

Integrative multiomics and in silico analysis revealed the role of ARHGEF1 and its screened antagonist in mild and severe COVID-19 patients

Prakash Jha¹ | Prithvi Singh² | Shweta Arora³ | Armiya Sultan⁴  | Arnab Nayek⁵ | Kalaiarasan Ponnusamy⁶ | Mansoor Ali Syed³ | Ravins Dohare²  | Madhu Chopra¹

¹Laboratory of Molecular Modeling and Anticancer Drug Development, Dr. B. R. Ambedkar Center for Biomedical Research, University of Delhi, New Delhi, India

²Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India

³Department of Biotechnology, Translational Research Lab, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi, India

⁴Department of Biosciences, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi, India

⁵Department of Biochemistry, All India Institute of Medical Sciences, New Delhi, India

⁶Synthetic Biology Lab, School of Biotechnology, Jawaharlal Nehru University, New Delhi, India

Correspondence

Ravins Dohare, Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi 110025, India.
Email: ravinsdohare@gmail.com

Madhu Chopra, Laboratory of Molecular Modeling and Anticancer Drug Development, Dr. B. R. Ambedkar Center for Biomedical Research, University of Delhi, New Delhi 110007, India.
Email: mchopradu@gmail.com

Funding information

Indian Council of Medical Research, Grant/Award Numbers: 5/3/8/13/ITR-F/2018-ITR, BMI/11(89)/2020; Science and Engineering Research Board, Grant/Award Number: DST/INSPIRE/03/2016/000026

Abstract

COVID-19 is a sneaking deadly disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The rapid increase in the number of infected patients worldwide enhances the exigency for medicines. However, precise therapeutic drugs are not available for COVID-19; thus, exhaustive research is critically required to unscramble the pathogenic tools and probable therapeutic targets for the development of effective therapy. This study utilizes a chemogenomics strategy, including computational tools for the identification of viral-associated differentially expressed genes (DEGs), and molecular docking of potential chemical compounds available in antiviral, anticancer, and natural product-based libraries against these DEGs. We scrutinized the messenger RNA expression profile of SARS-CoV-2 patients, publicly available on the National Center for Biotechnology Information–Gene Expression Omnibus database, stratified them into different groups based on the severity of infection, superseded by identification of overlapping mild and severe infectious (MSI)-DEGs. The profoundly expressed MSI-DEGs were then subjected to trait-linked weighted co-expression network construction and hub module detection. The hub module MSI-DEGs were then exposed to enrichment (gene ontology + pathway) and protein–protein interaction network analyses where Rho guanine nucleotide exchange factor 1 (ARHGEF1) gene conjectured in all groups and could be a probable target of therapy. Finally, we used the molecular docking and molecular dynamics method to identify inherent

Prakash Jha and Prithvi Singh contributed equally to this study as co-first authors.

hits against the ARHGEF1 gene from antiviral, anticancer, and natural product-based libraries. Although the study has an identified significant association of the ARHGEF1 gene in COVID19; and probable compounds targeting it, using in silico methods, these targets need to be validated by both in vitro and in vivo methods to effectively determine their therapeutic efficacy against the devastating virus.

KEYWORDS

ARHGEF1, gene co-expression network, molecular docking, molecular dynamics, MSI-DEGs, SARS-CoV-2

1 | INTRODUCTION

Eighteen years after the outbreak of severe acute respiratory syndrome (SARS) caused by SARS-coronavirus (SARS-CoV), the world is now facing a new dare posted by version 2.0 of CoV causing similar symptoms like SARS emerged from Wuhan city, China, which was initially named SARS-CoV-2 and later renamed by the World Health Organization as COVID-19 on February 11, 2020.^{1–3} As of July 1, 2021, the total number of laboratory-confirmed cases of COVID-19 infection in the world were 182 988 210, including 3 962 976 associated deaths.⁴ The first COVID-19 patient-reported from Seafood Market in Wuhan was possibly infected by the direct or indirect transmission of COVID-19 virus from bats or pangolins and later the epidemic spread to the whole world via human-to-human transmission.^{5,6} The spread toward Asian countries, especially in India, has further raised the concern to the wider population.⁷

CoV belongs to the largest group of viruses and comes under the *Coronaviridae* family and the *Coronavirinae* subfamily, found in birds and mammals. CoVs are classified into four genera: α , β , δ , and γ where COVID-19 refer to β -CoV which also consists of Middle East respiratory syndrome (MERS)-CoV and SARS-1.^{8(p. 2),9}

Images from transmission electron microscopy revealed that CoV gives a crown-like appearance with spike proteins projecting out from the spherical virion surface. COVID-19 shares about 50% similarity with MERS whereas approximately 79% with SARS-CoV indicates that SARS-CoV-2 is closer to SARS-CoV. All the human CoV consists of 3'-polyadenylated tails which encodes structural proteins (nucleocapsid protein, membrane proteins, spike proteins, and envelope proteins) responsible for maintaining viral life cycle, and 5'-methylated cap region is for nonstructural proteins, which are important for replication of the virus.^{10(p. 19),11,12}

Among various human CoVs, SARS-CoV-2 caused a severe mortality rate close to 1.38, and the most common symptoms in positive cases were cough, fever, headache, fatigue, diarrhea, and blood-stained mucous. In the

current scenario, the characterization of novel biomarkers which may play an important role in prediction and observation in the status of SARS-CoV-2 disease is required.^{13,14(p. 2)}

Many studies related to SARS-CoV-2 have been published in these 2 years covering aspects like high-throughput transcriptomic analysis, network-based dynamics, artificial intelligence-based drug repurposing, and novel therapeutic compounds identification followed by molecular dynamics (MD) validation. Our study is novel as it presents an age-based weighted network of SARS-CoV-2 transcriptomic dataset for robust identification of highly correlated hub module along with therapeutic targets identification thus enabling us to understand the disease mechanism at the molecular and systems level.

In the present study, SARS-CoV-2 associated messenger RNA (mRNA) expression profile was retrieved from National Center for Biotechnology Information (NCBI)-Gene Expression Omnibus (GEO) followed by identification of differentially expressed genes (DEGs) in different infection severity groups. The overlapping mild and severe infectious (MSI)-DEGs present in both groups were subjected to trait-linked weighted gene co-expression network (WGCN) construction and hub module detection. The hub module MSI-DEGs were forwarded to protein-protein interaction (PPI) network construction and enrichment analyses for predicting potential therapeutic targets. Lastly, we applied molecular docking and MD method against Rho guanine nucleotide exchange factor 1 (ARHGEF1) gene¹⁵ to discover potential hits for SARS-CoV-2 from antiviral, anticancer, and natural product-based libraries. The novelty of our work involves an amalgamation of transcriptomic, network analysis, PPI, and enrichment analyses of COVID-19 peripheral blood mononuclear cell (PBMC) patient samples for investigation of human target gene which might be responsible for controlling the severity of SARS-CoV-2 infection, unlike earlier research work.^{16,17} The potential candidate hits can be further selected for in vitro and in vivo studies in future work to speed up the drug discovery against SARS-CoV-2.

2 | MATERIALS AND METHODS

2.1 | SARS-CoV-2 microarray data collection and differential expression analysis

NCBI-GEO (<https://www.ncbi.nlm.nih.gov/geo/>)¹⁸ was queried to extract SARS-CoV-2-related mRNA expression profile with “COVID-19” and “SARS-CoV-2” being used as appropriate keywords. Our searching criteria for picking a suitable dataset was as follows: (1) the dataset must be “*expression profiling by array*” type and its samples must be from “*Homo Sapiens*”; (2) the dataset must be having both healthy and COVID-19 (varying severity levels) patient samples; (3) the dataset must be having diverse age group samples; (4) the dataset must be having processed and raw microarray data; and (5) the dataset submission date must be within 1 year. Any studies devoid of nonhuman samples, case reports, review articles, abstracts, and cell-line-based experimental study designs were excluded. Series matrix expression file of the chosen dataset was downloaded using GEO query R package.¹⁹ Mapping of probe IDs to their corresponding HUGO Gene Nomenclature Committee (HGNC) symbols was performed using feature data stored in the expression set object of dataset. The averaging expression value of genes mapping to more than one probe IDs was done to avoid redundancy. The COVID-19 patient samples in the dataset were bifurcated into groups based on their infection severity levels followed by the elimination of 50% low variance genes.²⁰ HGNC multisymbol checker (<https://www.genenames.org/tools/multi-symbol-checker/>) was utilized and only officially approved HGNC symbols were retained. A two-sample statistical *t*-test was utilized for calculating the *p*-value of each gene between control and COVID-19 samples (infection severity group-based) followed by obtaining their \log_2 (fold change) and Benjamini–Hochberg (BH) *p*-values utilizing the Limma R package.²¹ The genes corresponding to BH $p < 0.01$ and $|\log_2(\text{fold change})| > 2$ were regarded as DEGs. DEGs with $\log_2(\text{fold change}) > 2$ and $\log_2(\text{fold change}) < -2$ were bifurcated as up- and downregulated, respectively. The DEGs overlapping between different infection severity groups was considered as MSI-DEGs.

