



A Path-Based Analysis of Infected Cell Line and COVID-19 Patient Transcriptome Reveals Novel Potential Targets and Drugs Against SARS-CoV-2

Piyush Agrawal^{1†}, Narmada Sambaturu^{2†}, Gulden Olgun¹ and Sridhar Hannenhalli^{1*}

¹ Cancer Data Science Laboratory, National Cancer Institute, National Institutes of Health, Bethesda, MD, United States,

² IISc Mathematics Initiative, Indian Institute of Science, Bangalore, India

OPEN ACCESS

Edited by:

Lucia Lopalco,
San Raffaele Hospital (IRCCS), Italy

Reviewed by:

Mansi Srivastava,
Purdue University Indianapolis,
United States
Eleni Karakike,
National and Kapodistrian University of
Athens, Greece

*Correspondence:

Sridhar Hannenhalli
sridhar.hannenhalli@nih.gov

[†]These authors share first authorship

Specialty section:

This article was submitted to
Viral Immunology,
a section of the journal
Frontiers in Immunology

Received: 12 April 2022

Accepted: 07 June 2022

Published: 01 July 2022

Citation:

Agrawal P, Sambaturu N, Olgun G and
Hannenhalli S (2022) A Path-Based
Analysis of Infected Cell Line and
COVID-19 Patient Transcriptome
Reveals Novel Potential Targets and
Drugs Against SARS-CoV-2.
Front. Immunol. 13:918817.
doi: 10.3389/fimmu.2022.918817

Most transcriptomic studies of SARS-CoV-2 infection have focused on differentially expressed genes, which do not necessarily reveal the genes mediating the transcriptomic changes. In contrast, exploiting curated biological network, our PathExt tool identifies central genes from the differentially active paths mediating global transcriptomic response. Here we apply PathExt to multiple cell line infection models of SARS-CoV-2 and other viruses, as well as to COVID-19 patient-derived PBMCs. The central genes mediating SARS-CoV-2 response in cell lines were uniquely enriched for ATP metabolic process, G1/S transition, leukocyte activation and migration. In contrast, PBMC response reveals dysregulated cell-cycle processes. In PBMC, the most frequently central genes are associated with COVID-19 severity. Importantly, relative to differential genes, PathExt-identified genes show greater concordance with several benchmark anti-COVID-19 target gene sets. We propose six novel anti-SARS-CoV-2 targets ADCY2, ADSL, OCRL, TIAM1, PBK, and BUB1, and potential drugs targeting these genes, such as Bemcentinib, Phthalocyanine, and Conivaptan.

Keywords: SARS-CoV-2, transcriptome, cell lines, PBMCs (Peripheral Blood Mononuclear Cells), DEGs (Differentially Expressed Genes), network analysis

INTRODUCTION

COVID-19, a serious respiratory disease caused by the zoonotic virus Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2, or SC2 for short), has emerged as a global pandemic leading to ~315 million infections and 5 million deaths (data till date January 10th, 2022, as per WHO dashboard). Despite a huge body of research investigating the SC2 biology (1), host-virus mechanisms (2), potential drug targets and their inhibitors (3), we are far from a complete understanding of mechanisms underlying the varied COVID-19 symptoms, and the search for an effective therapy continues. As there are limited treatment options available for SC2, several drugs which are prescribed to treat infections by other viruses such as Middle East respiratory syndrome (MERS), severe acute respiratory syndrome (SARS), human immunodeficiency virus (HIV), etc.

have been tried on COVID-19 patients (4). Remdesivir, a viral RNA dependent RNA polymerase inhibitor used in the treatment of Ebola virus, was the first FDA-approved drug for the treatment of COVID-19 (5); however, it has not been broadly effective. More recently approved Paxlovid was found to be effective against multiple SC2 variants including Omicron (6). While several effective vaccines are now in general use, there continues to be an urgent need for more effective therapeutic options for infected patients, especially considering the highly varied, and sometimes long-term effects as well as side effects of the current therapies.

Viruses hijack the host system for their own survival and proliferation, especially by exploiting and manipulating host transcriptional machinery and gene regulation (7). A better understanding of the host transcriptomic response to the viral infection is thus widely recognized as critical in designing better therapeutics strategies (4). Most previous studies investigating infection-induced transcriptomic changes in the host tissues and immune cells focus on genes that are differentially expressed (DEGs) upon infection and perform a series of downstream analyses to decipher the underlying mechanisms (8). One issue with the DEG-centric approach is that certain genes are known to be differentially expressed in a wide variety of contexts and represent generic transcriptional response and are not specific to SC2 infection (9). Moreover, it is now well recognized that differential expression ought to be interpreted in the context of genetic networks and pathways, and indeed some of the works have investigated SC2 transcriptomic response data from a network perspective (10). These approaches nevertheless rely on gene-level differential expression as the lynchpin for the downstream network-assisted analyses. An alternative approach - PathExt - that we have recently shown to be superior to DEG-centric approaches (11), instead integrates transcriptomic data with curated gene networks and instead of identifying differentially expressed genes, identifies differentially active paths in the integrated network, and then identifies the central genes mediating the differential activities of the most perturbed paths. This alternative approach is based on the recognition that (i) gene expression is noisy and DEGs can therefore lead to false positives, and (ii) key regulatory genes that mediate global transcriptomic changes and thus present a potent target may themselves not be differentially regulated and will thus be missed by DEG-centric approaches.

Here, using PathExt, we comprehensively analyze and compare transcriptomic response to SC2 and other respiratory viruses (SARS-CoV-1, MERS, Influenza, RSV, and HPIV3) in multiple cell lines (A549, A549-ACE2, Calu3, Vero, MRC5 and NHBE), as well as in COVID-19 patient-derived peripheral blood mononuclear cells (PBMCs). While PathExt identifies largely distinct sets of central genes across cell lines and viruses, these genes nevertheless converge on common processes such as cytokine signaling, cell cycle, metabolism, etc.; however, we observe a much greater similarity in response across cell lines for the same virus than across viruses in the same cell line. We assess the complementarity and unique advantages of using PathExt compared to the conventional DEG-based approach and find that PathExt genes capture experimentally

identified anti-SC2 targets more accurately than DEGs, while also providing an overall greater enrichment of key biological processes. In PBMC data, we find that PathExt-identified central genes are associated with patient severity. Finally, we propose novel anti-SC2 therapeutic targets and their potential inhibitors that are either FDA-approved or currently in clinical trials.

RESULTS

Overview of the Approach

PathExt is our recently published tool; here we provide a brief intuitive overview of the approach. The aim of PathExt is to identify differentially active paths, in a prior knowledge-based gene network, while comparing transcriptomic data in two conditions, and shortlist genes among those paths which might be critically mediating the observed phenotypic change. As noted above, in contrast to conventional DEG-centered approaches, such mediating genes may not be differentially expressed themselves. The PathExt workflow is illustrated in **Figure 1A**. PathExt starts by integrating a knowledge-based curated gene network with sample-specific omics data from the conditions of interest; we employed a curated network - HPPIN - which integrates physical, regulatory, and metabolic interactions between genes or proteins (12). In each sample separately, nodes and edges of the network are weighted such that interactions involving differentially expressed genes are preferentially traversed by a shortest path algorithm. This integration is done in two ways, to emphasize either differential activation or repression. Shortest paths whose weights are statistically significant (based on permutation) can then be interpreted as statistically significant differentially active (or repressed) paths. Such paths are then cast as a sub-network referred to as the TopNet, either activated or repressed, based on the weighting scheme. PathExt then identifies central genes in each TopNet based on ripple centrality (13), which captures genes that can reach a large part of the TopNet along highly active (or repressed) paths. TopNet and the central genes are identified independently for each transcriptomic sample.

Here, we apply PathExt to analyze how the impact of SC2 infection (as well as other respiratory viruses) can vary in different cell lines, as well as in patient derived PBMCs (**Figure 1B**). We first compare SC2 infection across cell lines, and with other viral infections in the same cell lines. In PBMC data, we identify associations between PathExt results and patient demographics. We then show a high concordance between the results from both cell line and PBMC data, and benchmarks such as genes previously shown to be affected by SC2 infection as well as experimentally screened drugs. Based on this, we use the output of PathExt to propose novel drug targets as well as drugs against them.

Key Processes Mediating SC2 Infection Across Cell Lines

We compared the transcriptional responses between five cell lines infected by SC2. These included three lung epithelium derived cell lines - A549, A549 with increased ACE2 expression

