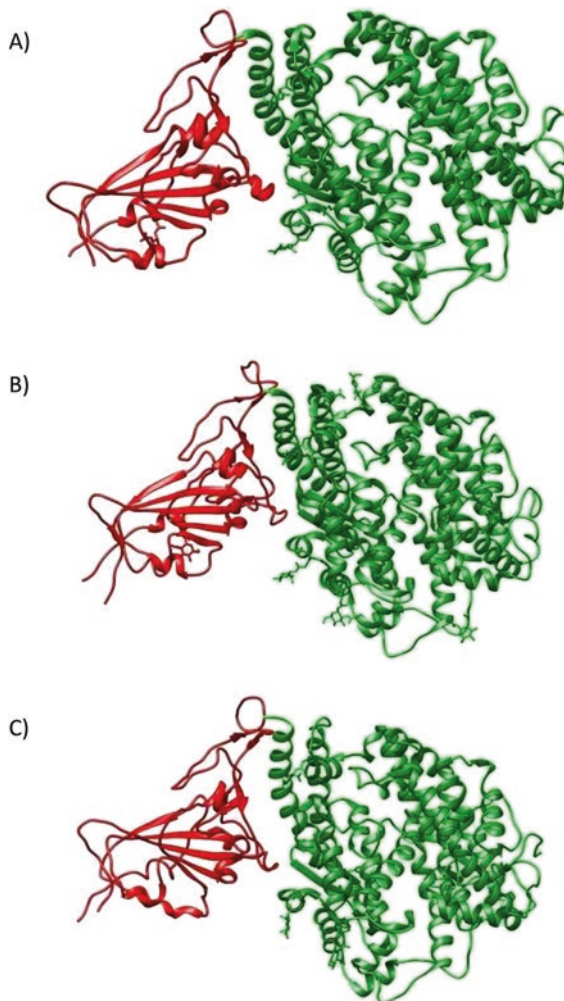
Accession number	lncRNA	Query length	Putative variant
MZ558096.1	NONHSAT156862.1	21,705–21,726	Deltacron
MZ558096.1	NONHSAT079728.2	21,793–21,814	Deltacron
MZ427312.1	NONHSAT156862.1	21,701–21,722	Gamma
MZ427312.1	NONHSAT079728.2	21,789–21,810	Gamma
MZ433432.1	NONHSAT156862.1	21,713–21,734	Beta
MZ433432.1	NONHSAT079728.2	21,801–21,822	Beta
MZ 297238.1	NONHSAT247026.1	12,481–12,504	Beta
OK189649.1	NONHSAT247026.1	12,494–12,517	Delta
ON017450.1	NONHSAT156862.1	21,701–21,722	Zeta
ON017450.1	NONHSAT079728.2	21,789–21,810	Zeta
MZ780476.1	NONHSAT156862.1	21,720–21,741	Beta
MZ780476.1	NONHSAT079728.2	21,808–21,829	Beta
ON017446.1	NONHSAT247026.1	12,461–12,484	Zeta

format [37]. AutoDock containing Molecular Graphics Laboratory (MGL) tools and Autodock4 was used for in silico docking, evaluating the binding energy (ΔG) and binding inhibition constant (K_i). We used MGL tools to set the parameters of ligand and protein by minimizing the energies and converted the files to Protein Data Bank, Partial Charge (Q), and Atom Type (T) (pdbqt) files for both ligands and proteins. The grid parameters were set for each protein with respect to each ligand separately by considering the X, Y, and Z coordinates, and the Grid Parameter Files (GPFs) were generated. Furthermore, the docking parameters for each protein with respect to each ligand were set considering Lamarckian and generic algorithms and the files were saved as dock parameter files (DPFs). Autogrid in Autodock commands were run using command prompt (Fig. 14.1). Chimera was used for visualization (<https://www.cgl.ucsf.edu/chimera/download.html>) in the analysis of spike protein from three different SARS-CoV-2 strains (6LZG: original Wuhan strain, 7TEW: Delta variant, and 7WPB: Omicron variant) (Fig. 14.2).

2.3 Selection of Ligands

The ligands were chosen based on binding patterns of host glycans to the spike protein RBD with steric hindrance checked for each [28]. We used the glycan structures from PubChem to obtain 2D and 3D structures [38]. These structures were downloaded in SDF format and then converted into PDB format (Table 14.2). The three ligands used were A2F *N*-glycan, 6 G1-glycan, and mannose.

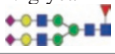

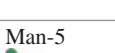
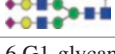
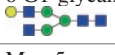
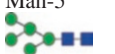


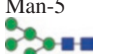
Fig. 14.2 Representation of SARS-CoV-2 spike protein 3D structures from (a) the March 2020 original strain (6LZG) and (b) the Delta (7TEW) and (c) Omicron (7WPB) variants



2.4 Phylogenetic and Pangolin Analyses

Clustal Omega was used to align multiple sequences [39]. The sequences from best hits selected from BLAST searching were converted into FASTA files which were uploaded in Clustal Omega to obtain the alignment results [40]. From this, we obtained guides and phylogenetic trees showing the evolution of the different strains. The guide tree data from the Clustal Omega analysis was uploaded to the Interactive Tree of Life (iTOL) online tool to obtain a circular phylogenetic tree [41]. We also used Pangolin (Phylogenetic Assignment of Named Global Outbreak Lineages) to assign lineages to genome sequences of SARS CoV-2 [42].