2.2 | WGCN construction and hub module identification

Weighted Gene Co-Expression Network Analysis (WGCNA) R package²² was utilized for establishing a WGCN from SARS-CoV-2 specific MSI-DEGs and identifying representative module genes having a high correlation with clinical characteristics (i.e., age). All the MSI-DEGs along

with their samples were passed through the *goodSamplesGenes* function to eliminate any missing values. The samples were clustered thereafter to eliminate any outliers. The age information of all samples was also taken into consideration before identifying modules. The *pickSoft-Threshold* function assisted in selecting a suitable soft-thresholding power (β) to which co-expression similarity will be raised for computing adjacency. In general, a suitable value of β was chosen in compliance with the approximate scale-free topology criterion. However, in some cases, the data does not follow appropriate scale-free topology due to the presence of a strong driver which makes a subgroup of samples globally distinct from the rest. This difference causes an elevated correlation amongst a large cluster of genes that invalidates the assumption of the scale-free topology approximation. In such cases, a suitable value of β needs to be chosen based on the sample size in accordance with the WGCNA guidelines (<https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/faq.html>) to make the resulting network conservative. For reduction in noise and false associations, the weighted adjacency matrix was converted into a topological overlap matrix (TOM) followed by computation of corresponding dissimilarity (diss-TOM). The *hclust* function was utilized to generate a hierarchical clustering tree (dendrogram) of genes in consideration with the dissTOM measure. To identify densely interconnected highly co-expressed gene patterns (i.e., modules) from the branches of a tree, a dynamic tree cut algorithm was incorporated. Module eigengene (ME) and dissimilarity measure between MEs were computed to merge the modules with highly co-expressed genes. ME dendrogram was checked based on Pearson correlation for merging multiple modules with comparable expression profiles. Correlation-based absolute module significance (i.e., average gene significance [GS] of participating genes in a given module) with the age of samples followed by module membership (MM) (correlation of the ME and the gene expression profile) for all modules were computed. Intramodular connectivity (k.in) can be explained as a measure of network connectivity with respect to nodes or genes of a specific module. The correlation of MM versus GS, GS versus k.in, and MM versus k.in was used to identify the most significant associations and the module having significantly highest correlation with age was chosen as the hub module.

2.3 | PPI network construction, pathway, and gene ontology term enrichment analyses

The MSI-DEGs present in our hub module were subjected to PPI network construction using Search Tool for the Retrieval of Interacting Genes (STRING,

<https://string-db.org/>) v11.5 web-based tool.²³ This PPI network was formed at the highest confidence (corresponding to interaction score >0.9) and afterward visualized using Cytoscape v3.9.0.²⁴ Pathway and GO term enrichment data for MSI-DEGs present in the hub module was compiled using various libraries (i.e., reactome for pathways and gene ontology [GO]-biological process [BP], GO-molecular function [MF], and GO-cellular compartment terms) available within Enrichr database.²⁵ Top 10 pathways and GO terms corresponding to $p < 0.05$ were regarded as statistically significant ones. The MSI-DEGs overlapping between constructed highest confidence PPI network, top 10 significant pathways, and GO terms were regarded as SARS-CoV-2 hub MSI-DEG(s), respectively.

2.4 | Identification of binding sites and preparation of protein

RhoGEF1 enzyme encoded by ARHGEF1 gene was the hub driver MSI-DEG in SARS-CoV-2 infectious process and was selected as a promising target for COVID-19. The binding sites of the identified protein (RhoGEF1), PDB ID: 3ODO, were searched using the machine learning method on PrankWeb server (<https://prankweb.cz/>)²⁶ due to very little information available about protein binding site. Before this, the protein structure was prepared using the automatic protein preparation module of Discovery Studio (DS) 2019. Hydrogen atoms were added followed by side-chain refinement and protonation states were determined. A receptor grid sphere with coordinates $X = 46.5266$, $Y = -41.5728$, $Z = 26.6118$, and radii = 12 Å was generated based on PrankWeb server information.

2.5 | High-throughput virtual screening using LibDock

LibDock²⁷ is a high-throughput screening program in which both ligands, as well as protein, were rigid. It enumerates the hotspots at the binding site of the protein by using apolar and polar probes which were used to match the ligands to form favorable interactions. Natural-Product-Like Compound Library (900 compounds), Anticancer Screening Library (6300 compounds), and Antiviral Compound Library (13 700 compounds) were retrieved from Life Chemicals (<https://lifechemicals.com/>) and COVID-19 related library (35 compounds) to screen ARHGEF1 inhibitors. Using the defined coordinates site, an active site for molecular docking was generated. Structure-based virtual screening

was executed by docking all the compound libraries at the prepared active site by using the LibDock module of DS 2019. The LibDock score was used as a criterion to rank the docked poses and grouped by name.

2.6 | Ligand preparation and ADMET prediction

The filtered ligands from LibDock docking were selected for the calculation of the ADMET properties (absorption, distribution, metabolism, excretion, and toxicity) and Ames mutagenicity using ADMET²⁸ and TOKAT²⁹ module of DS 2019. These compounds were further filtered based on their molecular properties using Lipinski's and Veber rule²⁸ implemented in DS 2019. Filtered compounds were subjected to energy minimization by using the steepest descent (5000 cycles) followed by the conjugate gradient (5000 cycles) to satisfy the minimum criteria of root mean square gradient of 0.001.³⁰

2.7 | Molecular docking

All the ligands that were successfully filtered out were submitted for molecular docking using the CDOCKER module of DS 2019 which is a CHARMM-based MD simulated docking tool in which ligands were allowed to be flexible while receptor to remain rigid.³¹ Crystal structure of ARHGEF1 (PDB ID: 3ODO)³² with the same binding site was selected for molecular docking. By retaining all the default parameters, filtered ligands were docked into the binding site of ARHGEF1. Different conformations for each ligand were generated and investigated based on CDOCKER energy. Next, we generated the nonbond interactions between the ARHGEF1, and ligand conformations obtained from CDOCKER docking with the help of Analyse ligand poses module of DS 2019.³³ Detailed knowledge of nonbond interactions is important in structure-based design, as it helps to select the favorable interactions between ligands and their receptor.

2.8 | Flexible docking

To cross-validate the identified binding site and molecular docking process, a flexible docking method was used. It may also generate more correct or accurate ARHGEF1 complexes which may be used as a starting structure for MD simulation. The top 15 complexes were selected after analyzing the nonbond interactions for redocking by using the flexible docking module of DS

2019 which docks the hit ligands with induced-fit protein optimization while simulating the protein flexibility. Favorable residues from analyzing ligand pose protocol were defined as flexible residues. The receptor conformations (50) were generated using ChiFlex³⁴ and subsequently, 255 conformations were generated for ligand. Next, ligand docking was done by the LibDock algorithm followed by side-chain refinement using ChiRotor.³⁴ The final refinement of complexes was done by using CDOCKER under the CHARMM forcefield.³⁵

2.9 | MD study

The top three hit-complexes were used as starting structures for MD simulation by using Gromacs 4.5.6 package³⁶ and Gromos96 forcefield.³⁶ Additionally, the top two hit complexes from the COVID-19 related library were also included in this study. The topology and parameter file for ligands were generated on the GlycoBioChem PRODRG2 Server (<http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg>).³⁷ The complexes were solvated with a cubic box of a simple point charge with 1 Å margin. Systems were next neutralized by adding Na⁺ and Cl⁻ ions to balance the system charges. Each complex was energy minimized with 10 000 steps of the steepest descent algorithm to remove any steric clashes. Then the system was equilibrated with 100 ps of NVT followed by 100 ps of NPT at a constant temperature of 300 K and a constant pressure of 1 atm. All bonds were constrained with LINC algorithm³⁸ and long-range electrostatics were calculated with the particle-mesh Ewald algorithm.³⁹ Finally, 50 ns of MD run was performed and snapshots were saved for every 5 ps. Final trajectories were analyzed in pymol and Gromacs tools (*g_energy*, *g_rmsf*, *g_rmsd*, *g_gyrate*, and *g_hhbond*) to calculate potential energy, root-mean-square fluctuation (RMSF), root-mean-square deviation (RMSD), the radius of gyration, and h-bond analysis.

2.10 | Molecular mechanics Poisson–Boltzmann surface area

The binding free energies for docked complexes were calculated using molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) to revalidate the MD simulation studies using the *g_mmpbsa* tool. We performed the calculations from the last 5 ns of each MD run and used all the set parameters. Next, the binding energies of the best three ligands-complexes were enumerated using the following equations as reported earlier⁴⁰:

$$\Delta G_{\text{bind,pred}} = \Delta G_{\text{complex}} - [\Delta G_{\text{Rec}} + \Delta G_{\text{lig}}], \quad (1)$$