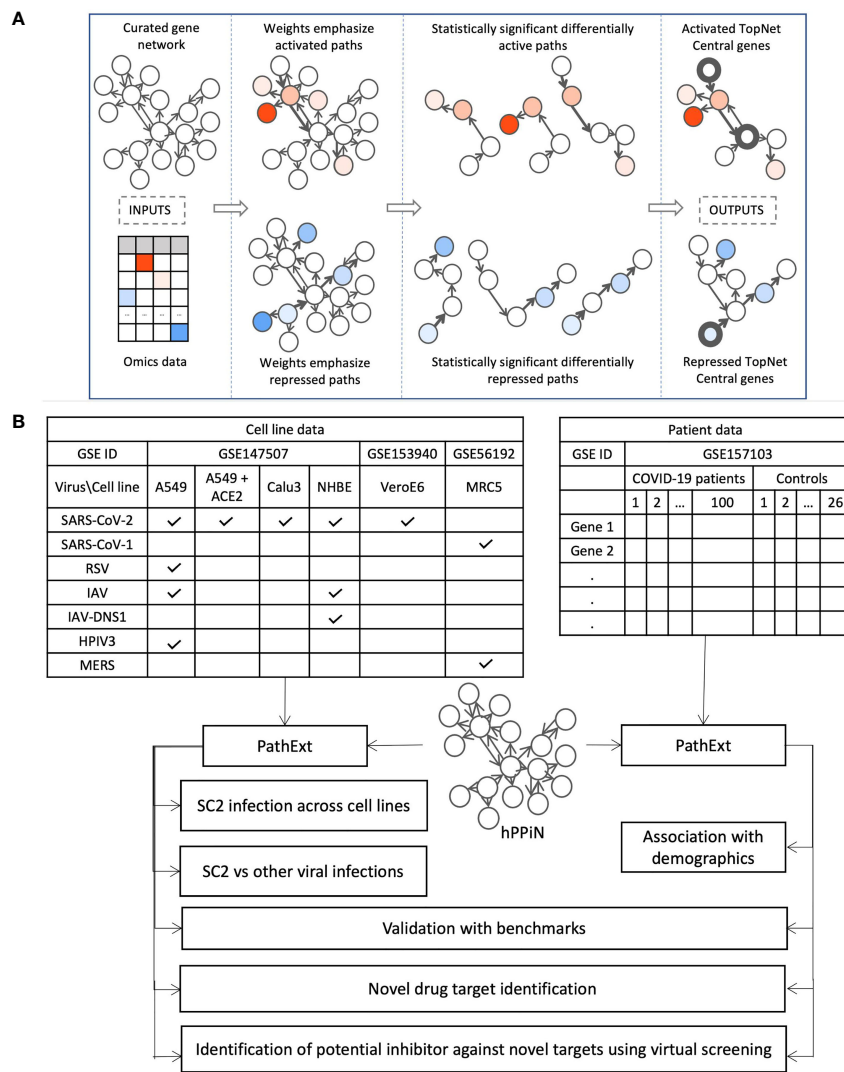


FIGURE 1 | Study Workflow. (A) Our PathExt tool accepts as input a curated gene network and gene expression data, to output two weighted sub-networks - an activated and a repressed TopNet comprising activated and repressed paths respectively, and a list of central genes in each TopNet. Activated genes are shown here in shades of red, and repressed genes are in shades of blue. PathExt integrates the inputs such that edges connecting genes with substantial change in expression are preferentially traversed by a shortest paths algorithm (Methods), illustrated here using wider arrows. Shortest paths which are statistically significant (permutation based) now represent differentially active (or repressed) paths and make up the TopNets in which PathExt identifies central genes based on ripple centrality. **(B)** We apply PathExt to analyze RNA-seq data from SARS-CoV-2 infection in both cell lines and patient PBMC data. PathExt outputs from the cell line data are used to compare cross-cell-line variation in SC2 infection response, and within-cell-line variation in response to other viruses. In patient PBMC data, we identify associations between PathExt results and demographics. We then validate all the results against benchmarks and use the TopNets and central genes to propose novel drug targets, as well as novel drugs for known targets.

(A549-ACE2), and Calu3, Vero cell line derived from African monkey kidney, and primary bronchial epithelium cell line – NHBE (14, 15). In each case, we applied PathExt to identify the top 100 central genes each in the activated and repressed TopNets (**Supplementary Table S1**); no repressed TopNet was detected in NHBE. We found that the top 100 genes were largely disjoint across cell lines (**Figures 2A, B**) with one exception, where the top 100 genes from the activated TopNets in Calu3 and Vero cell lines shared 79 common genes; indicating a high

cell-type specificity in response to SC2 infection. A list of the top 100 upregulated and downregulated DEGs is provided in the **Supplementary Table S2**.

We then performed functional enrichment on the top 100 genes identified from each dataset using PANTHER (16), and consolidated (Methods) the enriched biological process terms using REVIGO (17). In case of activated TopNets (**Figure 2C**), as expected, Calu3 and Vero cells show similar upregulated pathways such as I-kappaB kinase/NF-kappaB signaling,

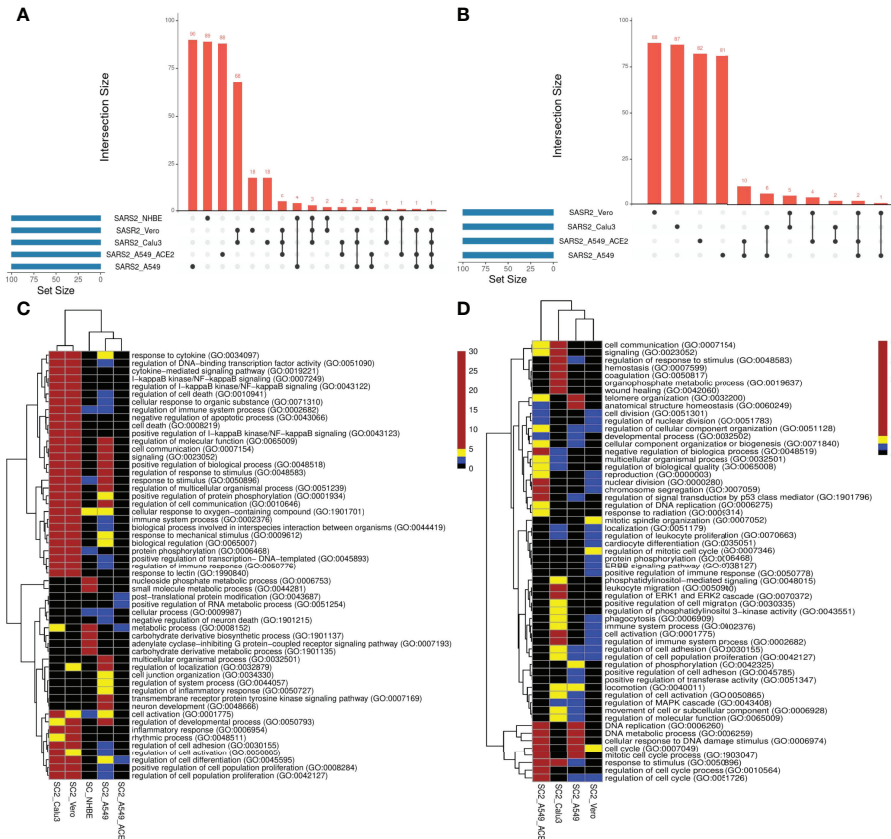


FIGURE 2 | SC2 infection across cell lines. We analyzed the transcriptional response across various cell lines post SC2 infection. We obtained the top100 central genes from each cell line for both TopNets (activated & repressed) and compared the gene commonality across cell lines as shown in upset plots for activated TopNets (A) and repressed TopNets (B). Next, we performed gene enrichment analysis and compared the top 10 parent GO terms enriched in various cell lines for activated TopNets (C) and repressed TopNets (D). Here we show only those GO terms that were significant in at least 2 cell lines. Complete list of enriched GO terms is provided in the supplementary tables. Heatmap is created by converting the FDR corrected p-value of each GO terms to $-\log_{10}$ scale. Significance of the terms is shown in various color ranges (0-1.3, 1.3-2.0, 2-5 and >5-30). All the non-significant processes are shown in black (value <1.3).

regulation of transcription factor (TF) activity, cytokine mediated signaling pathway, and response to stimulus. These processes are well supported by previous studies (18–20). Regulation of immune signaling pathways is another process prominently observed in COVID-19, especially production of various cytokines and chemokines (21, 22). Likewise, hyperactivation of NF-kappaB signaling pathway post SARS-CoV-2 infection has been observed (23) and NF-kappaB has been recognized as a potential pharmacological target to treat COVID-19 (24). Several host TF binding sites (TFBS) are present in SARS-CoV-2 genome which the virus exploits for its replication (25). Calu3, Vero, and A549 were enriched for inflammatory response, specifically, tyrosine kinase signaling, recapitulating previously established links between SARS-CoV-2 and tyrosine kinase signaling (26). In A549 cells with high ACE2 expression, the strongest enrichment was seen for the post-translational protein modification process. Although inflammatory and immune response pathways were not statistically enriched among the top 100 genes in A549-ACE2

cell line, key immune response genes were among the top 100 central genes, e.g., *STAT3*, *STAT5A*, *NFKB1*, *NFKBIA*, etc.

Post-translational modification (PTM) of viral proteins by the host machinery is well documented, such as glycosylation and palmitoylation of Spike protein, N or O-linked glycosylation of membrane protein, phosphorylation and ADP-ribosylation of nucleocapsid protein (27). Likewise, deimination of SARS-CoV-2 proteins is catalyzed by peptidylarginine deiminases (PADs) which leads to activation of neutrophils extracellular traps (NETs) in an ACE2-dependent manner (28). These findings suggest that a larger viral load, facilitated by high ACE2 expression, is likely to induce greater PTM. Gordon et al. (29) created a PPI network where 332 human proteins were interacting with 27 viral proteins. They showed that high viral load inside the host cells leads to increased human protein phosphorylation and host kinase modulation.

Other enriched processes such as metabolism and cell differentiation also have known roles in viral infections. These results suggest that the host response to viral infection may

depend on the infection and replication rates in various cell lines. For instance, the differential response in A549 may be because A549 cells are relatively less permissive to SARS-CoV-2 replication owing to lower concentration of *ACE2* receptor protein required for viral entry (30), while Calu3 and Vero cell lines exhibit a much higher expression of *ACE2*. This difference in infection rate leads to differences in the host immune responses as shown previously (15). Key genes in NHBE (bronchial epithelium cell line) were strongly associated with small molecule and nucleoside phosphate metabolic processes, known to be central for viral replication and survival (31). A complete list of all the upregulated enriched terms as well as the parent-child relations identified for various SARS-CoV-2 infected cell lines is provided in **Supplementary Table S3**.