Table 14.2 Docking results showing the binding energy and affinity of the ligands for the indicated amino acids

Protein	PBD ID	ligand	ΔG	K_i	AA residues
SARS-CoV-2 spike protein (reference sequence)	6LZG	A2F N-glycan 	-9.11 kcal/mol	210.60 nM	ASP364, CYS336, NAG601, GLY339, LEU441, ASN440
		6 G1-glycan 	-7.19 kcal/mol	5.37 μ M	ASN370, SER371, LEU368, PHE374, NAG601
		Man-5 	-4.89 kcal/mol	261.35 μ M	NAG601, PHE342
SARS-CoV-2 Delta spike protein	7TEW	A2F N-glycan 	-7.79 kcal/mol	1.94 μ M	NAG706, SER317, VAL316, NAG704, GLU312, LYS313
		6 G1-glycan 	-6.97 kcal/mol	7.75 μ M	NAG704, VAL316, LYS313
		Man-5 	-3.62 kcal/mol	2.22 μ M	NAG706, SER545, SER317, LYS313, ILE421
SARS-CoV-2 Omicron spike protein	7WPB	A2F N-glycan 	-7.79 kcal/mol	1.94 μ M	NAG902, ASN546, SER317, LYS313
		6 G1-glycan 	-6.37 kcal/mol	21.24 μ M	NAG902, VAL316, LYS313
		Man-5 	-4.29 kcal/mol	711.08 μ M	PHE374, TYR362, ILE431, LEU365, VAL364, PHE339

◆ = sialic acid, ● = galactose, ■ = GlcNAc, ● = mannose, ▲ = core fucose

Asp aspartate, *Cys* cysteine, *Gly* glycine, *Leu* leucine, *Asn* asparagine, *Ser* serine, *Phe* phenylalanine, *Val* valine, *Lys* lysine, *Ile* isoleucine, *Tyr* tyrosone, *NAG* N-acetylglucosamine

2.5 LncRNA Analysis

We used the Noncode RNA database [43] to enable retrieval of data and to compare lncRNA sequences with SARS-CoV-2 and host protein sequences using BLAST. The query sequence was given in FASTA format and the database used was NONCODE V6 animal. From the obtained hits, the ones with e-values less than zero, we chose human lncRNAs and sought to check the expression profile data to know where the particular lncRNA is expressed in human. Similarly, BLAST was performed for every other accession selected from NCBI earlier. Finally, data wrapper was used to design charts ranging from simple bars and lines to arrow, range, and scatter plots, which can be done using steps [44].

3 Results and Discussion

3.1 *Sequence Similarities and Dissimilarities for All SARS-CoV-2 Variants*

The SARS CoV-2 nucleotide sequences downloaded from the NCBI database were BLAST-searched against the RefSeq database and the resulting candidate hits are listed in Supplementary Table 14.ST2. The cut off for the e-value was set at <0 which indicates that the sequence is an exact match to the query. Thus lower e-values are indicative of better hits with respect to % identity and query coverage. The % identity for the 27 assemblies ranged between 89.5% and 100%. The assembled sequences were then assigned to specific SARS-CoV-2 variant sequences. For this, we employed a similar strategy as above and the best hits were tabulated in Supplementary Table 14.ST3. While a large number of sequences were mapped to Beta and Zeta, some were mapped to Gamma and a lower number to the Delta and Deltacron [45] variants.

3.2 *Pangolin Outbreak Lineages of 38 Different Variants*

The Pangolin tree yielded seeded guide alignments, and hidden Markov model (HMM) profile–profile techniques were used to generate alignments between three or more sequences. We considered the accession numbers of 38 different variants of the virus, with the rest of the information downloaded in FASTA format for multiple sequence alignment (Fig. 14.3a). The tree showed distinct clades with many variants sub-claded together. The ones which were sub-claded were assumed to belong to the same variants which we confirmed. We also checked assignment conflicts, ambiguity, and lineages showing the metadata files of the given accession number of viruses including the type of variants, dates, and regions where these variants were also depicted. From this, we finally considered 11 sequences as belonging to the Gamma, Delta, and Beta variants (Fig. 14.3b).