$$G = E_{\text{gas}} + G_{\text{sol-TS}}, \quad (2)$$

$$E_{\text{gas}} = E_{\text{int}} + E_{\text{ele}} + E_{\text{vdw}}, \quad (3)$$

$$G_{\text{sol}} = G_{\text{PB(GB)}} + G_{\text{sol-np}}, \quad (4)$$

$$G_{\text{sol-np}} = \gamma \text{SAS}. \quad (5)$$

3 | RESULTS

3.1 | SARS-CoV-2 microarray data collection and differential expression analysis

On the basis of searching criteria, we chose SARS-CoV-2 mRNA expression profile possessing accession number GSE164805. It comprised a total of 15 PBMC patient samples (i.e., 5 controls and 10 COVID-19 [5 mild + 5 severe]). Mapping probe IDs to their corresponding genes was followed by averaging expression values corresponding to duplicates which led to 44 822 unique genes. The COVID-19 patient samples in the dataset were bifurcated into MSI groups along with common controls in both groups leading to 10 samples in each group (i.e., Group 1: 5 mild + 5 controls and Group 2: 5 severe + 5 controls). The variance of each gene across samples in both groups was computed leaving 22 411 high variance genes in both the groups. After passing these genes through HGNC multisymbol checker, a total of 10 729 and 10 518 officially approved HGNC symbols were left in Groups 1 and 2, respectively. A total of 837 and 1935 genes were differentially expressed corresponding to BH $p < 0.01$ and $|\log_2(\text{fold change})| > 2$ in Groups 1 and 2. Amongst these DEGs, 427 and 1129 were downregulated while 410 and 806 were upregulated in Groups 1 and 2. Volcano plots showing the distribution of significant (colored points) as well as nonsignificant (black colored points) genes in Groups 1 and 2 were shown in Figure 1A,C. Figure 1B,D shows an annotation heatmap of top 10 up and downregulated DEGs in Groups 1 and 2. The sample type, gender, and age annotation bars were placed at the top of the heatmap. The controls clustered distinctly from mild and severe samples in both groups as evidenced from the heatmap. FST [$\log_2(\text{fold change}) = -5.96$] and TEX101 [$\log_2(\text{fold change}) = 10.73$] were most down and upregulated in Group 1. Whereas ADAT1 [$\log_2(\text{fold change}) = -6.75$] and TEX101 [$\log_2(\text{fold change}) = 9.96$] were most down and upregulated in Group 2. Among the top 10 up- and downregulated DEGs, ADAT1, HBD, and FST were commonly downregulated in both groups. Also, TEX101, FAM177B, and OR51E1 were commonly upregulated in both groups. TEX101 was the highest upregulated DEG in both groups. Group 1 samples constituted 20% females and

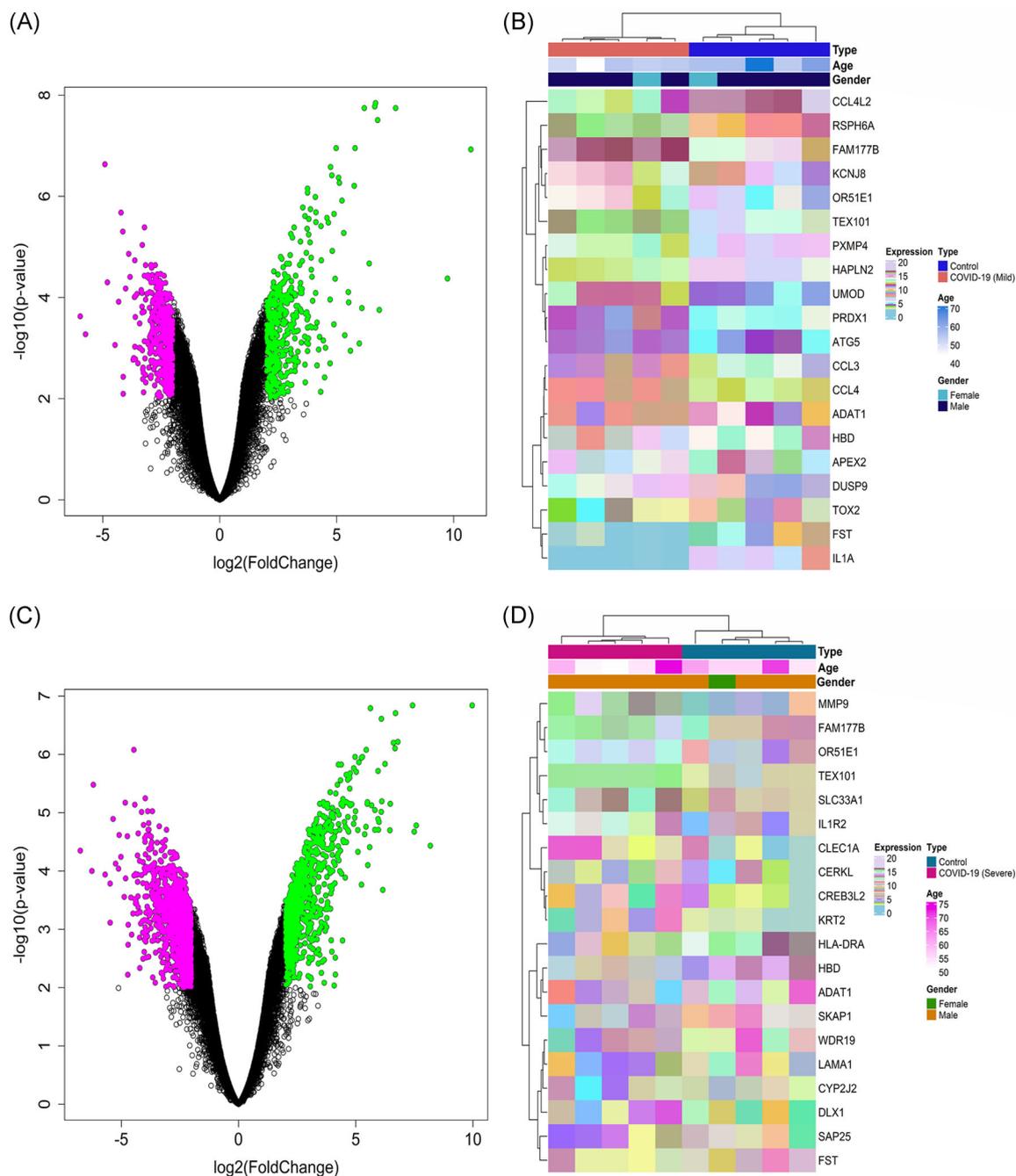


FIGURE 1 (A) Volcano plot showing the distribution of DEGs (downregulated: 427 + upregulated: 410) and nonsignificant (9892) genes across control and mild COVID-19 samples (i.e., Group 1). (B) Annotation heatmap showing the distribution top 10 up- and downregulated DEGs across Group 1 samples. The expression value of DEGs (in rows) is normalized across all the samples (in columns). Cluster dendrograms representing Euclidean distance-based hierarchical clustering for both rows and columns are presented along the left and top sides of the plot. Sample type (medium blue for controls and carmine for mild), gender (eastern blue for females and midnight blue for males), and age annotation bars are placed at the top of the heatmap. (C) Volcano plot showing the distribution of DEGs (downregulated: 1129 + upregulated: 806) and nonsignificant (8583) genes across control and severe COVID-19 samples (i.e., Group 2). (D) Annotation heatmap showing the distribution top 10 up- and downregulated DEGs across Group 2 samples. The expression value of DEGs (in rows) is normalized across all the samples (in columns). Cluster dendrograms representing Euclidean distance-based hierarchical clustering for both rows and columns are presented along the left and top sides of the plot. Sample type (Prussian blue for controls and ruby red for severe), gender (green for females and orange for males), and age annotation bars are placed at the top of the heatmap. The green- and magenta-colored points signify the up- and downregulated DEGs in volcano plots. While the black-colored points signify the nonsignificant genes. The x- and y-axes represent $\log_2(\text{fold change})$ and $-\log_{10}(p\text{-value})$. DEG, differentially expressed gene

80% males while Group 2 samples constituted 10% females and 90% males. Venn plot showing the 569 overlapping MSI-DEGs between both groups is shown in Figure S1A. Figure S1B shows a kernel density plot distribution of sample age for all 569 MSI-DEGs. Three samples (i.e., one control, one mild, and one severe) with age 54 were having the highest density as evidenced from the plot. All the samples in Group 2 were having an age ≥ 50 with 73 being the maximum age. Whereas all samples in Group 1 except one were had age ≥ 50 with 71 being the maximum age.

3.2 | WGCN construction and hub module identification

The expression data of all 569 MSI-DEGs along with their sample age information were used as an input in WGCNA. No obvious sample outliers were detected from the sample clustering dendrogram as shown in Figure S2. Our WGCN did not follow the scale-free topology criterion for choosing the appropriate value of β with respect to $R^2 \geq 0.8$. Hence, in accordance with the WGCNA guidelines, $\beta = 18$ (for less than 20 samples) was chosen. Hierarchical clustering tree and dynamic tree cut algorithm revealed a total of five color-coded gene modules (i.e., blue, turquoise, brown, green, and yellow) as shown in Figure 2A. The size of these five gene modules were as follows: blue = 109, turquoise = 231, brown = 102, green = 56, and yellow = 71. The multi-dimensional scaling plot of all modules in three scaling dimensions is shown in Figure 2B. There was no need for merging these modules due to the low merging height observed in the ME dendrogram. Figure S3A shows the gene co-expression network as a heatmap plot depicting TOM among all module genes. Figure 2C shows the Barplot of GS (correlated with age) with respect to each module. The GS values for the modules were as follows: blue = 0.14, turquoise = 0.27, brown = 0.15, green = 0.41, yellow = 0.23. MM versus GS correlation values along with their p -values for each module are shown in Table S1. The GS versus k.in values along with their p -values for each module is shown in Table S2. MM versus k.in values along with their p -values for each module are shown in Table S3. As evidenced from Tables S1–3, blue, yellow, and turquoise modules were eliminated from any further analysis since they were not statistically significant ($p > 0.05$). All these results indicated that the green module was the most promising and highly correlated (module significance = 0.41, GS versus MM = -0.49 , GS versus k.in = -0.4 , MM versus k.in = 0.43) as compared to the brown module. Therefore, the green module (comprising 56 MSI-DEGs) was chosen as our hub module. Figure 2D shows a scatterplot of GS for age with respect to MM in the green hub module. Figure S3B shows a scatterplot of GS for age with respect to MM in the brown

module. Figure S3C,D shows heatmap plots of green and brown module genes along with their corresponding ME levels.