In addition to biological process, we also looked at the molecular function associated with the top 100 central genes present in the Activated TopNets across these cell lines (**Supplementary Table S4**). In the case of A549, we saw the enrichment of molecular functions associated with DNA-binding transcription factor activity, calcium-dependent protein kinase C activity, protein binding, etc. All these processes have been shown to be important for SC2 infection. For instance, the complex of viral S protein and host *ACE2* receptor binds to calcium ions (EF-hand domain) and shows protein kinase activity (32). After binding, the spike protein is further cleaved by *TMPRSS2*, a transmembrane protein serine 2 leading to further downstream signaling processes (33). As observed with biological processes, molecular functions for Vero and Calu3 cell lines were very similar. They were enriched for the functions such as protein kinase activity, TNF receptor superfamily binding, cytokine receptor binding, etc. These findings are supported by previous published literatures (34). Lastly, for NHBE cell line, we saw molecular functions such as small molecule binding, carbohydrate derivative binding, etc. as shown in previous work (35). In stark contrast, for top 100 DEGs, only one statistically significant molecular function was enriched across all cell lines, namely, “sequence specific DNA binding function” in the Vero cell line. The enrichment of regulatory and signaling molecules among the PathExt genes underscores the rationale that PathExt attempts to identify key genes mediating the transcriptional response.

Figure 2D summarizes the pathways enriched among the top 100 central genes in the repressed TopNets. A549 cell lines were highly enriched in cell cycle, DNA replication, and DNA damage response; interestingly, the enrichment was much greater in A549 with *ACE2*. These observations recapitulate the established biology of viral infection. For instance, genes mediating DNA damage response – *ATM* and *ATM/Rad3*-related (*ATR*) – were among the top 100 central genes. Key genes in Vero cells are enriched for mitotic spindle organization, ERBB signaling, EGFR signaling, protein phosphorylation, phagocytosis, etc., all previously known to modulate SARS-CoV-2 infection response (36). Lastly, in Calu3 cells, we saw enrichment of several lipid metabolism pathways. Lipids are one of the major components of viral structure and play an important role in its entry and replication inside the host cell. Several studies have shown the potential role of lipids and lipid induced metabolic changes in coronavirus infection (37, 38).

Other enriched pathways include leukocyte migration and phagocytosis. The downregulated enriched biological terms identified for various SC2 infected cell lines is provided in **Supplementary Table S5**.

We also performed molecular function analysis for the top 100 central genes in repressed TopNets (**Supplementary Table S6**). For A549 cell line, functions such as catalytic activity and single-strand DNA binding, were enriched. These processes are associated with SC2 attachment, infection and downstream signaling events as shown before (39). Likewise, for A549 cell line with high *ACE2* expression, “single-stranded DNA binding” was enriched. Genes associated with this function (e.g., *ATM* and *ATR*) are essential for response to DNA damage and repair, DNA metabolism, and maintaining genomic stability (40). In case of Calu3 cell lines, enriched functions included kinase activity, purine ribonuclease triphosphate binding, signaling receptor binding, etc. These functions play an important role in viral entry, membrane trafficking, signaling, etc. (34).

We investigated the commonality between the activated and repressed TopNets for the cell lines. We first confirmed that the top 100 central genes of the activated and repressed TopNets of the same cell line were disjoint, with a few exceptions: out of 100 genes, 1 is common in Calu3 cell line, 2 in A549, 3 in A549 with high *ACE2* expression, and 5 in the Vero cell line, suggesting dual activation and repressive roles of certain key genes; for instance, *ATM*, a protein kinase which plays key roles in many processes such as cell cycle progression, cell metabolism and growth, oxidative stress and chromatin remodeling, and is upregulated as well as downregulated in different cancers (41). Indeed, *ATM* is known to regulate these pathways in COVID-19, as shown in previous study (42). Despite disjoint central genes between activated and repressed TopNets, the enriched pathways among the central genes show greater commonality, such as protein phosphorylation, suggesting that these key processes may be involved in both activation and repression of different downstream processes.

Comparison of SARS-CoV-2 Infection With Other Viral Infections

Next, we compared the host transcriptional response to SC2 infection with those for other respiratory viruses in Calu3, Vero, MRC5, NHBE and A549 cell lines where there was data for an additional respiratory virus (RSV, IAV, HPIV3, SARS-CoV-1 and MERS). Top 100 central genes from each viral TopNet (activated & repressed) were used for comparison. As shown in **Figures 3A, B**, while there are small but significant overlaps in most cases, the responses could be considered largely virus specific.

Amongst activated TopNets, the SC2 response shares at least 10 central genes with responses to influenza virus, RSV, SARS-CoV-1 and MERS virus in different cell lines. SC2 shared 29 genes with MERS in the MRC5 cell line. Influenza virus in NHBE (bronchial epithelium) lacking NS1 shared 27 genes with SC2-infected Calu3 (Lung epithelium) and 25 genes with SC2-infected Vero cell line. The commonly shared genes between SARS-CoV-2 and most of the other viruses included *MAPK1*, *IL7*, *LYN*, *STAT3*, *TRAF6*, *NFKB1*. These genes are associated

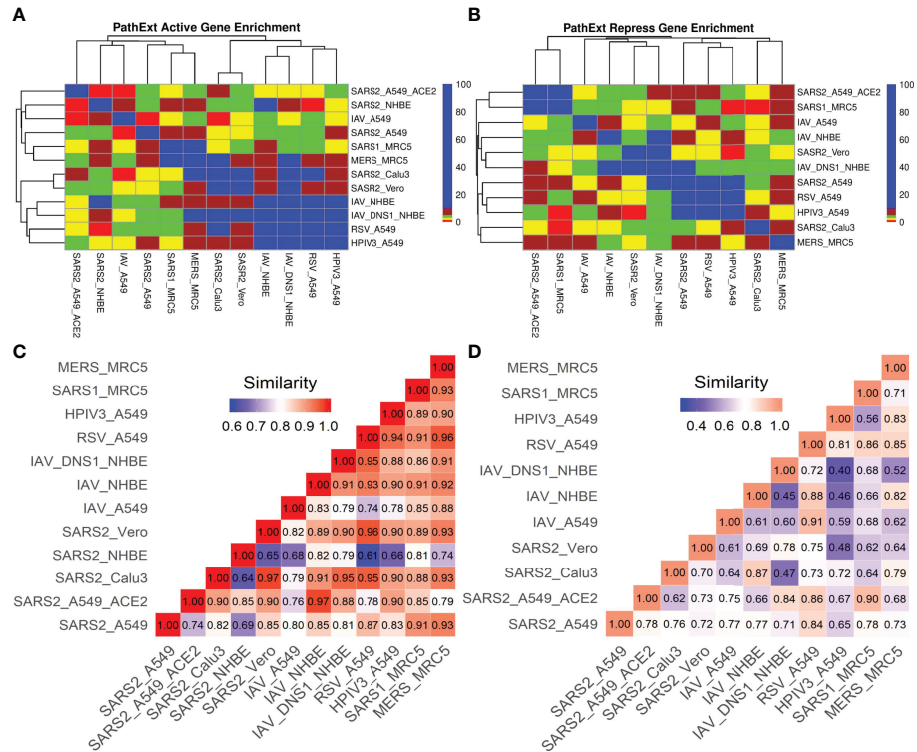


FIGURE 3 | SC2 infection comparison with other viruses. We analyzed the transcriptional response to different viruses including SC2 in various cell lines. We obtained the top 100 central genes from each cell line for activated & repressed TopNets and compared the gene commonality across various viruses in different cell lines as shown in gene enrichment plot (Observed/Expected overlap) for activated TopNets (**A**) and repressed TopNets (**B**). Semantic similarity among the enriched biological processes observed in different viruses activated and repressed networks (**C**, **D** respectively) shows higher similarity.

with pathways such as regulation of MAPK, response to cytokine, positive regulation of cell differentiation, and innate and adaptive immune response, all of which are supported by previous experimental work (15). These overlaps among the top 100 genes are significant and suggest shared responses across viruses. In repressed TopNets, a similar overlap trend was seen. SC2 shares at least 10 genes with all the viruses in different cell lines except MERS. Among the top 100 central genes in the repressed TopNet, SC2 shares 13 common genes with RSV and HPIV3 in the A549 cell line. Some of the commonly shared genes between SC2 and other viruses included *E2F1*, *BRCA1*, *RAD21*, *CDK*, *DDAH2*. These genes are associated with pathways such as cell cycle process, apoptotic process, cellular nitrogen compound metabolic process, DNA repair, etc., again revealing known shared responses across viruses (15, 43).

Next, we assessed the similarity in responses between SC2 and other viruses at the pathway level, using PANTHER and REVIGO. Given the enriched terms in the TopNets for all the viruses, we computed the semantic similarity using the GOSemSim package in R (44). For the activated TopNet, SC2 shares high similarity with nearly all the viruses considered in the study (**Figure 3C**). Highest similarity of 0.97 is observed between the processes enriched in SC2-infected A549 with high ACE2 expression, and Influenza-infected NHBE. Activated TopNets

across viruses share cytokine signaling, NF-kappaB signaling, inflammatory response, DNA binding transcription factor activity, and protein phosphorylation, all of which are well established host responses to respiratory viral infections. The processes which were uniquely enriched for SC2 infection include ATP metabolic process, G1/S transition of mitotic cell cycle, post-translational protein modification, and carbohydrate derivative metabolic process. Some recent published studies also account for these processes associated with SC2 infection (42, 45). A complete list of enriched biological pathways in activated TopNets for SC2 and other viruses is provided in the **Supplementary Table S3**.

Compared to activated TopNets, repressed TopNets exhibited greater divergence in terms of enriched processes (**Figure 3D**), but consistent with activated TopNets, we observed a greater similarity across different cell lines infected by SARS-CoV-2. Key enriched processes across viruses include DNA replication, telomere organization, DNA metabolic process, cellular response to DNA damage stimulus, regulation of MAPK cascade, negative regulation of apoptotic process, regulation of cell migration, nuclear division, regulation of cell cycle, etc., all of which have literature support (46). However, the unique processes enriched with SC2 infection include activation of phospholipase C activity, cardiomyocyte differentiation, leukocyte activation and migration,

and phagocytosis. Complete list of the enriched biological pathways in repressed TopNets for SC2 and other viruses is provided in the **Supplementary Table S5**. Overall, while as expected the functional response to SC2 infection is similar with the response to other respiratory viruses, our analysis reveals unique aspects of SC2 response. Functional relevance of the unique processes enriched among central genes in host response to SC2 response, however, will require further experimental follow up.