3.3 *Noncoding RNA Sequences Known to Be Trespassed*

We searched the NONCODE database to identify lncRNA sequences within the complete genome sequences of the selected viral accessions in FASTA format. From the resulting table, we considered human lncRNA sequences that showed 100% sequence identity (Fig. 14.4). Binary values of 0 and 1 were ascribed to a lncRNA sequence if absent or present, respectively, in specific viral accession numbers and entries were summed row- and column-wise to obtain the final lists of matching lncRNAs. This showed that MW562722.1 had the lowest sum of 17

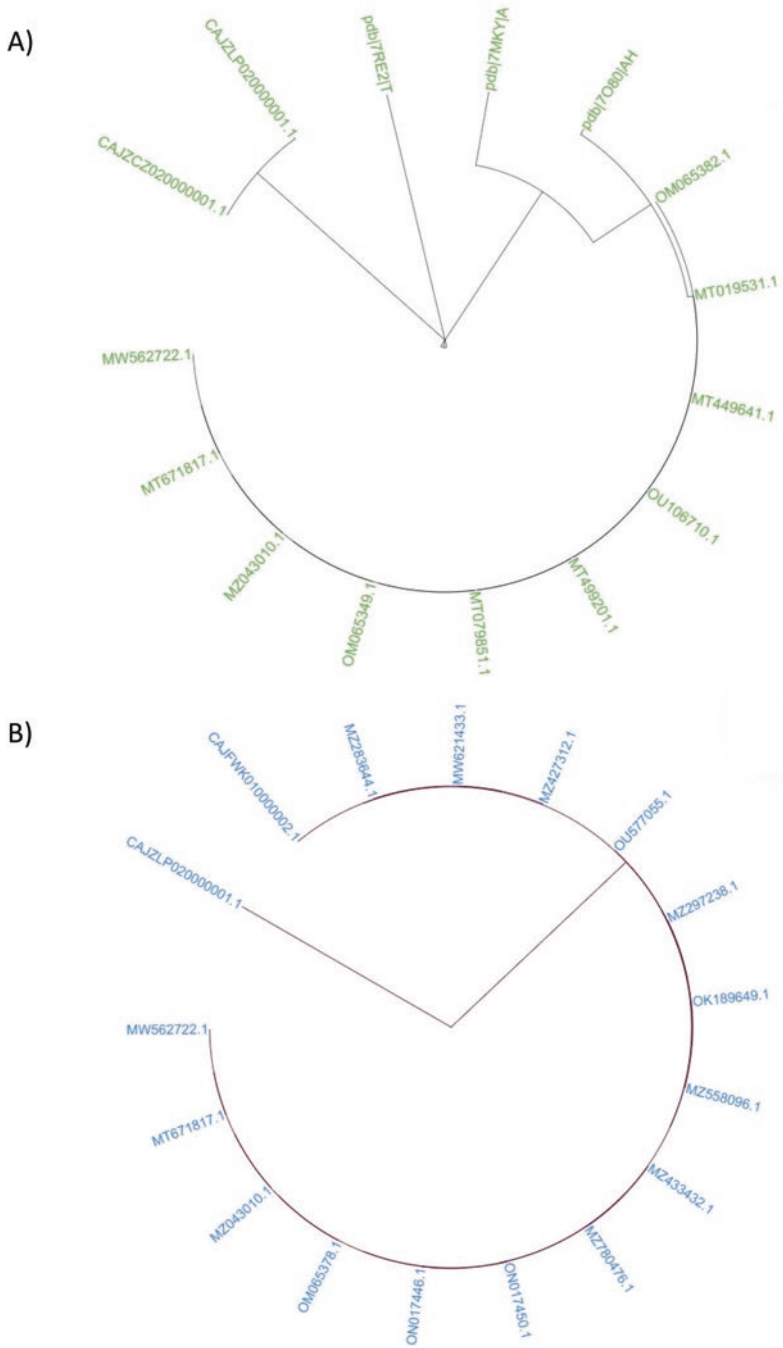


Fig. 14.3 Circular phylogenetic trees of (a) all similar SARS-CoV-2 sequences and (b) dissimilar sequences emerging from different SARS-CoV-2 variants

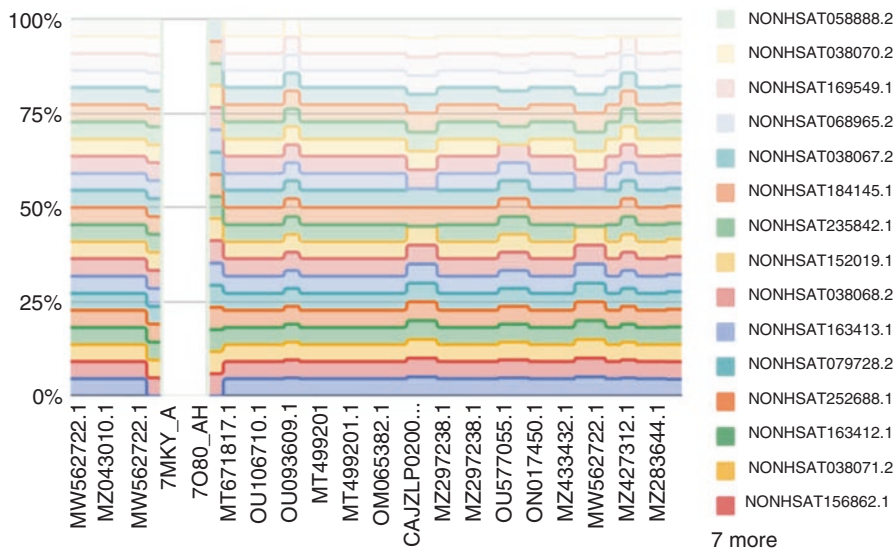


Fig. 14.4 Visualization of lncRNAs (NONHSAT codes) versus viral accessions. The Y-axis shows percentage of lncRNAs present for each viral accession (X-axis). This analysis revealed no lncRNAs in two viral accessions: 7MKY_A (SARS-CoV-2 chain A, RNA 66-MER) and 7O80_AH (SARS-CoV-2 chain AH, mRNA). Note that the figure shows only 15 of the 22 lncRNAs