3.3 | PPI network construction, pathway, and GO term enrichment analyses

All the 56 MSI-DEGs present in our hub module were given as an input to the STRING web-based tool where a total of 4 (3 upregulated + 1 downregulated) MSI-DEGs participated in the PPI network corresponding to STRING interaction score >0.9 . The PPI network as shown in Figure 3A comprised four nodes and four edges. Also, a total of 12, 14, and 10 out of 56 MSI-DEGs within our hub module participated in the top 10 significant pathways, GO-BP, and GO-MF terms. The most significant pathway, GO-BP, and GO-MF terms were G alpha (12/13) signaling events ($p = 1.21 \times 10^{-3}$), regulation of mitotic spindle checkpoint ($p = 1.15 \times 10^{-4}$), and ATP-dependent DNA helicase activity ($p = 2.77 \times 10^{-3}$). Sankey plot showing the association of top 10 significant GO-BP and GO-MF terms with corresponding 18 genes is shown in Figure 3B. Chord plot showing the association of 12 MSI-DEGs with the top 10 significant pathways is shown in Figure 3C. The width of interaction edges connecting the pathways and MSI-DEGs varied according to the p -values. PSMF1, ARHGEF9, GNG4, and AP2M1 appeared in a maximum number of pathways (i.e., 4) whereas F10, ABCG2, BRIP1, GJD3, RASAL2, and DUSP9 appeared in a minimum number of pathways (i.e., 1). Venn plot as shown in Figure 3D revealed ARHGEF1 as the overlapping hub MSI-DEG between PPI, pathway, and GO term (i.e., GO-BP, GO-MF) genesets.

3.4 | Identification of binding sites and preparation of protein

Since there is no crystal structure available for our predicted target ARHGEF1 with co-crystallized ligand and very little knowledge available about binding site, the refined or prepared structure of ARHGEF1 protein (PDB ID:3ODO)³² was examined on PrankWeb server²⁶ and topographic features of protein were assessed through P2Rank program. Out of the three predicted pockets, we selected pocket 1 based on rank score. The amino acids predicted in pocket 1 were Thr427, Ala430, His431, Met434, Arg551, Leu552, Asp556, Met557, Thr560, Gln563, Arg564, and Lys567. On the basis of these amino acids found in the pocket (score = 5.22) (Table S4 and Figure S4), we defined the binding site for structure-based virtual screening.

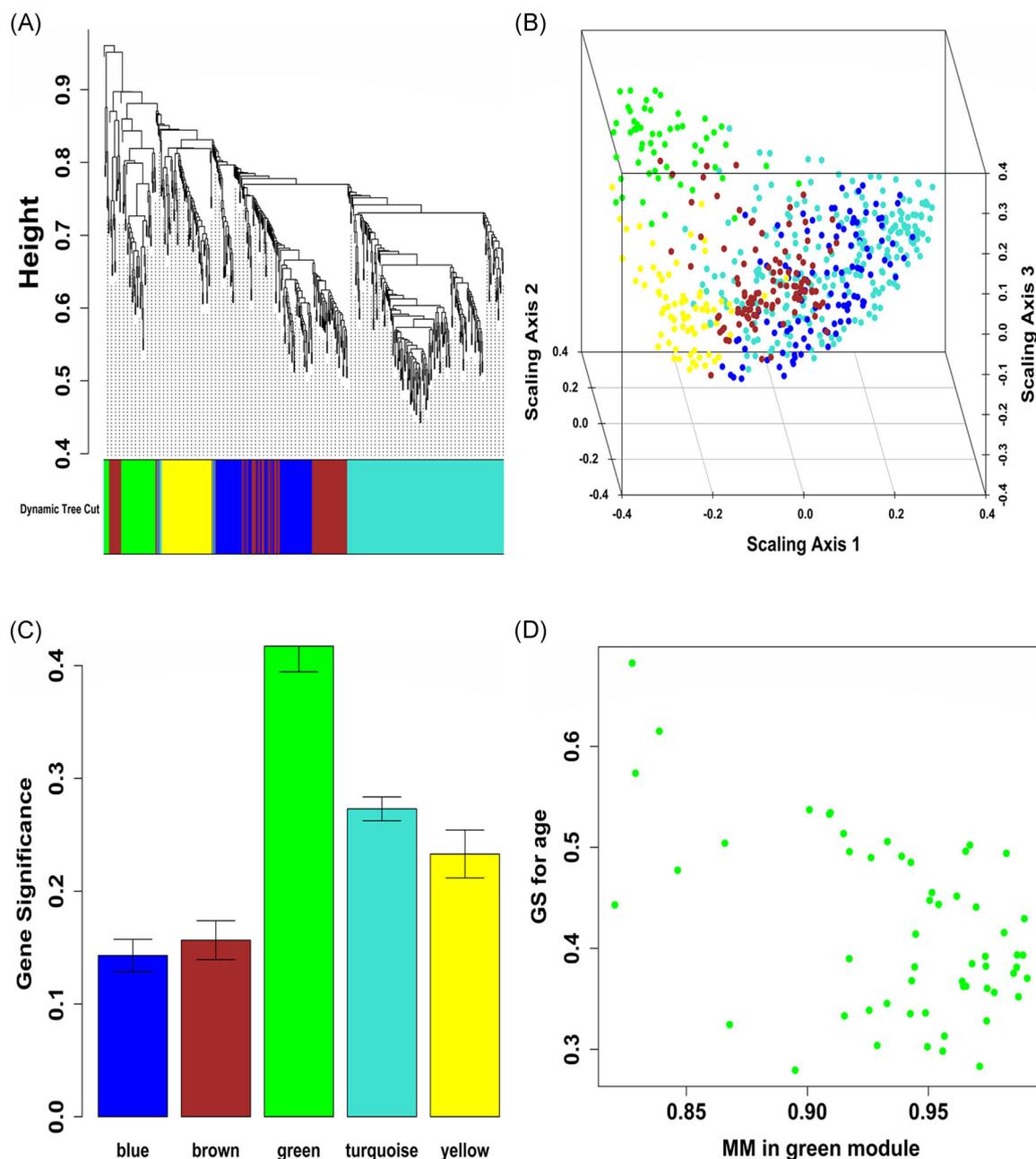


FIGURE 2 (A) Hierarchical clustering dendrogram of MSI-DEGs clustered based on the dissimilarity measure (dissTOM) and five color-coded modules (obtained using Dynamic Tree Cut). The modules contained highly similar expression profiles with the following sizes: turquoise (231), blue (109), brown (102), yellow (71), and green (56). (B) Three-dimensional MDS plot where each colored point denotes a gene belonging to the module of the corresponding color. (C) Barplot showing the distribution of average gene significance (GS), that is, “module significance” and error bars across turquoise, blue, brown, yellow, and green modules. (D) Scatterplot showing significantly ($p < 0.05$) highest correlation of GS for age with module membership (MM) in the green module. DEG, differentially expressed gene; MDS, multidimensional scaling; MSI, mild and severe infectious; TOM, topological overlap matrix

3.5 | High-throughput virtual screening using LibDock

To identify novel compounds against ARHGEF1 through Dbl- and pleckstrin-homology (DH/PH) binding pocket, a docking-based virtual screening was performed using the LibDock module of DS 2019. Total 20 900

purchasable compounds comprising antiviral, anticancer, and natural products were taken from Life Chemicals (<https://lifechemicals.com/>) database and 35 COVID-19 related inhibitors from Selleck chem library which were currently tested for COVID therapy (<https://www.selleckchem.com/covid-19-related-products.html>) were also selected as a positive control. A total of 741 542