PathExt Provides Unique Insights Compared With DEG Analyses

Next, to assess advantage or complementarity of PathExt relative to DEG-centric approach, we compared the genes and pathways identified by PathExt with DEGs and their enriched pathways. Recall that PathExt identifies central genes that potentially mediate differential expression of other genes but may not be differentially expressed themselves. To test this, we selected the SC2-infected Calu3 cell line data and analyzed the log-fold change in expression of the top 100 central genes identified by PathExt TopNets and DEGs. As shown in **Figure 4A**, log-fold change of the genes present in the PathExt TopNets are much smaller than those of the DEGs, and therefore would systematically go undetected in a differential expression analysis. Similar results were seen for other SC2 infected cell lines as well (**Supplementary Figure S1**).

Next, in each SC2-infected cell line, top 100 upregulated genes were compared with the top 100 central genes in activated TopNet, and top 100 downregulated genes were compared with the top 100 central genes in repressed TopNet. As shown in **Figures 4B, C**, PathExt genes and DEGs were largely distinct. Given minimal overlap between PathExt central genes and DEGs, we compared them in terms of enriched processes. Remarkably, in stark contrast with PathExt central genes, the 100 most downregulated DEGs do not reveal any pathway enrichment in any of the cell lines, and the upregulated DEGs showed pathway enrichment only in Calu3 and Vero cell lines (**Supplementary Table S7**); in Vero cell line only one process was enriched and is therefore not discussed. As expected, pathways enriched among upregulated DEGs in Calu3 showed highest similarity with those in activated TopNet in Calu3 (**Figure 4D**), despite very little overlap in terms of genes. In Calu3, while the common pathways among PathExt and DEGs (**Supplementary Figure S2**) includes response to cytokine production, and immune system process, several pathways were uniquely revealed by PathExt, including regulation of DNA binding transcription factor activity, I-kappaB kinase/NF-kappaB signaling, regulation of cell death, and cellular response to organic substances. These pathways are well-associated with the COVID-19 infection as shown in multiple studies (21, 25), highlighting the relative advantage of PathExt over the conventional DEG approach.

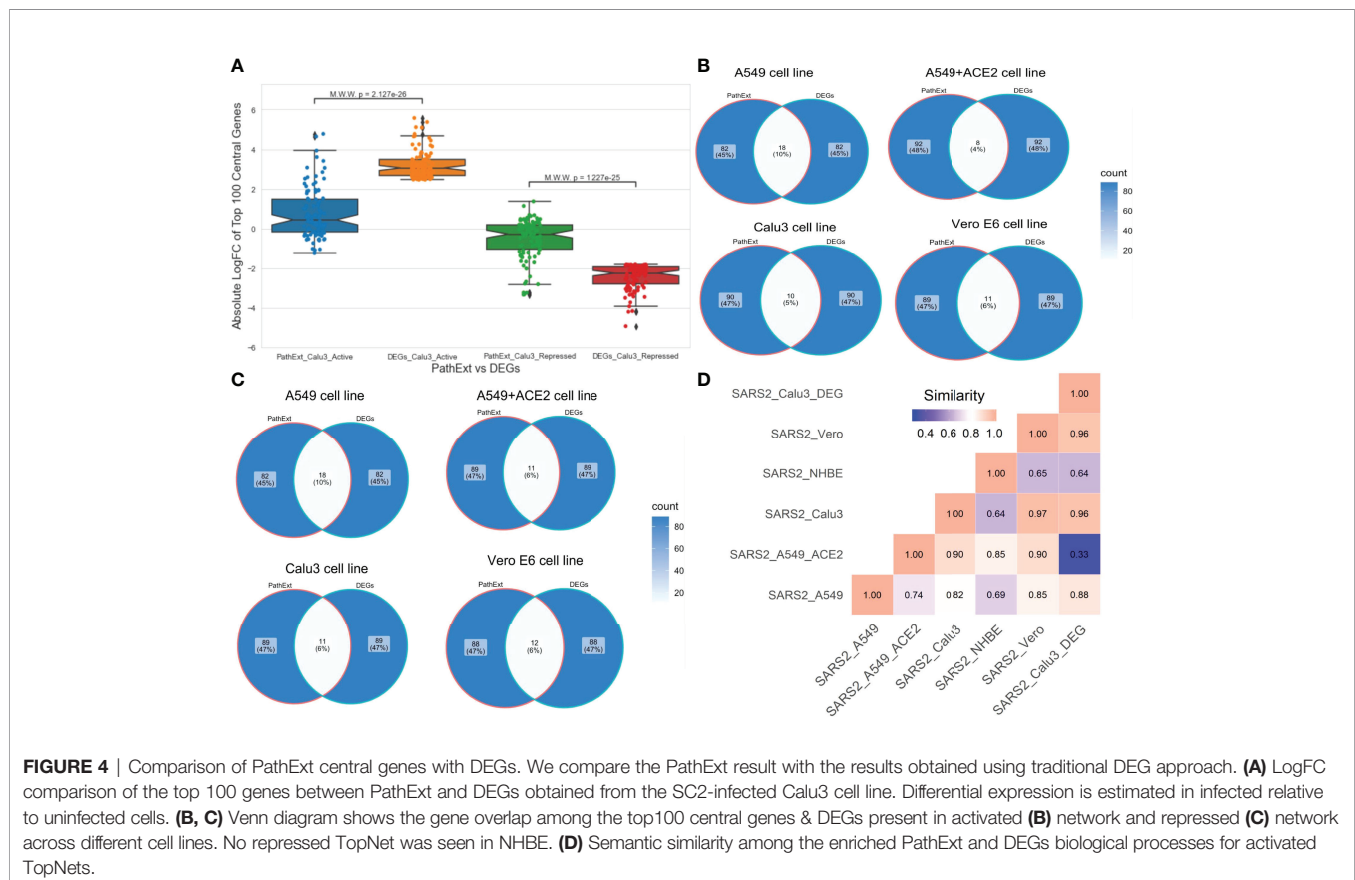


FIGURE 4 | Comparison of PathExt central genes with DEGs. We compare the PathExt result with the results obtained using traditional DEG approach. **(A)** LogFC comparison of the top 100 genes between PathExt and DEGs obtained from the SC2-infected Calu3 cell line. Differential expression is estimated in infected relative to uninfected cells. **(B, C)** Venn diagram shows the gene overlap among the top100 central genes & DEGs present in activated **(B)** network and repressed **(C)** network across different cell lines. No repressed TopNet was seen in NHBE. **(D)** Semantic similarity among the enriched PathExt and DEGs biological processes for activated TopNets.

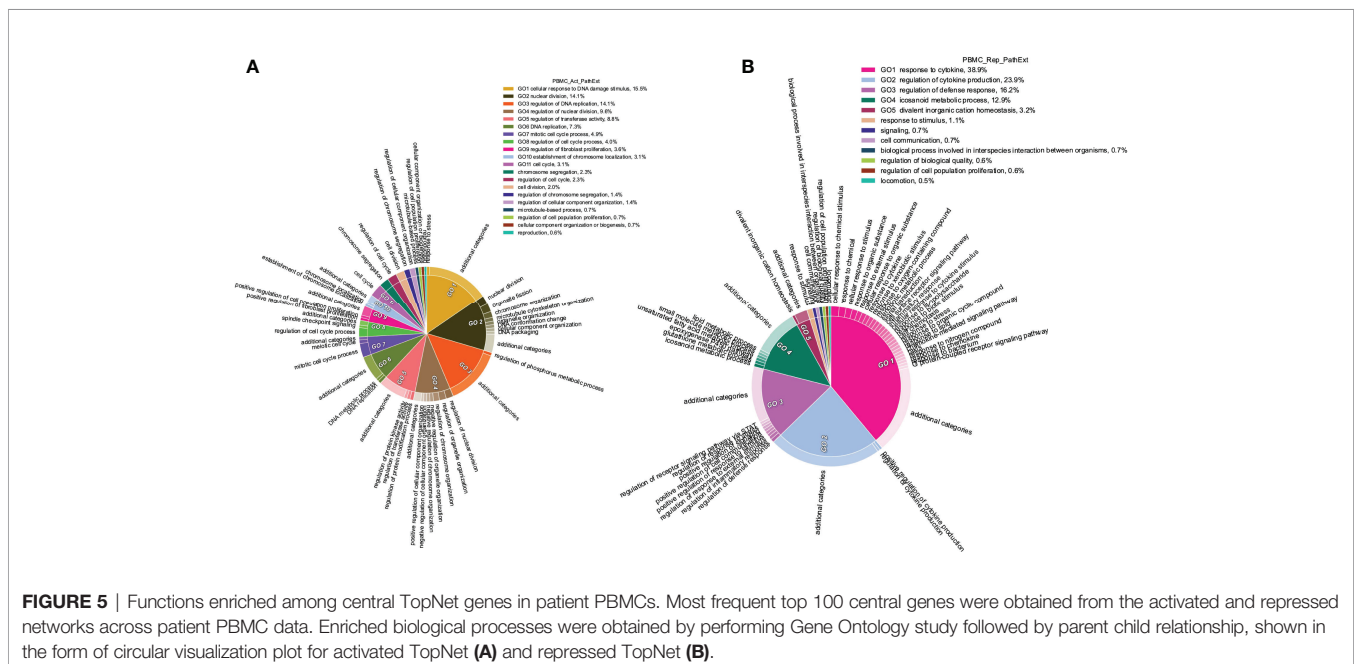
Application of PathExt to COVID-19 Patient PBMC