lncRNA sequences (NONHSAT252687.1, NONHSAT209697.1, NONHSAT209698.1, NONHSAT169548.1, NONHSAT155452.1, NONHSAT209695.1, NONHSAT156862.1, NONHSAT038071.2, NONHSAT163412.1, NONHSAT252688.1, NONHSAT079728.2, NONHSAT163413.1, NONHSAT038068.2, NONHSAT152019.1, NONHSAT235842.1, NONHSAT184145.1, NONHSAT038067.2). We also found strains with a sum of 22 as the highest number of lncRNA sequences. Some of the lncRNA sequences such as NONHSAT163412.1 and NONHSAT163413.1 were repeated in more than 30 viral sequences. We also identified the loci of the genes which code for the spike protein by examining sequences of the identified strains using the Pangolin tool. We then selected these stains and retrieved the sequences from the NCBI database and searched for those with 100% identity with lncRNA sequences. This resulted in the identification of seven viral accession numbers which had similarities to some lncRNAs (Table 14.1). Finally, the hypothesized transgression hypothesis is depicted in Fig. 14.5.

3.4 Molecular Interaction Studies

Our in silico molecular docking approach investigated the binding of the SARS-CoV-2 spike protein variants to specific carbohydrate groups. It revealed that the original Wuhan SARS-CoV-2 spike protein (6LZG) bound to A2F N-glycan with a

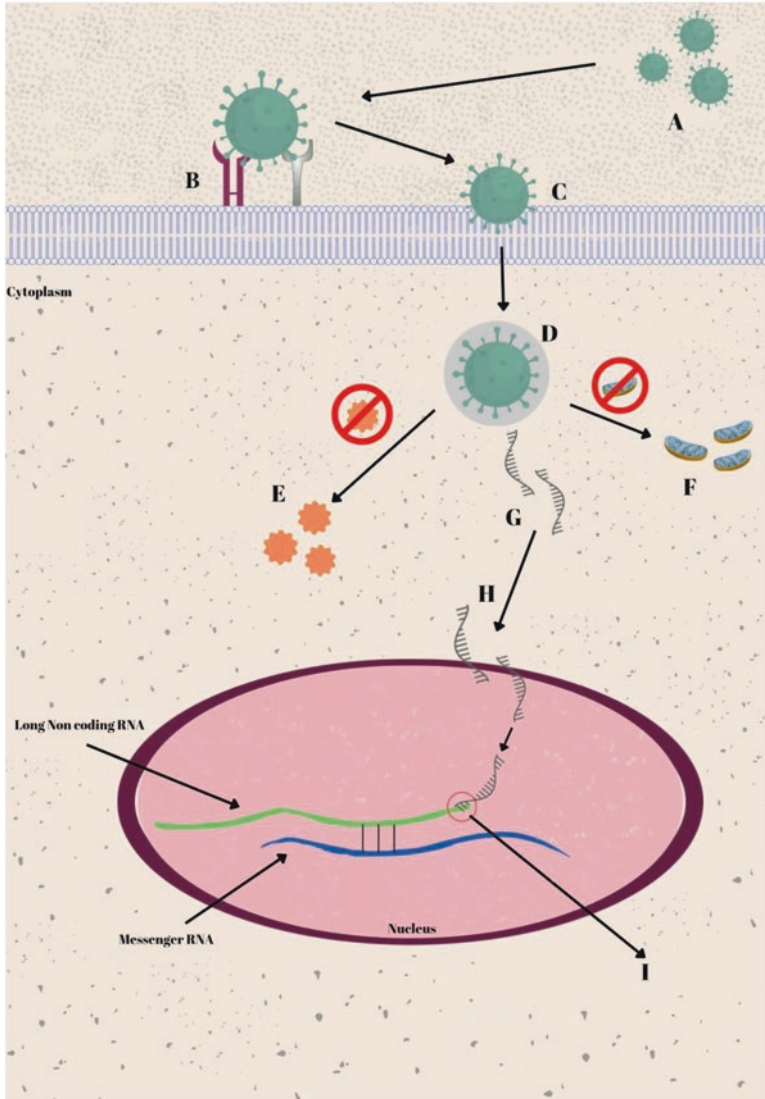


Fig. 14.5 Hypothesis of how lncRNAs transgress: (A) SARS-CoV-2 particles enter the body; (B) the spike protein binds to the host ACE2 receptor, followed by cleavage by TMPRSS2 protease which activates the fusion process; (C) the virus fuses with host cell membrane; (D) the virus enters the cell by endocytosis; (E) the virus destroys or deactivates interferons and interleukins responsible for innate immunity; (F) the virus hijacks mitochondria; (G) the virus undergoes uncoating and release of RNA; (H) the viral RNA enters the nuclei; and (I) the viral RNA transgresses specific lncRNAs of the host cell and takes the neighboring genes under its control