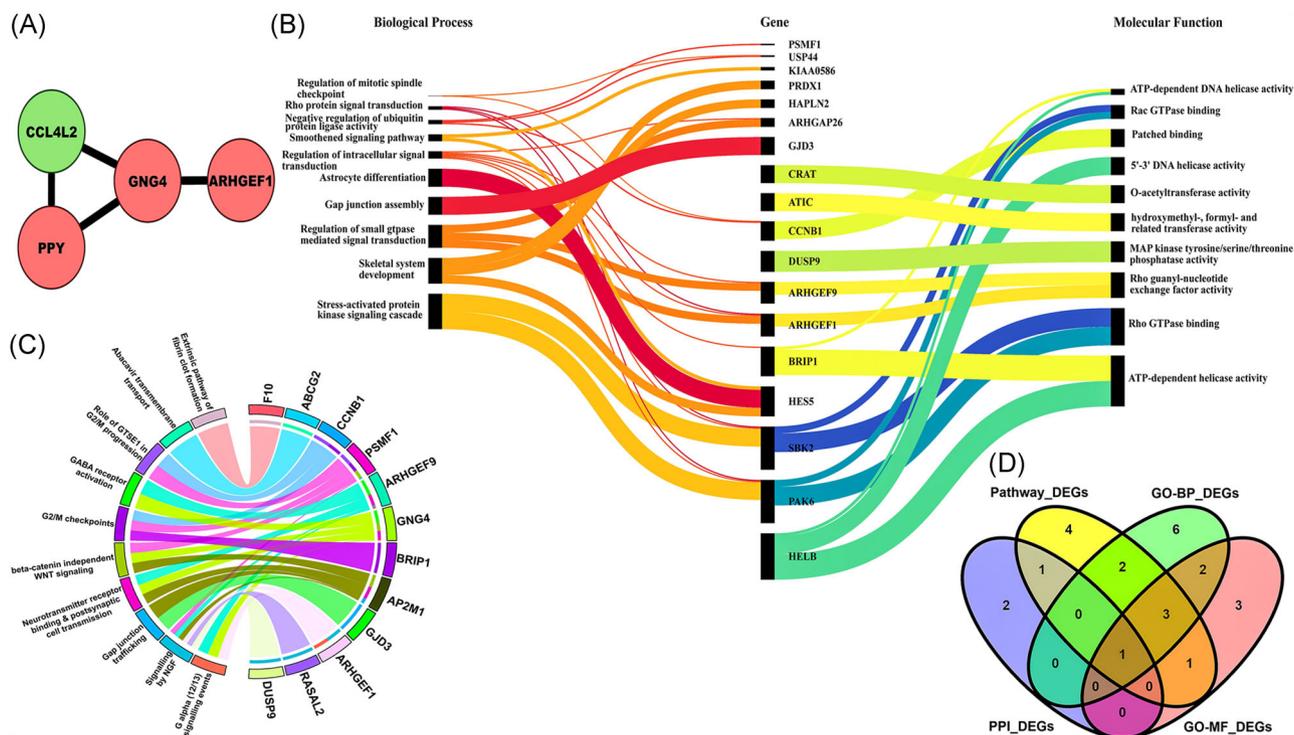


FIGURE 3 (A) Protein-protein interaction (PPI) network comprising four nodes and four edges corresponding to STRING interaction score >0.9 , that is, “highest confidence score.” The red- and green-colored nodes represent up- and downregulated proteins. (B) Sankey plot showing the association of top 10 significant GO-BP and GO-MF terms with corresponding 18 genes. The width and color intensity of interaction edges varied with respect to the *p*-values. (C) Chord plot showing the connection of 12 MSI-DEGs (on right semicircle) with 10 significant pathways (on left semicircle) via colored edges. The width of edges varied with respect to *p*-values. The edges initiated from unique colored strips present on the right semicircle (indicating genes) and converged to unique colored strips present on the left semicircle (indicating pathways). (D) Venn plot showing the overlapping hub MSI-DEG (1) between PPI, GO-BP, GO-MF, and pathway genesets. BP, biological process; DEG, differentially expressed gene; GO, gene ontology; MF, molecular function; MSI, mild and severe infectious

conformations of ligand were generated which aligned to the polar or apolar region of the protein known as hot-spots and 1 491 483 poses were docked successfully into the predicted binding site of the ARHGEF1. All the docked poses were shortlisted based on the highest LibDock score.

3.6 | Ligand preparation and ADMET prediction

To become a successful drug, ADMET, TOPKAT, and physiochemical properties of hit-compounds were predicted by using different modules of DS 2019. The top 300 compounds were selected for toxicity prediction followed by ADMET properties. Next, the oral bioavailability of selected hit compounds was calculated by Lipinski's rule of 5 and Veber's rule. Total 197 compounds following the TOPKAT, ADMET, Lipinski, and Veber's rule were subjected to energy minimization via steepest and conjugate algorithm. As COVID-19 related compounds were already approved for other targets, there is no need to

perform ADMET prediction. Therefore, final hit compounds were having satisfying drug-likeness properties and were predicted to have good bioavailability.

3.7 | Molecular docking

A total of 197 filtered compounds, were docked into the predicted binding site ($X = 46.5266$, $Y = -41.5728$, $Z = 26.6118$, and radii = 12 \AA) of ARHGEF1 protein (PDB ID:3ODO). Out of 197 compounds, 125 ligands were able to dock to the defined receptor cavity. Additionally, only 5 out of 35 COVID-19 related compounds were able to interact with ARHGEF1 using a favorable cDock energy score. All the ligands were arranged according to cDock energy score whereas positive cDock energy scored compounds were removed from the list. The molecular interaction of the binding poses was analyzed with analyse ligand poses in DS 2019. From all the binding poses, a total of 2381 favorable interactions, 1088 hydrogen bonds, 1052 hydrophobic interactions, and only 19 interactions were unfavorable. The top five

residues with favorable interactions were: Lys567, Arg551, Met434, Ala430, and Leu552 (Figure S5).

3.8 | Flexible docking

Flexible docking or induced fit docking overcome the problem of false-negative results and was more precise and accurate but it took more time to dock one ligand into the receptor cavity than other methods. The top 10 ligands from CDOCKER docking results and 5 ligands from COVID-19 related compounds were redocked via flexible docking method into the above binding site using DS 2019. The obtained docking poses were sorted based on their LibDock score (first-level screening) followed by cDocker energy (second-level screening). The cDocker energy score after flexible docking (third-level screening) of each pose was improved tremendously and enumerated in Table 1 and interactive residues of ARHGEF1 with top three hits were shown in Table 2 whereas docking scores of top hits from COVID-19 related compounds were enumerated in Table S5 and top two ligand interactions were shown in Table S6. The receptor–ligand interactions were analyzed in DS 2019 and, for instance, the interactions of the top three ligands, F3222-4380, F3394-0943, and F6548-1087 are explained in Figure 4 whereas interactions of the top two ligands (COVID-19 related library) are explained in Figure S6. It can be seen in Figure 4A, F3222-4380 binds with the active site of ARHGEF1 through interactions: hydrogen bonds (His431, Arg550, and Arg551), pi-cation/alkyl/pi-alkyl (Ala430, Arg433, Leu552, Lys567, and Arg564), and van der Waals interactions (Ser394, Thr427, Met434, Val437, Gln563, and Thr560). F3394-0943 interacted via hydrogen bond (Thr427, Arg551, and

Lys567), pi-cation/alkyl/pi-alkyl (Ala430, His431, Arg433, Arg550, Arg564, and Lys567), and with van der Waals interactions (Met434, Val437, Leu552, and Gln563) (Figure 4B). Finally, the interaction between F6548-1087 and ARHGEF1 was also observed, the interactions were hydrogen bonds (His431, Arg564, and Lys567), pi-cation/alkyl/pi-alkyl (Ala430, Arg433, and Leu552), and van der Waals (Glu423, Thr427, Met434, Val437, Arg550, Arg551, and Thr560) (Figure 4C).

3.9 | MD simulation

In computer-aided drug design, MD simulation studies are used as a validation tool that predicts the binding stability of the complexes throughout the simulation run. Herein, the top three docked complexes with ARHGEF1 were examined via a 50 ns MD simulation run. Trajectories were extracted to analyze the dynamic properties of each receptor–ligand complex. Additionally, MD simulations of the top two ligands from COVID-19 related library (Mitoxantrone and Camostat) complexed with ARHGEF1 were also performed for validation purposes. Each ligand shares the common binding pocket and exhibited similar kinds of interactions throughout the MD run. Additionally, five properties were also extracted from 50 ns MD trajectory for each hit compound, which includes (a) energy, (b) RMSD, (c) RMSF, (d) radius of gyration, and (e) hydrogen bonds.

The average potential energy score was found higher in ARHGEF1-Lead3 as $-7.91041e5$ kcal/mole in comparison to control protein (ARHGEF1) which has an average potential energy of $-8.76562e5$ kcal/mole. The ARHGEF1-Lead1 and ARHGEF1-Lead2 had average potential energy of $-1.16112e6$ and $-1.16114e6$ which

TABLE 1 Molecular docking score of the top 10 selected hits

Compound	Name	First-level screening LibDock score	Second-level screening -CDOCKER energy	Third-level screening Flexible (-CDOCKER energy)	Class
Lead 1	F3222-4380	132.739	34.7124	50.0296	Antiviral
Lead 2	F3394-0943	135.581	30.5814	45.7904	Anticancer
Lead 3	F6548-1087	129.865	28.3122	44.6191	Anticancer
Lead 4	F3394-0911	135.581	30.5814	43.8092	Anticancer
Lead 5	F0372-0536	137.649	34.5305	43.4841	Antiviral
Lead 6	F0375-0251	138.437	26.6627	42.1999	Anticancer
Lead 7	F1342-0042	131.347	32.7999	41.738	Antiviral
Lead 8	F3407-3720	126.415	30.3855	41.2499	Anticancer
Lead 9	F0886-0075	136.183	27.0587	40.27	Antiviral
Lead 10	F0886-0050	130.1	32.6969	36.0879	Antiviral