Next, we applied PathExt to 100 COVID-19 positive and 26 negative control individuals' PBMC transcriptomic data from (47), revealing activated TopNets in 96 samples and repressed TopNets in all 100 samples. As above, we first identified the top 100 most central genes in each TopNet, and then integrating across samples, we obtained the top 100 most frequent central genes separately for the activated and the repressed TopNets. Genes *PBK*, *CDC6*, and *BUB1*, were found to be the most frequent central genes among the activated TopNets occurring in 56%, 53%, and 48% of the 96 samples, respectively. Likewise, in the repressed TopNets, *IL6* was the most frequent, occurring in 49% of the samples, followed by *POMC* in 48% of the 100 samples; a complete list of top genes is provided in **Supplementary Table S8**. As shown in **Figure 5A**, pathways enriched among the frequent central genes in activated TopNets are predominantly related to cell cycle, which is an expected response to infection by the host immune system (48), as well as potentially linking SC2 infection with cancer (49). Previous studies have observed that cancer patients are more vulnerable to SC2 infection leading to adverse outcomes, likely due to compromised immune system (50). Some of the pathways such as uncontrolled production of cytokines, type-I interferon (*IFN-I*), dysregulation of immune checkpoint signaling, etc. are common in both COVID-19 and cancer. Complete list of enriched terms associated with top 100 most frequent central genes in the activated TopNets are provided in the **Supplementary Table S9**. Further, instead of using only the 100 most frequent central genes, if we consider all the unique central genes, we see enrichment of additional established pathways such as receptor signaling *via* JAK-STAT, cellular response to interleukin-1, etc. Molecular function enrichment analysis revealed cell cycle and replication processes such as single-stranded DNA binding,

cyclin dependent protein serine/threonine kinase regulator activity, etc. These are well known functions which occur post SC2 infection (4). Interestingly, top 100 upregulated DEGs show only one molecular function enriched, namely, 2'-5'-oligonucleotide synthetase activity (**Supplementary Table S10**).

In contrast, pathways enriched among the frequent central genes in the repressed TopNets included response to cytokine, regulation of cytokine production, regulation of defense response and icosanoid metabolic process (**Figure 5B**). Suppression of regulatory mechanisms that keep cytokine production in check is consistent with observed cytokine storms in COVID-19 patients and ensuing damage of organs such as the liver leading to downregulation of various metabolic processes (51). Likewise, SC2 can suppress the host immune defense by downregulating the T-cell function (52), cytokine production and cell-cell adhesion (51). Complete list of enriched terms associated with the top 100 most frequent central gene is provided in the **Supplementary Table S11**. Molecular functions associated with top 100 central genes include receptor ligand activity, heme binding, chemokine binding, etc. Once again, these are well known functions which occur post SC2 infection. However, downregulated DEGs were found to be enriched only for retinol binding function. Complete list of molecular function associated with repressed TopNet in PBMC is provided in **Supplementary Table S12**.

The cellular response in COVID-19 patient PBMC likely reflects immune response to systemic infection by SC2. However, there is some evidence that SC2 can infect immune cells as well (53–55), and therefore, the cellular response in PBMC could in part reflect endogenous response by the infected PBMC cells. To distinguish between these two possibilities, we identified the enriched pathways separately for the PBMC central genes shared with those in other cell lines (reflecting endogenous



response to infection) and the ones unique to PBMC (potentially reflecting immune response). For activated TopNet, we therefore compared the 1,830 unique genes appearing among the top 100 central genes in any TopNet across patient PBMC samples with the 395 unique genes among the 100 central genes across 5 cell lines. The 156 common genes were enriched in the similar pathways as observed in cell lines, such as regulation of DNA-binding transcription factor activity, regulation of I-kappaB kinase/NF-kappaB signaling, etc. The remaining PBMC-specific 1,674 genes were enriched for the pathways such as response to cytokine, humoral immune response, cellular response interleukin-4, and leukocyte proliferation, representing immune response to systemic infection. Likewise, for the repressed TopNet, we compared 1,877 unique genes in PBMC TopNets with the 368 genes across cell lines. The 148 common genes were enriched for the pathways such as positive regulation of cell migration, response to DNA damage stimulus, phosphatidylinositol metabolic process, etc. The remaining PBMC-specific 1,729 genes were enriched for T-helper-2 cell differentiation, cellular response to interleukin-18, positive regulation of MHC class II biosynthetic process, regulation of eosinophil migration, regulation of neutrophil mediated cytotoxicity, suggesting an overall suppression of immune response to systemic infection. These results are consistent with previous studies (56, 57). Zhou et al. have shown that an individual who has gone through COVID-19 and has started testing SC2-negative may still exhibit cold symptoms due to decreased expression of adaptive immune related genes, especially those related to T and B cells and HLA molecules, making them susceptible to secondary infections. They also showed that the suppression of the adaptive immune system could be due to dysregulated host response and not because of immune checkpoint molecules such as PD-1, PD-L1, CTLA4, etc. (58). Overall, PathExt reveals immune cell response to SC2 infection and

discriminates, to some extent, the potential cell-endogenous responses by infected PBMC and response to systemic infection. While it is possible that unique central genes in PBMC could simply reflect the inter-cell type heterogeneity in response to infection, we note that the fraction of unique genes in PBMCs are far greater than those in other cell types and we speculate that some of the unique gene genes and pathways in PBMCs may reflect systemic immune response. All enriched GO processes for activated and repressed TopNets in patient PBMCs common with cell lines are provided in **Supplementary Table S13** whereas enriched GO processes for activated and repressed TopNets unique to patient PBMCs are provided in **Supplementary Table S14**.

Next, we investigated the association between PathExt-identified most frequent central genes in PBMCs and the available demographics and clinical features of the patients – age (≤ 60 years or >60 years), sex (male vs female), and severity (ICU vs non-ICU). We found that most frequent genes in activated TopNets were more frequently observed in severe cases (**Figure 6A**). We did not see a direct association between frequent central genes and sex or age. Frequency and distribution of the PathExt identified genes in patients with severe illness is consistent with previous findings (47). To further probe into association of frequent central genes with severity, we identified genes that were uniquely central in either ICU or non-ICU patients, with a minimum frequency of 5 (**Supplementary Table S15**). Central genes in ICU patients included genes like *CDC25C*, *LOX*, *TGFB3*, *CCNF*, etc. and were enriched for mitotic cell cycle phase transition and regulation of cell death. Central genes in non-ICU patients included genes such as *CCR10*, *IRF6*, *CCL4*, etc. and were enriched for cytokine-mediated signaling pathway, cell communication, positive regulation of ERK1 and ERK2 cascade (**Supplementary Table S16**).

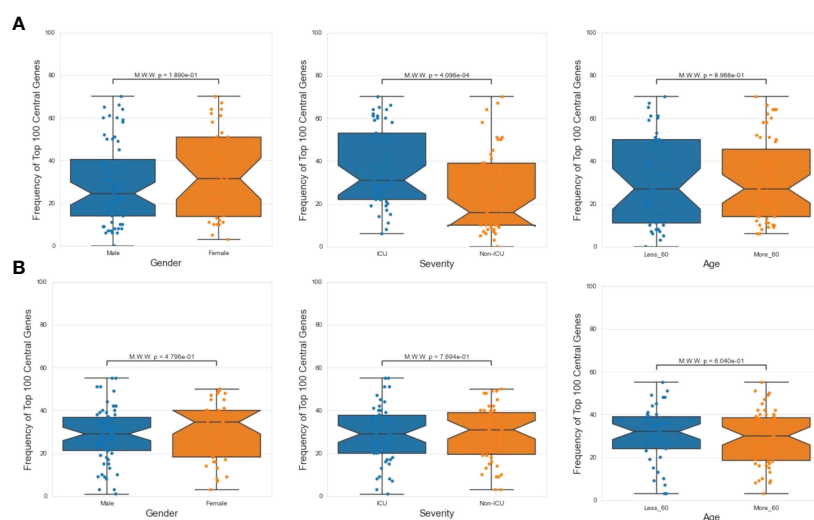


FIGURE 6 | Demographic features analysis. Mann Whitney Test was performed to check statistical significance between top 100 central genes and various demographic features (age, sex and severity). In case of PathExt identified top genes, “Severity” was found to be the only statistically significant group among activated TopNet genes (**A**). However, no group was statistically significant in case of repressed TopNet (**B**).

The analysis above identifies frequent central genes in ICU and non-ICU patients. We further aimed to assess whether the TopNet neighbors of the central genes are similar across samples, which would indicate a homogeneity in the perturbed paths mediated by a central gene. Toward this, we quantified, for each central gene independently, all sample-pair overlap (quantified by Jaccard Index or JI) between the sample-specific TopNet neighbors of the central gene. We did this separately for ICU and non-ICU patients; **Supplementary Figure S3** shows the distribution of JIs for the central genes in ICU and non-ICU patients separately, suggesting a greater homogeneity of response mediated by the central genes in the ICU patients.

We did not notice any association between frequent genes in the repressed TopNets and demographic features (**Figure 6B**). For DEGs, top central genes were more associated with male in activated TopNets and with severe cases in repressed TopNets (**Supplementary Figure S4**).

As a point of comparison, we also identified the upregulated and downregulated DEGs from the COVID-19 patients. As above, we selected the top 100 upregulated and downregulated genes in each patient sample and then the top 100 most frequent upregulated and downregulated genes across all 100 patients. First, we note that while frequent central genes in activated TopNets exhibit slightly lower fold changes compared to frequent DEGs, the frequent central genes in repressed TopNets exhibit far lower fold changes compared to DEGs (**Supplementary Figure S5**).