low free energy (ΔG) of -9.11 kcal/mol and high affinity (K_i) of 210.60 nM, at the indicated RBD amino acids (Fig. 14.6 and Table 14.2). 6 G1-glycan bound to the 6LZG spike protein with a ΔG of -7.19 kcal/mol and K_i of 5.37 μ M as indicated. Mannose was also bound with a ΔG of -4.89 kcal/mol and a low affinity (261.35 μ M).

The same analysis of the SARS-CoV-2 Delta spike protein (7TEW) revealed binding to A2FN-glycan with a ΔG of -7.79 kcal/mol and a K_i of 1.94 μ M. 6G1-glycan bound with a ΔG of -6.97 kcal/mol and K_i of 7.75 μ M, and mannose bound to the Delta variant with ΔG equal to -3.62 kcal/mol and high binding affinity of 2.22 μ M (Fig. 14.7 and Table 14.2).

Finally, SARS-CoV-2 Omicron spike protein (7WPB) bound to A2F *N*-glycan with a ΔG of -7.79 kcal/mol and K_i of 1.94 μ M, 6G1-glycan was bound with a ΔG of -6.37 kcal/mol and a K_i of 21.24 μ M, and mannose was bound at -4.29 kcal/mol with a low affinity of 711.08 μ M (Fig. 14.8).

We next differentiated the number of H-bonds formed between the host glycan and the spike RBD with the reference spike protein (6LZG). This showed that A2F *N*-glycan formed six H-bonds, 6 G1-glycan formed five H-bonds, and mannose had two H-bonds. The Delta spike protein (7TEW) interacts with A2F *N*-glycan forming six H-bonds, while interaction with 6 G1-glycan gave three H-bonds and mannose had five H-bonds. Lastly, the Omicron spike protein (7WPB) interacted with A2F

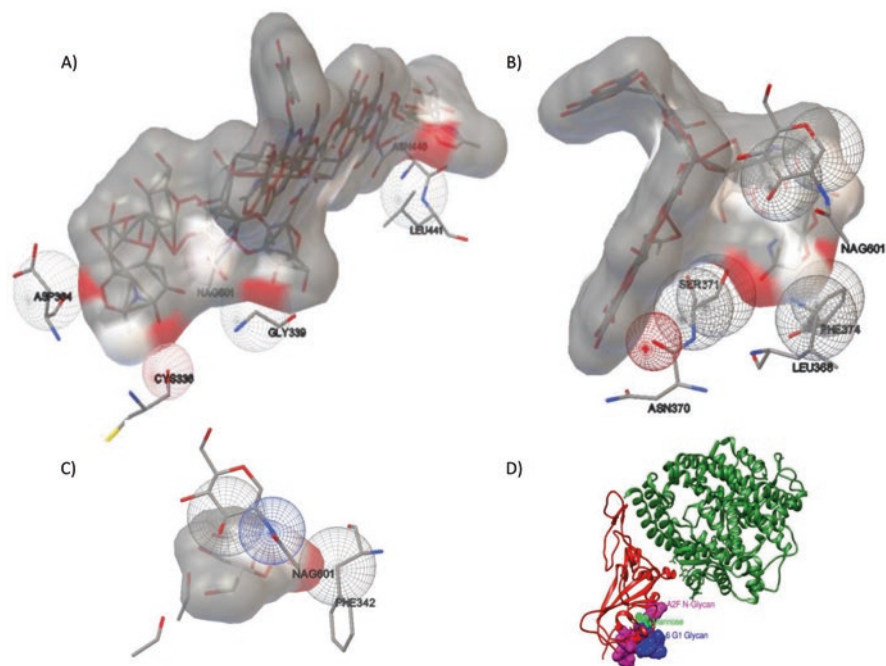


Fig. 14.6 2D representation of (a) A2F *N*-glycan, (b) 6 G1 glycan, and (c) mannose structures of 6LZG SARS-CoV-2, and (d) 3D representation of 6LZG SARS-CoV-2 with glycan complexes