TABLE 2 Nonbond interactions between RhoGef1 and top three compounds

Compound name	Residues in contact	Types of interaction	Distance (Å°)
Lead 1 (F3222-4380)	His431	Carbon hydrogen bond	2.33
	Arg550	Carbon hydrogen bond	2.76
	Arg551	Conventional hydrogen bond	3.79
	Ala430	Alkyl	4.39
	Arg433	Pi-alkyl	4.69
	Leu552	Alkyl	4.97
	Arg564	Pi-cation	4.58
	Lys567	Pi-alkyl	2.55
Lead 2 (F3394-0943)	Thr427	Carbon hydrogen bond	3.59
	Arg551	Conventional hydrogen bond	5.73
	Lys567	Conventional hydrogen bond	5.12
	Ala430	Pi-alkyl	3.79
	His431	Pi-alkyl	4.58
	Arg433	Pi-alkyl	4.38
	Arg550	Pi-alkyl	5.49
	Arg564	Pi-cation	4.61
Lys567	Pi-alkyl	6.19	
Lead 3 (F6548-1087)	His431	Carbon hydrogen bond	4.77
	Arg564	Conventional hydrogen bond	4.83
	Lys567	Conventional hydrogen bond	5.54
	Ala430	Pi-alkyl	5.02
	Arg433	Pi-alkyl	5.25
	Leu552	Pi-alkyl	4.97

represents the stability of the complexes in terms of energy (Figure 5A). The potential energy of ARHGEF1-mitoxantrone and ARHGEF1-camostat complex were also energetically stable (Figure S7A). To measure the structural and conformational stability of the protein, the RMSD of the backbone atoms (Figure 5B and Figure S7B) for each complex was computed with respect to the initial docking structure. In all the receptor–ligand complexes and control protein, backbone RMSD showed deviation less than 0.4 nm. ARHGEF1-Lead1, ARHGEF1-Lead3, and ARHGEF1-mitoxantrone showed similar kinds of trends like control protein, whereas ARHGEF1-camostat and ARHGEF1-Lead2 complex was the least stable complex in comparison to other complexes and control proteins. Furthermore, the flexibility

of the amino acid residues was calculated using the RMSF profile. From Figure 5C and Figure S7C, RMSF of the backbone atoms was examined where residues of all complexes were found in the acceptable range of RMSF values, except in the case of F3394-0943 against ARHGEF1 in which there is a minimal amount of fluctuation from Glu586 to Leu599, but not in the binding region of the protein. From Figure 5D and Figure S7D, it was observed that the average R_g of all the complexes was below 2 nm which indicates the compactness of the system. Moreover, the number of hydrogen bonds that had a crucial role in the stability in the protein–ligand complex was analyzed from simulation trajectories. All the complexes remained stable (Figure 5E and Figure S7E), throughout the 50 ns run signified rigidity in the

COVID-19 related library) were performed to revalidate the MD simulation and molecular docking exercises and are tabulated in Table 3. All the top five complexes exhibited negative binding energy in the MM-PBSA threshold, respectively. However, among three compounds, Lead 1 (F3222-4380) exhibits a better net binding energy score which may be comparable with the flexible CDOCKER energy score (-50.0296 kcal/mole) (Table 2). The difference in net binding energy among all the complexes is due to the difference in polar solvation energy score which plays a minimum role in protein–ligand interactions.

4 | DISCUSSION

COVID-19 pandemic has radically affected the human population. Now, more than a year into the COVID-19 has passed but still, no specific therapeutic option is available for its treatment. This is one of the main reasons that new cases and deaths are still reported throughout the world. Therefore, identification of new therapeutic targets and development of therapeutic agents in COVID-19 are vigorously being pursued. On the basis of the life cycle of SARS-CoV-2 and the interaction of other molecular factors in COVID-19, we can propose several targets for drug development as well as propose therapeutic agents (existing drug

molecules and new ones) for the treatment of COVID-19. This can be achieved through computational approaches that significantly speed up the identification of therapeutic targets and the development of therapeutic molecules in drug discovery. Many new therapeutic targets and therapeutic molecules have been proposed in COVID-19.^{41–43(p. 2),44,45} But only a few have shown inhibition potency in bioassays, warranting new drug targets and drugs for the treatment of COVID-19. Some of the repurposed drug molecules such as remdesivir have been registered for COVID-19 therapy although based on the overall response of this drug globally till now states that it has several side effects. In addition to this, several vaccines have been approved for clinical use in the prevention of COVID-19.

To serve the same purpose, we applied the WGCNA to identify biologically relevant therapeutic targets in PBMC COVID-19 patient samples which were categorized into two groups, that is, Group 1: 5 mild + 5 controls and Group 2: 5 severe + 5 controls. WGCNA is a well-known approach for gene co-expression network analysis.^{46,47} We specifically aimed to identify the common therapeutic target (i.e., MSI-DEGs) in both groups. The DEGs overlapping between two groups was considered as MSI-DEGs. We found 569 MSI-DEGs which along with different age groups of patients were subjected to WGCN construction and hub module identification. In the WGCNA, we identified five unique gene modules. The

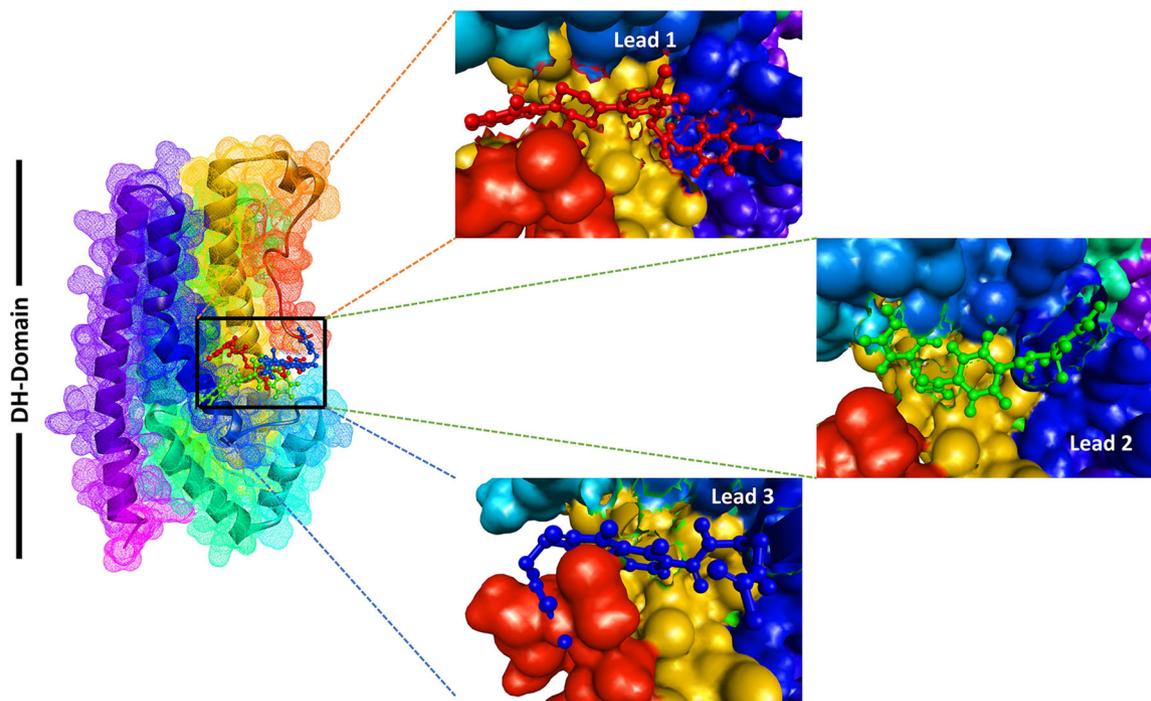


FIGURE 6 Binding modes of the screened ligand into the active site of protein after 50 ns of MD run. DH, Dbl homology; MD, molecular dynamics

TABLE 3 MM-PBSA energetic profile for the top three complexes

Compound Name	ΔE_{VDW} (kcal/mol)	ΔE_{ELE} (kcal/mol)	ΔG_{SOL-PB} (kcal/mol)	ΔG_{SOL-NP} (kcal/mol)	$\Delta G_{bind-PBSA}$ (kcal/mol)
Lead 1 (F3222-4380)	-50.35	-7.20	20.76	-4.13	-40.92
Lead 2 (F3394-0943)	-48.89	-22.75	42.69	-4.47	-33.42
Lead 3 (F6548-1087)	-45.51	-31.60	54.63	-4.02	-26.50

Abbreviations: MM, module membership; PBSA, Poisson-Boltzmann surface area.

green module (containing 56 MSI-DEGs) which was significant and highly correlated (Tables S1–S3) was selected as our hub module. Among these MSI-DEGs three upregulated (PPY, GNG4, and ARHGEF1) and one downregulated (CCL4L2) participated in the PPI network. While analyzing these MSI-DEGs for the most significant pathway, GO-BP, and GO-MF terms, we found that 12 MSI-DEGs participated in the top 10 significant pathways, 14 MSI-DEGs participated in the top 10 GO-BP, and 10 MSI-DEGs participated in top 10 GO-MF terms. The MSI-DEG(s) overlapping between constructed highest confidence PPI network, top 10 significant pathways, and GO terms were considered as SARS-CoV-2 specific hub MSI-DEG(s), respectively. We revealed ARHGEF1 as the overlapping hub MSI-DEG between PPI, pathway, and GO terms (i.e., GO-BP and GO-MF). ARHGEF1 gene encodes for the RhoGEF1 protein.^{48,49} RhoGEF1 is a member of a group of four RhoGEF proteins known to be activated by G protein-coupled receptors (GPCRs) coupled to the G12 and G13 heterotrimeric G proteins.^{49,50} We report that the most significant pathways associated with ARHGEF1 are G alpha (12/13) signaling events and signaling by nerve growth factor. The GO-MF terms showed that ARHGEF1 is associated with Rho guanyl-nucleotide exchange factor activity (GO: 0005089), while GO-BP terms revealed its association with Rho protein signal transduction (GO:0007266), regulation of intracellular signal transduction (GO:1902531), and regulation of small GTPase mediated signal transduction (GO:0051056).

The involvement and overlapping of ARHGEF1 between PPI, significant pathways, and GO terms reflect its translational product as a suitable target for the therapeutics and therapeutic management of COVID-19. Studies have reported that ARHGEF1, an intracellular signaling molecule, is involved in restraining signaling events from GPCRs to RhoA. It is crucially implicated in GPCR signaling events by possessing two regulatory domains – regulator of G-protein signaling (RGS) domain that combines with $G\alpha_{13}$ -containing heterotrimeric G-proteins and mitigates GPCR signaling; and tandem DH/PH domains that participate in RhoA activation by encoding Rho guanine nucleotide exchange factor (GEF).