Next, we analyzed the commonality among the PathExt and DEGs top 100 genes. We found 28 genes to be common among the activated TopNets and upregulated DEGs and 8 were common among the repressed TopNets and downregulated DEGs. Next, we analyzed the similarity at the pathway level. While many processes enriched among frequent central genes were also enriched among upregulated DEGs (e.g., cell cycle) (**Supplementary Figure S6**), the PathExt central genes uniquely revealed cellular response to DNA damage stimulus, regulation of transferase activity, regulation of fibroblast proliferation, establishment of chromosome localization, etc. However, the frequent downregulated DEGs did not identify any significant enriched process, underscoring the relative advantage of the PathExt approach.

PathExt Reveals Previously Ascertained Anti-SC2 Target Genes Far Better Than DEGs

Next, we assessed the extent to which the PathExt-identified genes and DEGs recapitulate previously proposed anti-SC2 target genes based on experimental screens. Toward this, we compared the top 100 central genes identified in SC2-infected cell lines and patient PBMC data against previously published benchmark datasets of anti-SC2 targets. In total, we compiled 11 gene sets from our study: 6 for activated TopNets (5 for the cell lines and 1 for the PBMC cohort) and 5 for repressed TopNets (4 for the cells lines and 1 for the PBMC cohort); while for each cell line we considered the top central genes, for PBMC, we selected the 100 most frequent central genes across patients. We compiled 9 benchmark genes sets from previously published reports which

includes CRISPR gene-knockout studies, viral-host protein-protein interactions (PPI) and targets associated with experimentally screened drugs in various cell lines and animal models (Methods). Next, we assessed overlap between our 11 gene sets and the 9 different benchmark datasets using Fisher's Exact test, resulting in 99 tests. In 18 of the 99 comparisons (expectation is ~5 at p-value threshold of 0.05) PathExt gene sets significantly overlapped with the benchmark gene sets (**Figure 7A**). In sharp contrast, analogous sets of DEGs significantly overlapped with gold sets in only 5 cases, as expected by random chance (**Figure 7B**), again underscoring the relative advantage of PathExt.

Identifying Novel Potential Anti-SC2 Targets and Drugs

Next, to identify novel drug targets against SC2, we removed the already known targets (Methods) against SC2 and other viruses (considered in this study) from the TopNets (activated & repressed). Based on the frequency of these unique genes in the TopNets of various SC2 infected cell lines, we proposed novel anti-SC2 targets namely *ADCY2*, *ADSL* (mediating activated TopNet), and *OCRL*, and *TIAM1* (mediating repressed TopNet). Similar analysis performed for the PBMC data reveals that genes like *PBK* and *BUB1* can be potential new targets. Even though the inter-cell type overlap at the level of key genes is not high, the overlap at the pathway level is much greater (**Figures 2C, D**). Insofar targeting the key genes may impact the specific pathways, the impact of targeting these genes may nevertheless be broad. Complete list of the potential targets observed in both cell line and PBMC is provided in the **Supplementary Table S17**. The gold standard dataset used to ascertain novel targets was created in March 2021. Instead of using an updated dataset, we decided to further assess the accuracy of our proposed novel targets in a prospective manner: a quick survey of the literature published since March 2021 validated some of the targets that we have identified, for example *ESRRA* (59), *PTGDR2* (60), *EGFR* (59), etc. This serves as a prospective validation of the targets proposed by PathExt. Finally, we performed virtual screening (Methods) to identify potential inhibitors against the proposed novel targets in our study. We propose the top 5 potential drug molecules for each target (cell line and PBMC) in **Table 1**.

All the drugs listed in **Table 1** are either FDA approved or are currently undergoing clinical trial against various diseases, e.g., Nilotinib to treat chronic myeloid leukemia (61), Tirilazad against acute ischemic stroke (62), etc. Extensive drug repurposing has been carried out recently to treat SC2 infection and corroborate our proposal. Nilotinib has been shown to be effective against covid-19 (63); Conivaptan have been shown to be effective against covid-19 based on *in silico* studies, where conivaptan targets viral non-structural protein 9 (Nsp9) (64). Likewise, Phthalocyanine has been shown to be effective against preventing covid-19 in randomized trials (65). Bemcentinib has been shown to be effective against SC2 (66) and is being tested (trial NCT04890509) for efficacy in hospitalized covid-19 patients. Bemcentinib was designed for targeting AXL, a tyrosine kinase which signals *via* PI3K (67); however, our analysis identifies it as a lead molecule

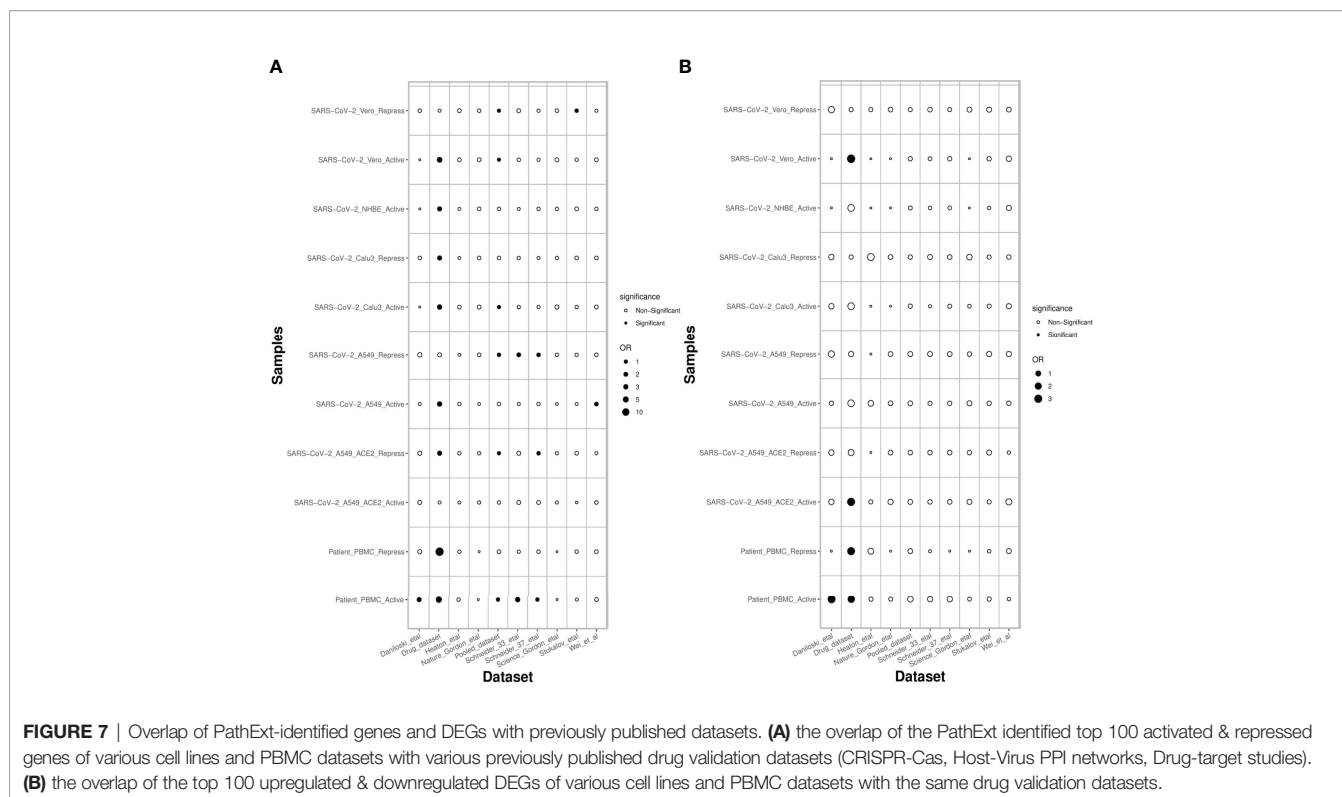


TABLE 1 | Proposed small molecules potentially inhibiting the novel anti-SC2 targets.

Sr No.	Target	Proposed Drug Molecules
Cell Line		
1	ADCY2	Phthalocyanine, Tirilazad, Temoporfin, Telcagepant, Laniquidar
2	ADSL	Nilotinib, Lixivaptan, Telcagepant, Doramapimod, Bemcentinib
5	OCRL	Conivaptan, Implipitapide, Ergotamine, Fluspirilene, Tolvaptan
7	TIAM1	Cipargamin, MK-3207, Adozelesin, Hypericin, Tariquidar
PBMC		
8	BUB1	Hypericin, Erismodegib, Hemin, Irinotecan, TMC-647055
10	PBK	Conivaptan, Bemcentinib, Dihydroergotamine, Phthalocyanine, UK-432097

against ADSL and PBK. We use GeneMANIA software (68) to probe potential relationship between ADSL or PBK with AXL. As shown in **Figure 8A**, there is a physical interaction between AXL and PBK *via* PIK3R2. Likewise, interaction can be seen among AXL and ADSL in **Figure 8B**, suggesting that Bemcentinib’s effect may be mediated by multiple genes within a closely linked gene module. Our computationally generated hypotheses however need experimental validations through knockout or induction studies.