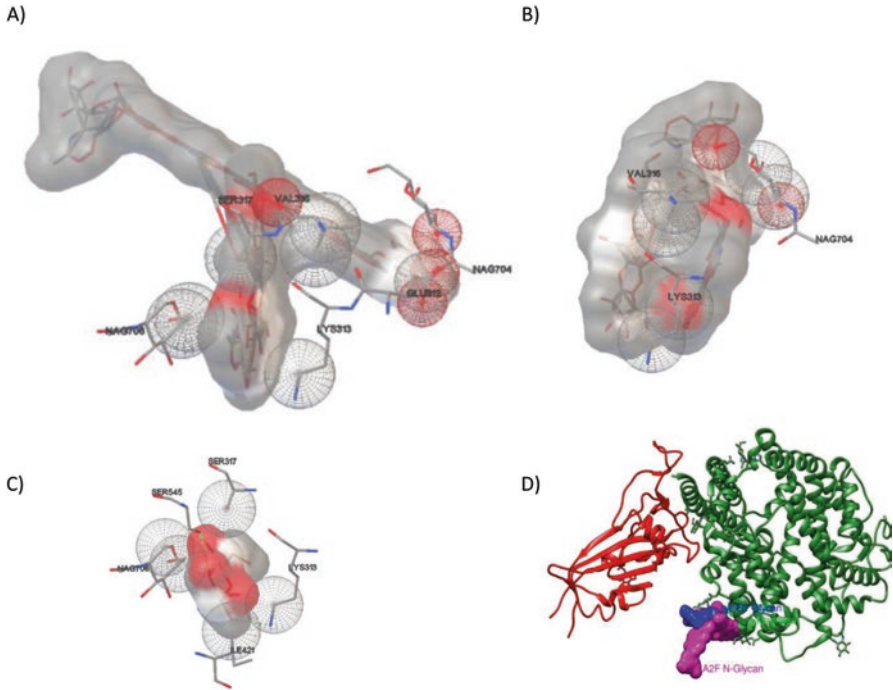


Fig. 14.7 2D representation of (a) A2F *N*-glycan, (b) 6 G1 glycan, and (c) mannose structures of 7TEW SARS-CoV-2, and (d) 3D representation of 7TEW SARS-CoV-2 with glycan complexes

N-glycan forming four H-bonds, while interaction with 6 G1-glycan gave three H-bonds and mannose had six H-bonds.

In summary, the analysis revealed that A2F *N*-glycan had the lowest binding affinity for the Delta and Omicron spike protein RBD sites. However, A2F *N*-glycan and 6 G1-glycan are bound with a lower free energy at RBD sites for all spike proteins compared to mannose. 6 G1-glycan had marginally lower affinity for the Omicron spike RBD (21.24 μM) compared to the Wuhan strain and the Delta variant. In contrast, mannose is bound to the Delta variant with markedly higher affinity (2.22 μM) compared to the original strain ($K_i = 261.35 \mu\text{M}$) and the Omicron variant ($K_i = 711.08 \mu\text{M}$).

4 Conclusions

The sequence similarity and dissimilarity approaches helped us to increase our understanding of how the SARS-CoV-2 variants achieve different binding, infectivity, and transmission properties in host cells. In the first part of the study, we identified key lncRNAs that could play a role in these transgression effects, and in the

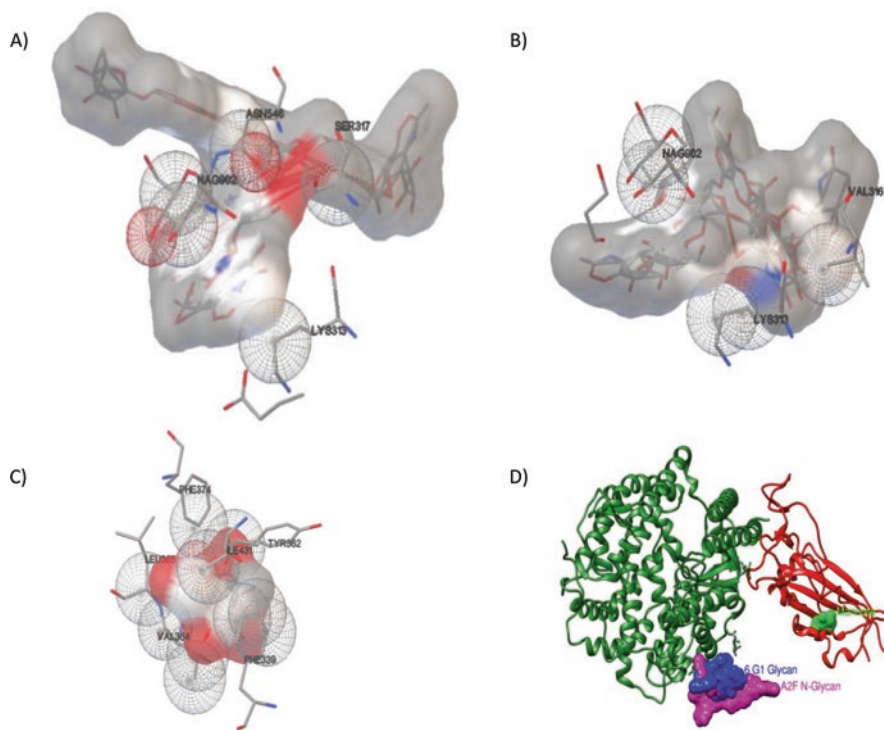


Fig. 14.8 2D representation of (a) A2F *N*-glycan, (b) 6 G1 glycan, and (c) mannose structures of 7WPB SARS-CoV-2, and (d) 3D representation of 7WPB SARS-CoV-2 with glycan complexes