Activation of RhoA, in turn, regulates several cellular processes and hence ARHGEF1 also regulates cell migration.^{50(p. 13),51(p. 13)} Several reports have demonstrated its role in cell migration and motility.^{52,53} Moreover, it also plays a critical role in pulmonary immunity. It was demonstrated by a study where ARHGEF1 existed as a tetramer in a B cell line, however, interrupting intramolecular association, lead to constitutive activation of RGS and RhoGEF. On the other hand, deficiency of ARHGEF1 increased adhesion of B lymphocytes, while its constitutive activation caused increased adhesion of B lymphocytes. In addition, it has also been found that ARHGEF1 is a key player in the production of Th2 cytokines by T cells, in the case of airway hyperresponsiveness and inflammation. Furthermore, ARHGEF1 deficient mice exhibited increased leukocytes and pulmonary components such as increased airspace and matrix metalloproteinase (MMP) activity with decreased lung elastase; typically found in patients with chronic obstructive pulmonary disease.^{54–58} It is also found to be involved in a novel signaling pathway, comprising of thromboxane A2 that is endowed in pulmonary inflammation. High or low levels of ARHGEF1 determines thromboxane receptor-dependent production of MMP9 and macrophage association of fibronectin.⁵⁹

Additionally, there is no ligand complexed crystal structure available for ARHGEF1 at present. Likewise, there is no clear-cut information available regarding the ligand-binding sites present on ARHGEF1. Therefore, to target the ARHGEF1 for the therapeutics of COVID-19, we analyzed its structure and predicted three binding pockets. Pocket 1 was defined as a suitable binding site based on rank score and composition of amino acids namely Thr427, Ala430, His431, Met434, Arg551, Leu552, Asp556, Met557, Thr560, Gln563, Arg564, and Lys567. The majority of these amino acid residues provide a plinth for the binding of potent inhibitors of many therapeutic targets in CoV diseases.^{41(p. 2),42,60}

Next, we applied molecular docking and MD method against ARHGEF1 gene¹⁵ to discover potential hits for SARS-CoV-2 from the antiviral library, anticancer library, and natural product-based library. We found that the three top ligands – Lead F3222-4380, F3394-0943, and

F6548-1087 possessed significant interactions with the active site of ARHGEF1, including hydrogen bonds, p-cation, and Van der Waals Interactions, suggesting their potential in modulating functional roles of ARHGEF1. To validate our findings and endorse the molecular docking study, we have selected the top 3 hit compounds for MD simulation. The RMSD profile from F3222-4380, F3394-0943, and F6548-1087 did not exceed 0.4 nm as their averages were 0.2134, 0.2967, and 0.2336 nm, respectively, which represents their structural stability. The F3222-4380 and F6548-1087 complex were energetically more stable than the control protein. On the other hand, RMSF analysis exhibited that F3394-0943 fluctuated more in the region of Glu586 and Leu599 in comparison to F3222-4380 and F6548-1087, overall, all the hit compounds have RMSF profile have less than 1 nm. These potential candidate hits could act as probable compounds that may target the important gene ARHGEF1, implicated in COVID-19, and can be further selected for in vitro and in vivo studies in future work to speed up the drug discovery against SARS-CoV-2. Additionally, we performed molecular docking and dynamics method against COVID-19 related library which serves as a positive control in our study. However, our novel hit compounds had better cDock energy and MM-PBSA profile in comparison to mitoxantrone and camostat (COVID-19-related compounds). Overall, our study has identified some important components involved in pulmonary immunity, and understanding the MMs behind the ARHGEF1 mediated regulation of immunity may reveal potential therapeutic targets to limit chronic inflammation. Besides, chemical compounds targeting ARHGEF1 could pave the way towards the development of medicine for the calamitous disease.

In conclusion, as we navigate through the findings, our study has identified some important components involved in pulmonary immunity. This study showed a significant association of the ARHGEF1 gene with COVID-19, therefore highlighting it as a probable drug target for conventional treatments of this pandemic disease. The understanding of the MMs behind the ARHGEF1 mediated regulation of immunity may limit chronic inflammation in COVID-19 patients. This study also found some actionable drug molecules with a rationale against the ARHGEF1 gene. These molecules may be effective alone or in combination with other drug molecules in COVID-19. The in vitro and in vivo validation are warranted to determine their therapeutic efficacy against the devastating virus.

ACKNOWLEDGMENTS

The authors would like to thank the Department of Biotechnology, Government of India (Grant Number: BT/PR40153/BTIS/137/8/2021) for providing

Bioinformatics Infrastructure Facility (BIF) at Dr. B. R. Ambedkar Center for Biomedical Research (ACBR). The authors would also like to acknowledge Jamia Millia Islamia, All India Institute of Medical Sciences, and Jawaharlal Nehru University for providing infrastructure, journal access, high-computing and internet facilities. Prakash Jha would like to thank the Department of Science and Technology, Government of India for awarding him DST-INSPIRE fellowship (Grant Number: DST/INSPIRE/03/2016/000026). Prithvi Singh and Shweta Arora would like to thank the Indian Council of Medical Research for awarding them Senior Research Fellowship (Grant Numbers: BMI/11(89)/2020 to Prithvi Singh and 5/3/8/13/ITR-F/2018-ITR to Shweta Arora).

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Prakash Jha, Prithvi Singh, Ravins Dohare, and Madhu Chopra: Conceptualization. **Prakash Jha, Prithvi Singh, Arnab Nayek, and Kalaiarasan Ponnusamy:** Methodology. **Prakash Jha, Prithvi Singh, and Arnab Nayek:** Software. **Prakash Jha and Prithvi Singh:** Formal analysis. **Prakash Jha, Prithvi Singh, Arnab Nayek, and Kalaiarasan Ponnusamy:** Data curation. **Prakash Jha, Prithvi Singh, Armiya Sultan, and Shweta Arora:** Writing – original draft. **Prakash Jha, Prithvi Singh, Armiya Sultan, Shweta Arora, Arnab Nayek, Kalaiarasan Ponnusamy, Ravins Dohare, Mansoor Ali Syed, and Madhu Chopra:** Writing – review and editing. **Ravins Dohare and Madhu Chopra:** Supervision. **Ravins Dohare and Madhu Chopra:** Project administration.

DATA AVAILABILITY STATEMENT

The dataset used in our study was downloaded from National Center for Biotechnology Information–Gene Expression Omnibus under accession number GSE164805 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164805>, accessed on May 1, 2021).

ORCID

Armiya Sultan  <http://orcid.org/0000-0001-5373-8824>

Ravins Dohare  <http://orcid.org/0000-0002-0602-3719>

REFERENCES

1. Lai C-C, Shih T-P, Ko W-C, Tang H-J, Hsueh P-R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *Int J Antimicrob Agents*. 2020;55:105924. doi:10.1016/j.ijantimicag.2020.105924