DISCUSSION

We applied PathExt to investigate the response to SC2 infection in different cell lines (lung epithelium, bronchial epithelium and kidney cells), as well as in patient derived PBMCs. We also compared the response to SC2 with those for other respiratory viruses. Some of the key processes enriched among the PathExt-

identified genes involve the immune system, cytokine signaling, metabolism, DNA replication and transcription, lipid mediated processes, etc. Our analysis revealed both similarity and dissimilarity in response to SC2 across different cell lines, reflecting different strategies evolved in different cell types likely governed by their specific regulatory networks. For instance, as A549 cell line, with low baseline ACE2 expression, has a lower viral load, its response, in terms of the enriched pathways among central TopNet genes, were different in comparison to Vero and Calu3 cell lines. Several previous studies have shown the links between viral load dynamics and disease severity. Viral load affects the host gene expression and downstream response pathways (55). Blanco et al. have shown that only 0.1% viral reads were detected post SC2 infection in A549, whereas in Calu3 cell line, 15% of the reads were detected. They also observed different host transcriptional landscape in different cell lines based on virus replication rate post SC2 infection (15). Comparing the response to SC2 with

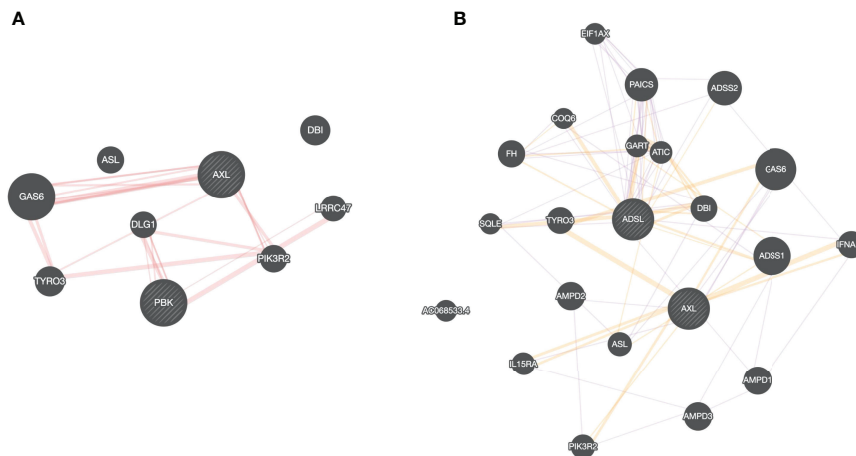


FIGURE 8 | Drug-target association. Based on our virtual screening study, we identified, Bemcentinib as a potential inhibitor for PBK and ADSL. Bemcentinib is a well-known drug against AXL, and we saw that this gene (AXL) is connected to our proposed target PBK (A) and ADSL (B) suggesting that the drug effect may be mediated by multiple genes within a closely linked gene module.

those of other respiratory viruses (Influenza, SARS1, MERS, HPIV3 and RSV), we noted that the central TopNet genes are largely virus specific. However, the broad biological processes enriched among these genes exhibit a much greater degree of similarity across viruses, suggesting a virus-specific host response network that nevertheless affects common phenotypic response. Some common processes enriched among all the viruses include cytokine signaling, defense response, cell cycle regulation, inflammatory response, etc. However, processes like ATP metabolic process, carbohydrate derivative metabolic process, cardiomyocyte differentiation, phagocytosis, leukocyte activation and migration, etc. were found to be specific to SC2 response. Though activated and repressed TopNets shared very few central genes, they shared several processes enriched among the central genes, including cell cycle, phosphorylation, metabolism, etc., suggesting that these biological processes likely mediate both upregulation and downregulation of various genes in the global transcriptomic response to SC2 infection.

A previous study by Thair et al. (69) has reported the comparison of differentially expressed genes in SARS-CoV-2 infection with other viral infection models. There are key similarities between their findings and ours in terms of enriched processes. For instance, both our studies identified cytokine signaling, inflammatory response, DNA binding transcription factor activity, etc. However, PathExt uniquely identified ATP metabolic process, carbohydrate derivative metabolic process, cardiomyocyte differentiation, etc. to be active in SARS-CoV-2 infection.

Conventional DEG-centered approaches can miss the genes which are not significantly differentially expressed but are nevertheless responsible for mediating, for instance, based on their post-translational modification state, the global transcriptional response in a given condition. PathExt addresses this limitation by focusing on genes that critically mediate significant path-level perturbations in the network. Thus, as expected, many of the central genes identified by PathExt exhibit much lower expression

fold change relative to DEGs. Interestingly, some genes are identified as central in both activated as well as repressed TopNets (a feature unique to PathExt), consistent with pleiotropic function of genes. PathExt identified genes are largely disjoint from DEGs. In contrast to top PathExt genes, the top upregulated DEGs show significantly enriched pathways only in the Calu3 and Vero cell lines, with only one pathway enriched in the Vero cell line. The top downregulated DEGs exhibited no enrichment at all in any cell line. Though PathExt based approach have several advantages over conventional DEGs based approach, one of its limitations, indeed of any network-based approach, is the dependence on the knowledge of the protein-protein interaction knowledge-based network, which may be biased toward more studied proteins and have both false positives and false negatives.

Similarly in patient PBMC dataset, while the central genes both in activated TopNet as well as in upregulated DEGs were enriched for cell cycle and defense response, the downregulated DEGs did not show any enrichment while key genes in repressed TopNet were enriched for cytokine production. The pathways enriched among central genes in PBMC TopNets likely reflect both the innate endogenous response of immune cells to viral infection as well as response to systemic infection. However, a vast majority of central genes identified in PBMCs are specific to PBMCs, suggesting that transcriptional changes in PBMCs are largely in response to systemic infection. Demographic features were not statistically significant except “severity” category in Activated TopNet of PathExt. Central genes unique to ICU patients were associated with cell cycle processes whereas unique genes associated with non-ICU patients were enriched for cytokine mediated processes. Central genes in ICU patients exhibited a much greater homogeneity in their TopNet neighborhood across patients compared with central genes in non-ICU patients, suggesting that disease severity may be mediated by a more conserved genes and processes. Importantly, PathExt was substantially more effective in identifying potential anti-SC2 drug targets. PathExt-identified targets were enriched in 18% of the gold

set comparisons, compared with only 5% (NULL expectation) for DEGs.

Previous studies have noted that genes with variable expression are more likely to be detected as differentially expressed in multiple contexts and do not reveal context-specific functional responses (9). Compared to reliance only on the differential expression, by exploiting the knowledge-based gene networks, and focusing on identifying key genes associated with significantly perturbed paths, PathExt represents a complementary, and in important ways, more effective approach. While previous works have exploited protein networks to infer transcriptomic perturbations, they have still relied on significantly differentially expressed genes and interpreted them in the context of the network (70, 71). We have previously demonstrated (11), superiority of PathExt over such integrative approaches that nevertheless rely on significant differential expression.

In summary, our work (1) further establishes the utility of PathExt, (2) provides a comparative analysis of key genes potentially driving the global transcriptomic response to SC2 and other respiratory viruses and across multiple cell lines and patient PBMCs, (3) identifies target genes validated in previously published benchmark anti-COVID-19 target gene datasets, far better than DEGs, (4) proposes novel targets against COVID-19, and (5) proposes FDA approved drugs or drugs in clinical trials, against the novel targets.

METHODS

Data Collection and Processing

We collected 12 datasets from 2 studies, one published (15) and another unpublished (72), which include pre- and post-infection transcriptome in 5 different cell lines (NHBE, A549, Calu3, Vero, MRC5), each infected by one or more of the 6 viruses (SARS-CoV-2, SARS-CoV-1, MERS, RSV, HPIV3, and Influenza virus). In addition, we also obtained PBMC transcriptomes from 100 COVID-19 patients and 26 non-infected individuals as controls (47). In total, we had 13 datasets from 3 different studies, details of which are provided in **Table 2**. For each dataset (except Vero and PBMCs), we downloaded the raw reads using prefetch (73); most datasets were single end reads except Vero for which paired

end reads were provided. Files were split using fastq-dump command (73), trimmed using Trim Galore-0.6.6 (74) at default parameters and the reads mapped to the human transcriptomic index version hg38 using SALMON v.1.12 (75). We proceeded only with those samples for which at least 60% of the reads were mapped. In the case of Vero cell line and patient PBMC data, we directly downloaded the TPM (transcripts per million) values provided from the GEO.

Gene Expression Normalization and Node Weight Computation

For every cell line, mean gene expression was computed for the treated and control samples. Genes with TPM value 0 or greater than 100 were removed. For patient PBMC data, each case was analyzed individually and for the control, we took the mean expression of 26 control samples. Data was further filtered by removing the genes whose expression value was either 0 or above 100 in at least 50% of the PBMC case samples. Lastly, we also excluded the genes from the cell lines and PBMC data, which were not present in our human protein-protein interaction network. After the above filters, there were a total of 7,740 genes for cell lines (except Vero), 6,852 for Vero cell line. In case of patient PBMC datasets, the main dataset comprises 12,417 genes (**Supplementary Table S18**). All data was quantile normalized as done in previous study (76). Lastly, the normalized data was used to compute node weights which was then used to compute differentially expressed paths using PathExt software. Node (gene) weights to identify activated TopNet were computed as follows. While $\log(\text{fold change})$ is reasonable choice for node weight, given the dependence of the magnitude of $\log(\text{fold change})$ on the expression value, we instead computed the expected $\log(\text{fold change})$ for a given control expression and then used the difference of observed and expected $\log(\text{fold change})$ as the node weight. To compute the expected absolute $\log(\text{fold change})$, we regressed absolute $\log(\text{fold change})$ values across all genes against the gene expression in control samples using Loess fit, implemented in R (77). For activated TopNets the fold change was computed in cases relative to control, and for repressed TopNets the fold change was computed in control relative to cases. Computed node weights for Activated and Repressed TopNets in SC2 infected cell line (other

TABLE 2 | Summary of datasets used in the study.