second part, we focused on the sequence differences in spike proteins from the Delta and Omicron variants with regard to glycan binding in the host. Taken together, the findings revealed that the sequence differences in the variants of concern can affect glycosylation of the SARS-CoV-2 spike and host proteins which, in turn, can impact on the various transgression pathways. For example, such changes could increase infectivity by enhancing interactions with the ACE2 receptor or block the effect of neutralizing antibodies by disrupting their binding to the virus. In the current study, we examined the effects on three *N*-glycan structures (A2F, 6-G1, and high mannose) which differed with respect to fucosylation and terminal sugar composition, and showed differential binding with the SARS-CoV-2 Delta and Omicron spike RBDs compared to that of the original Wuhan strain. We suggest that the methods described in this study could be used to predict the virulence and transmissibility of new SARS-CoV-2 variants as these emerge. This would enable implementation of appropriate response measures and help to prepare us for the next pandemic.

Acknowledgments The authors gratefully acknowledge Revered Amma, Mata Amritanandamayi, Chancellor of Amrita Vishwa Vidyapeetham.

Author Contributions NN visualized and performed the major structural docking analysis; AM, HH, and AJ along with AR, MJ, PA, SR, and AAN wrote the first draft. All other authors contributed with lateral texts. PS, AA, and RS mentored the project. PS proofread the manuscript before all authors agreed to it.

References

1. Zhao P, Praissman JL, Grant OC, et al (2020) Virus-receptor interactions of glycosylated SARS-CoV-2 spike and human ACE2 receptor. *Cell Host Microb* 28(4):586–601.e6. <https://doi.org/10.1016/j.chom.2020.08.004>
2. Bagdonaite I, Wandall HH (2018) Global aspects of viral glycosylation. *Glycobiology* 28(7):443–467
3. Vigerust DJ, Shepherd VL (2007) Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol* 15(5):211–218
4. Seabright GE, Doores KJ, Burton DR, Crispin M (2019) Protein and Glycan Mimicry in HIV Vaccine Design. *J Mol Biol* 431(12):2223–2247
5. ACE2_HUMAN. <https://www.uniprot.org/uniprotkb/Q9BYF1/entry>
6. TMPS2_HUMAN. <https://www.uniprot.org/uniprotkb/O15393/entry>
7. Murin CD, Wilson IA, Ward AB (2019) Antibody responses to viral infections: a structural perspective across three different enveloped viruses. *Nat Microbiol* 4(5):734–747
8. Alter G, Ottenhoff THM, Joosten SA (2018) Antibody glycosylation in inflammation, disease and vaccination. *Semin Immunol* 39:102–110
9. Varki A (2017) Biological roles of glycans. *Glycobiology* 27(1):3–49
10. Talabnin K, Talabnin C, Ishihara M, Azadi P (2018). Increased expression of the high-mannose M6N2 and NeuAc3H3N3M3N2F tri-antennary N-glycans in cholangiocarcinoma. *Oncol Lett* 15(1):1030–1036
11. Schwedler C, Grzeski M, Kappert K, et al (2022) Coronavirus Disease 2019-Related Alterations of Total and Anti-Spike IgG Glycosylation in Relation to Age and Anti-Spike IgG Titer. *Front Microbiol* 13:775186. <https://doi.org/10.3389/fmicb.2022.775186>
12. Alijotas-Reig J, Esteve-Valverde E, Belizna C, et al (2020) Immunomodulatory therapy for the management of severe COVID-19. Beyond the anti-viral therapy: A comprehensive review. *Autoimmun Rev* 19(7):102569. doi: <https://doi.org/10.1016/j.autrev.2020.102569>
13. Tufan A, Avanoğlu Güler A, Matucci-Cerinic M (2020) COVID-19, immune system response, hyperinflammation and repurposing antirheumatic drugs. *Turk J Med Sci* 50(SI-1):620–632
14. Soy M, Keser G, Atagündüz P, et al (2020) Cytokine storm in COVID-19: pathogenesis and overview of anti-inflammatory agents used in treatment. *Clin Rheumatol* 39(7):2085–2094
15. Bindoli S, Felicetti M, Sfriso P, et al (2020) The amount of cytokine-release defines different shades of Sars-Cov2 infection. *Exp Biol Med* (Maywood) 245(11):970–976
16. Rabaan AA, Al-Ahmed SH, Muhammad J, et al (2021) Role of Inflammatory Cytokines in COVID-19 Patients: A Review on Molecular Mechanisms, Immune Functions, Immunopathology and Immunomodulatory Drugs to Counter Cytokine Storm. *Vaccines* (Basel) 9(5):436. doi: <https://doi.org/10.3390/vaccines9050436>
17. Ramasamy S, Subbian S (2021) Critical Determinants of Cytokine Storm and Type I Interferon Response in COVID-19 Pathogenesis. *Clin Microbiol Rev* 34(3):e00299–20. doi: <https://doi.org/10.1128/CMR.00299-20>
18. Shajahan A, Pepi LE, Rouhani DS, et al (2021) Glycosylation of SARS-CoV-2: structural and functional insights. *Anal Bioanal Chem* 413(29):7179–7193
19. Land A, Braakman I (2001) Folding of the human immunodeficiency virus type 1 envelope glycoprotein in the endoplasmic reticulum. *Biochimie* 83(8):783–790