2. Sharma A, Tiwari S, Deb MK, Marty JL. Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): a global pandemic and treatment strategies. *Int J Antimicrob Agents*. 2020;56:106054. doi:10.1016/j.ijantimicag.2020.106054
3. Severe Acute Respiratory Syndrome (SARS). (n.d.). [WWW Document]. Accessed November 7, 2021. <https://www.who.int/westernpacific/health-topics/severe-acute-respiratory-syndrome>
4. COVID Live Update: 187,393,771 cases and 4,045,487 deaths from the coronavirus—Worldometer. (n.d.). [WWW Document]. Accessed November 7, 2021. <https://www.worldometers.info/coronavirus/>
5. Li H, Liu S-M, Yu X-H, Tang S-L, Tang C-K. Coronavirus disease 2019 (COVID-19): current status and future perspectives. *Int J Antimicrob Agents*. 2020;55:105951. doi:10.1016/j.ijantimicag.2020.105951
6. Zheng J. SARS-CoV-2: an emerging coronavirus that causes a global threat. *Int J Biol Sci*. 2020;16:1678-1685. doi:10.7150/ijbs.45053
7. #IndiaFightsCorona COVID-19 in India, Vaccination, Dashboard, Corona Virus Tracker | mygov.in. (n.d.). [WWW Document]. Accessed November 7, 2021. <https://www.mygov.in/covid-19>
8. Alluwaimi AM, Alshubaith IH, Al-Ali AM, Abohelaika S. The coronaviruses of animals and birds: their zoonosis, vaccines, and models for SARS-CoV and SARS-CoV2. *Front Vet Sci*. 2020;7:582287. doi:10.3389/fvets.2020.582287
9. Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. *Virol J*. 2019;16:69. doi:10.1186/s12985-019-1182-0
10. Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*. 2021;19:141-154. doi:10.1038/s41579-020-00459-7
11. Brahim Belhaouari D, Fontanini A, Baudoin J-P, et al. The strengths of scanning electron microscopy in deciphering SARS-CoV-2 infectious cycle. *Front Microbiol*. 2020;11:2014. doi:10.3389/fmicb.2020.02014
12. Romano M, Ruggiero A, Squeglia F, Maga G, Berisio R. A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells*. 2020;9:1267. doi:10.3390/cells9051267
13. Saleki K, Yaribash S, Banazadeh M, et al. Interferon therapy in patients with SARS, MERS, and COVID-19: a systematic review and meta-analysis of clinical studies. *Eur J Pharmacol*. 2021;906:174248. doi:10.1016/j.ejphar.2021.174248
14. Zhang X-Y, Huang H-J, Zhuang D-L, et al. Biological, clinical and epidemiological features of COVID-19, SARS and MERS and AutoDock simulation of ACE2. *Infect Dis Poverty*. 2020;9:99. doi:10.1186/s40249-020-00691-6
15. Cook DR, Rossman KL, Der CJ. Rho guanine nucleotide exchange factors: regulators of Rho GTPase activity in development and disease. *Oncogene*. 2014;33:4021-4035. doi:10.1038/onc.2013.362
16. Cava C, Bertoli G, Castiglioni I. Potential drugs against COVID-19 revealed by gene expression profile, molecular docking and molecular dynamic simulation. *Future Virol*. 2021;16:527-542. doi:10.2217/fvl-2020-0392
17. Cava C, Bertoli G, Castiglioni I. A protein interaction map identifies existing drugs targeting SARS-CoV-2. *BMC Pharmacol Toxicol*. 2020;21:65. doi:10.1186/s40360-020-00444-z
18. Clough E, Barrett T. The Gene Expression Omnibus database. *Methods Mol Biol*. 2016;1418:93-110. doi:10.1007/978-1-4939-3578-9_5
19. Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics*. 2007;23:1846-1847. doi:10.1093/bioinformatics/btm254
20. Hackstadt AJ, Hess AM. Filtering for increased power for microarray data analysis. *BMC Bioinformatics*. 2009;10:11. doi:10.1186/1471-2105-10-11
21. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47. doi:10.1093/nar/gkv007
22. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559. doi:10.1186/1471-2105-9-559
23. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47:D607-D613. doi:10.1093/nar/gky1131
24. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498-2504. doi:10.1101/gr.1239303
25. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44:W90-W97. doi:10.1093/nar/gkw377
26. Jendele L, Krivak R, Skoda P, Novotny M, Hoksza D. PrankWeb: a web server for ligand binding site prediction and visualization. *Nucleic Acids Res*. 2019;47:W345-W349. doi:10.1093/nar/gkz424
27. Rao SN, Head MS, Kulkarni A, LaLonde JM. Validation studies of the site-directed docking program LibDock. *J Chem Inf Model*. 2007;47:2159-2171. doi:10.1021/ci6004299
28. Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL. Quantifying the chemical beauty of drugs. *Nat Chem*. 2012;4:90-98. doi:10.1038/nchem.1243
29. Zhu Y, Han Y, Ma Y, Yang P. ADME/toxicity prediction and antitumor activity of novel nitrogenous heterocyclic compounds designed by computer targeting of alkylglycerone phosphate synthase. *Oncol Lett*. 2018;16:1431-1438. doi:10.3892/ol.2018.8873
30. Yuan G, Li T, Hu W. A conjugate gradient algorithm and its application in large-scale optimization problems and image restoration. *J Inequalities Appl*. 2019;2019:247. doi:10.1186/s13660-019-2192-6
31. Wu G, Robertson DH, Brooks CL, Vieth M. Detailed analysis of grid-based molecular docking: a case study of CDOCKER-A CHARMM-based MD docking algorithm. *J Comput Chem*. 2003;24:1549-1562. doi:10.1002/jcc.10306
32. Chen Z, Guo L, Sprang SR, Sternweis PC. Modulation of a GEF switch: autoinhibition of the intrinsic guanine nucleotide

- exchange activity of p115-RhoGEF. *Protein Sci.* 2011;20:107-117. doi:10.1002/pro.542
33. Li X, Li Y, Cheng T, Liu Z, Wang R. Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. *J Comput Chem.* 2010;31:2109-2125.
34. Spassov VZ, Yan L, Flook PK. The dominant role of side-chain backbone interactions in structural realization of amino acid code. ChiRotor: a side-chain prediction algorithm based on side-chain backbone interactions. *Protein Sci.* 2007;16:494-506. doi:10.1110/ps.062447107
35. Abdel-Hamid MK, McCluskey A. In silico docking, molecular dynamics and binding energy insights into the bolinaquinone-clathrin terminal domain binding site. *Mol Basel Switz.* 2014;19:6609-6622. doi:10.3390/molecules19056609
36. Pronk S, Páll S, Schulz R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics.* 2013;29:845-854. doi:10.1093/bioinformatics/btt055
37. Schüttelkopf AW, van Aalten DMF. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr.* 2004;60:1355-1363. doi:10.1107/S0907444904011679
38. Childers MC, Daggett V. Validating molecular dynamics simulations against experimental observables in light of underlying conformational ensembles. *J Phys Chem B.* 2018;122:6673-6689. doi:10.1021/acs.jpcc.8b02144
39. Abraham MJ, Gready JE. Optimization of parameters for molecular dynamics simulation using smooth particle-mesh Ewald in GROMACS 4.5. *J Comput Chem.* 2011;32:2031-2040. doi:10.1002/jcc.21773
40. Kollman PA, Massova I, Reyes C, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res.* 2000;33:889-897. doi:10.1021/ar000033j
41. Jin Z, Du X, Xu Y, et al. Structure of M^{Pro} from SARS-CoV-2 and discovery of its inhibitors. *Nature.* 2020;582:289-293. doi:10.1038/s41586-020-2223-y
42. Jin Z, Zhao Y, Sun Y, et al. Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. *Nat Struct Mol Biol.* 2020;27:529-532. doi:10.1038/s41594-020-0440-6
43. McKee DL, Sternberg A, Stange U, Laufer S, Naujokat C. Candidate drugs against SARS-CoV-2 and COVID-19. *Pharmacol Res.* 2020;157:104859. doi:10.1016/j.phrs.2020.104859
44. Sultan A, Ali R, Sultan T, Ali S, Khan NJ, Parganiha A. Circadian clock modulating small molecules repurposing as inhibitors of SARS-CoV-2 M^{Pro} for pharmacological interventions in COVID-19 pandemic. *Chronobiol Int.* 2021;38:971-985. doi:10.1080/07420528.2021.1903027
45. Wang M, Cao R, Zhang L, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res.* 2020;30:269-271. doi:10.1038/s41422-020-0282-0
46. Xiao J, Wang R, Cai X, Ye Z. Coupling of co-expression network analysis and machine learning validation unearthed potential key genes involved in rheumatoid arthritis. *Front Genet.* 2021;12:604714. doi:10.3389/fgene.2021.604714
47. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol.* 2005;4:17. doi:10.2202/1544-6115.1128
48. Aasheim H-C, Pedoutour F, Smeland EB. Characterization, expression and chromosomal localization of a human gene homologous to the mouse Lsc oncogene, with strongest expression in hematopoietic tissues. *Oncogene.* 1997;14:1747-1752. doi:10.1038/sj.onc.1200994
49. Hart MJ, Sharma S, elMasry N, et al. Identification of a novel guanine nucleotide exchange factor for the Rho GTPase. *J Biol Chem.* 1996;271:25452-25458. doi:10.1074/jbc.271.41.25452
50. Kozasa T, Jiang X, Hart MJ, et al. p115 RhoGEF, a GTPase activating protein for Galpha12 and Galpha13. *Science.* 1998;280:2109-2111. doi:10.1126/science.280.5372.2109
51. Hart MJ, Jiang X, Kozasa T, et al. Direct stimulation of the guanine nucleotide exchange activity of p115 RhoGEF by Galpha13. *Science.* 1998;280:2112-2114. doi:10.1126/science.280.5372.2112
52. Girkontaite I, Missy K, Sakk V, et al. Lsc is required for marginal zone B cells, regulation of lymphocyte motility and immune responses. *Nat Immunol.* 2001;2:855-862. doi:10.1038/ni0901-855
53. Rubtsov A, Strauch P, Digiacoio A, Hu J, Pelanda R, Torres RM. Lsc regulates marginal-zone B cell migration and adhesion and is required for the IgM T-dependent antibody response. *Immunity.* 2005;23:527-538. doi:10.1016/j.immuni.2005.09.018
54. Francis SA, Shen X, Young JB, Kaul P, Lerner DJ. Rho GEF Lsc is required for normal polarization, migration, and adhesion of formyl-peptide-stimulated neutrophils. *Blood.* 2006;107:1627-1635. doi:10.1182/blood-2005-03-1164
55. Hartney JM, Brown J, Chu HW, Chang LY, Pelanda R, Torres RM. Arhgef1 regulates alpha5beta1 integrin-mediated matrix metalloproteinase expression and is required for homeostatic lung immunity. *Am J Pathol.* 2010;176:1157-1168. doi:10.2353/ajpath.2010.090200
56. Hu J, Strauch P, Rubtsov A, Donovan EE, Pelanda R, Torres RM. Lsc activity is controlled by oligomerization and regulates integrin adhesion. *Mol Immunol.* 2008;45:1825-1836. doi:10.1016/j.molimm.2007.11.002
57. Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med.* 2007;176:532-555. doi:10.1164/rccm.200703-456SO
58. Brown JP, Taube C, Miyahara N, et al. Arhgef1 is required by T cells for the development of airway hyperreactivity and inflammation. *Am J Respir Crit Care Med.* 2007;176:10-19. doi:10.1164/rccm.200702-270OC
59. Guan Y, Torres RM, Hartney JM. The influence of Arhgef1 on pulmonary leukocyte function. *Immunol Res.* 2013;55:162-166. doi:10.1007/s12026-012-8360-0

60. Zhang L, Lin D, Sun X, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science*. 2020;368:409-412. doi:10.1126/science.abb3405

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Jha P, Singh P, Arora S, et al. Integrative multiomics and in silico analysis revealed the role of ARHGEF1 and its screened antagonist in mild and severe COVID-19 patients. *J Cell Biochem*. 2022;123:673-690. doi:10.1002/jcb.30213