20. Elphick GF, Querbes W, Jordan JA, et al (2004) The human polyomavirus, JCV, uses serotonin receptors to infect cells. *Science* 306(5700):1380–1383
21. Huang C, Tan Z, Zhao K, et al (2021) The effect of N-glycosylation of SARS-CoV-2 spike protein on the virus interaction with the host cell ACE2 receptor. *iScience*. 24(11):103272. <https://doi.org/10.1016/j.isci.2021.103272>
22. Haniu M, Horan T, Arakawa T, et al (1996) Disulfide structure and N-glycosylation sites of an extracellular domain of granulocyte-colony stimulating factor receptor. *Biochemistry* 35(40):13040–13046
23. Li W, Moore MJ, Vasilieva N, et al (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426(6965):450–454
24. Gates B (2020) Responding to Covid-19 - a once-in-a-century pandemic? *New Engl J Med* 382(18):1677–1679
25. Mehdipour AR, Hummer G (2021) Dual nature of human ACE2 glycosylation in binding to SARS-CoV-2 spike. *Proc Natl Acad Sci USA* 118(19):e2100425118. <https://doi.org/10.1073/pnas.2100425118>
26. Allen JD, Watanabe Y, Chawla H, et al (2021). Subtle Influence of ACE2 Glycan Processing on SARS-CoV-2 Recognition. *J Mol Biol*433(4):166762. <https://doi.org/10.1016/j.jmb.2020.166762>
27. Clausen TM, Sandoval DR, Spliid CB, et al (2020). SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *Cell*183(4):1043–1057.e15. <https://doi.org/10.1016/j.cell.2020.09.033>
28. Baker A, Richards SJ, Congdon T, et al (2021) Correction to: The SARS-COV-2 Spike Protein Binds Sialic Acids, and Enables Rapid Detection in a Lateral Flow Point of Care Diagnostic Device. *ACS Cent Sci* 7(2):379–380
29. Guttman M, Amit I, Garber M, et al (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458(7235):223–227
30. Peng X, Gralinski L, Armour CD, et al (2010) Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling. *mBio* 1(5):e00206-10. <https://doi.org/10.1128/mBio.00206-10>
31. Shukla N, Prasad A, Kanga U, et al (2021) SARS-CoV-2 transgressing LncRNAs uncovers the known unknowns. *Physiol Genomics* 53(10):433–440
32. Vishnubalaji R, Shaath H, Alajez NM (2020) Protein Coding and Long Noncoding RNA (lncRNA) Transcriptional Landscape in SARS-CoV-2 Infected Bronchial Epithelial Cells Highlight a Role for Interferon and Inflammatory Response. *Genes (Basel)* 11(7):760. <https://doi.org/10.3390/genes11070760>
33. Turjya RR, Khan MA, Mir Md Khademul Islam AB (2020) Perversely expressed long non-coding RNAs can alter host response and viral proliferation in SARS-CoV-2 infection. *Future Virol* 15(9):577–593
34. Jin G, Sun J, Isaacs SD, et al (2011) Human polymorphisms at long non-coding RNAs (lncRNAs) and association with prostate cancer risk. *Carcinogenesis* 32(11):1655–1659
35. Aznaourova M, Schmerer N, Schmeck B, Schulte LN (2020) Disease-Causing Mutations and Rearrangements in Long Non-coding RNA Gene Loci. *Front Genet* 11:527484. <https://doi.org/10.3389/fgene.2020.527484>
36. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/>. Accessed May 30, 2022
37. Molview. <https://molview.org/>
38. National Library of Medicine; National Center for Biotechnology Information; PubChem. <https://pubchem.ncbi.nlm.nih.gov>
39. Clustal Omega; Multiple Sequence Alignment. www.ebi.ac.uk/Tools/msa/clustalo/. Accessed May 20, 2022
40. Sievers F, Higgins DG (2014) Clustal omega. *Curr Protoc Bioinformatics* 48:3.13.1–16
41. ITOL; Interactive Tree of Life. <https://itol.embl.de/>. Accessed May 18, 2022

42. Pangolin COVID-19 Lineage Assigner. <https://pangolin.cog-uk.io/>. Accessed May 28, 2022
43. Noncode. <http://www.noncode.org/>. Accessed May 20, 2022
44. Datawrapper. <https://www.datawrapper.de/>. Accessed May 24, 2022
45. Evans JP, Qu P, Zeng C, et al (2022) Neutralization of the SARS-CoV-2 Deltacron and BA.3 Variants. *N Engl J Med* 386(24):2340–2342