Dataset Name	Description	Sample Size	Platform	Reference
SC2 ¹ _Vero	SC2 infected African green monkey kidney epithelial cell line	6	Bulk RNASeq	Riva et.al. (14)
SC2_Calu3	SC2 infected human lung adenocarcinoma cell line	6	Bulk RNASeq	Blanco et.al. (15)
SC2_NHBE	SC2 infected primary normal human bronchial epithelium cell line	6	Bulk RNASeq	Blanco et.al. (15)
SC2_A549	SC2 infected human alveolar basal epithelial cells	6	Bulk RNASeq	Blanco et.al. (15)
SC2_A549_ACE2	SC2 infected human alveolar basal epithelial cells with higher ACE2 expression	6	Bulk RNASeq	Blanco et.al. (15)
INF ² _A549	Influenza infected human alveolar basal epithelial cells	6	Bulk RNASeq	Blanco et.al. (15)
INF_NHBE	Influenza infected primary normal human bronchial epithelium cell line	6	Bulk RNASeq	Blanco et.al. (15)
INF_DNS1_NHBE	Influenza (lacking gene NS1) infected primary normal human bronchial epithelium cell line	6	Bulk RNASeq	Blanco et.al. (15)
HPIV3 ³ _A549	HPIV3 infected human alveolar basal epithelial cells	6	Bulk RNASeq	Blanco et.al. (15)
RSV ⁴ _A549	RSV infected human alveolar basal epithelial cells	4	Bulk RNASeq	Blanco et.al. (15)
SC1 ⁵ _MRC5	SC1 infected diploid cell culture line composed of fibroblast	6	Bulk RNASeq	GSE56192 (72)
MERS ⁵ _MRC5	MERS infected diploid cell culture line composed of fibroblast	6	Bulk RNASeq	GSE56192 (72)
SC2_PBMCs	SC2 infected peripheral blood mononuclear cells	126	Bulk RNASeq	Overmyer et.al. (47)

1: SARS-CoV-2; 2: Influenza; 3: Human Parainfluenza Virus; 4: Respiratory Syncytial Virus; 5: SARS-CoV-1; 6: Middle East Respiratory Syndrome.

than Vero) is provided in **Supplementary Table S19**, for Vero cell line is provided in **Supplementary Table S20** and for the PBMCs is provided in the **Supplementary Table S21**. Node weight computed for remaining viruses is provided in **Supplementary Table S22**.

Given the node weights across samples, the PathExt tool computes the significant paths in two steps: (1) Top 0.1% shortest paths are selected and (2) Statistical significance of those selected paths is estimated based on data randomization and multiple testing corrected q-value threshold. We selected those percentiles and q-values which provide at least 300 nodes for a given TopNet. Once the TopNets were generated, we computed the ripple centrality score for each gene in the TopNet and top 100 central genes were selected for the further analysis.

GO Enrichment Analysis

Enriched pathways in a given TopNet (activated and repressed) were analyzed using the top 100 central genes using PANTHER software. Customized reference was used as a background during the enrichment analysis, where we considered only those genes which were used for the TopNet creation in this study. The reference was different for the Vero cell line, cell lines other than Vero and the patient PBMC data. 'GO biological process complete' was selected as Annotation Data Set, 'Fisher's Exact' as Test Type, and 'Calculate False Discovery Type' as Correction method. As there could be redundant terms present in the result, we removed them by performing parent-child relation study using REVIGO software. The GO term and its corresponding FDR value was provided as an input with the resulting list option to be 'Medium (0.7)'. Also, in the 'Advanced options', we selected 'Yes' in the remove obsolete GO terms, 'Homo sapiens' as the working species, and the 'SimRel' as semantic similarity measures. Finally, the circular plot representing the parent-child GO terms was created using the CirGo software (78). This tool requires the output of the REVIGO as an input. 'GOSemSim' package was used to compute the semantic similarity among the processes enriched in various cell lines, among different viruses and the patient PBMC data. In GOSemSim package, to calculate the similarities between two GO terms, we used Wang method which uses the topology of GO DAG graph structures to compute the semantic similarities. To combine two sets of GO terms, we utilized 'rcmax' method which considers the average of maximum similarity on each row and column on the matrix consisting of the similarities among two sets of GO terms. Heatmap plots for the similarity analysis are visualized using ggplot2 R package (79).

Comparing Most Central Genes Across Cell Lines and Viruses

Upset plot was generated by providing the list of genes of all the samples in a csv file format as an input to the server. Intervene server (80) at default parameters was used for generating the upset plots. For gene enrichment heatmap, we computed the Observed/Expected score. 'pheatmap' package in R was used for generating the heatmaps (81).

Identification of Novel Drug Targets and Their Potential Inhibitors

First, we identified known drug targets of SC2 (till March 2021) from various published reports which includes CRISPR gene-

knockout studies, viral-host protein-protein interactions (PPI) and experimentally screened drugs in various cell lines and animal models. See **Supplementary Table S23**, for the references used for preparing this gold standard benchmarking datasets. These studies reported the host genes important for the viral replication and function. CRISPR-Cas studies provided the important therapeutic targets based on the gene knockdown studies and their downstream effects, whereas protein-protein interaction studies shed light on the targets physically interacting with the viral proteins. *In vitro* and *in vivo* studies in the cell lines and animal models provided the information about the genes which could be potentially targeted by different drugs. We excluded all these targets from our list of predicted targets. To provide specific recommendations for SC2, we also excluded the targets which were present in the TopNets (Activated & Repressed) for any other virus such as Influenza, HPIV3, RSV, SC1 and MERS. For other viruses, we removed all the genes present in the TopNets for the respective viral response in various cell lines. This left us with the targets unique to SC2 and we then prioritized the remaining central TopNet genes based on their frequency across cell line samples as novel potential drug targets. Similar approach was followed for the PBMC data.

Next, to propose potential inhibitors against our proposed novel targets, we performed a virtual screening process using AutoDock Vina software (82). The 3D structures of the targets were downloaded from the RCSB-PDB (83) and further refined using Open Babel software (84). Active site information of the protein molecules was computed using P2RANK software (85). Next, a drug library was created for the virtual screening process, where we considered only those drug molecules which are either FDA approved or are under clinical trials; SMILES formatted files of these drugs were downloaded from the ZINC database (86) and were further converted to MOL2 file format using openbabel. The ligand and the receptor files were prepared in the 'pdbqt' file format required by vina for the docking purpose. The center and the grid size of the receptor molecules was computed using UCSF Chimera software (87), based on the P2RANK software active site result. Lastly, based on the AutoDock Vina affinity score and Root Mean Square Deviation (RMSD) value, we proposed the drug molecules which could potentially inhibit the function of these new targets.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

PA created the dataset. PA and NS developed the code for the analysis. PA, NS and SH analyzed the results. PA, GO, and NS created the tables and figures. PA and SH wrote the manuscript.

SH supervised the study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by funding from the Intramural Research Program, National Institutes of Health, National Cancer Institute, Center for Cancer Research.

ACKNOWLEDGMENTS

Authors are thankful to Prof. Adolfo Garcia-Sastre and Dr. Kris White from Icahn School of Medicine at Mount Sinai for their excellent suggestions for improving the manuscript. The work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>). Authors also thank other lab members for their constructive suggestions for improving the quality of the work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2022.918817/full#supplementary-material>

REFERENCES

- Yang H, Rao Z. Structural Biology of SARS-CoV-2 and Implications for Therapeutic Development. *Nat Rev Microbiol* (2021) 19:685–700. doi: 10.1038/S41579-021-00630-8
- Wu M, Chen Y, Xia H, Wang C, Tan CY, Cai X, et al. Transcriptional and Proteomic Insights Into the Host Response in Fatal COVID-19 Cases. *Proc Natl Acad Sci U S A* (2020) 117:28336–43. doi: 10.1073/PNAS.2018030117
- Gil C, Ginex T, Maestro I, Nozal V, Barrado-Gil L, Cuesta-Gejjo MÁVerifytat, et al. COVID-19: Drug Targets and Potential Treatments. *J Med Chem* (2020) 63:12359–86. doi: 10.1021/ACS.JMEDCHEM.0C00606
- Shah VK, Fimal P, Alam A, Ganguly D, Chattopadhyay S. Overview of Immune Response During SARS-CoV-2 Infection: Lessons From the Past. *Front Immunol* (2020) 11:1949. doi: 10.3389/FIMMU.2020.01949
- Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, et al. Remdesivir for the Treatment of Covid-19 - Final Report. *N Engl J Med* (2020) 383:1813–26. doi: 10.1056/NEJMJA2007764
- Greasley SE, Noell S, Plotnikova O, Ferre RA, Liu W, Bolanos B, et al. Structural Basis for Nirmatrelvir *In Vitro* Efficacy Against SARS-CoV-2 Variants. *bioRxiv* (2022). doi: 10.1101/2022.01.17.476556
- Lim Y, Ng Y, Tam J, Liu D. Human Coronaviruses: A Review of Virus-Host Interactions. *Diseases* (2016) 4:26. doi: 10.3390/DISEASES4030026
- Lee JS, Park S, Jeong HW, Ahn JY, Choi SJ, Lee H, et al. Immunophenotyping of COVID-19 and Influenza Highlights the Role of Type I Interferons in Development of Severe COVID-19. *Sci Immunol* (2020) 5:eabd1554-69. doi: 10.1126/SCIIMMUNOL.ABD1554
- Crow M, Lim N, Ballouz S, Pavlidis P, Gillis J. Predictability of Human Differential Gene Expression. *Proc Natl Acad Sci U S A* (2019) 116:6491–500. doi: 10.1073/PNAS.1802973116/-/DCSUPPLEMENTAL
- Wong HSC, Guo CL, Lin GH, Lee KY, Okada Y, Chang WC. Transcriptome Network Analyses in Human Coronavirus Infections Suggest a Rational Use of Immunomodulatory Drugs for COVID-19 Therapy. *Genomics* (2021) 113:564–75. doi: 10.1016/J.YGENO.2020.12.041
- Sambaturu N, Pusadkar V, Hannehalli S, Chandra N. PathExt: A General Framework for Path-Based Mining of Omics-Integrated Biological Networks. *Bioinformatics* (2021) 37:1254–62. doi: 10.1093/BIOINFORMATICS/BTAA941
- Mishra S, Shukla P, Bhaskar A, Anand K, Baloni P, Jha RK, et al. Efficacy of β -Lactam/ β -Lactamase Inhibitor Combination is Linked to WhiB4-Mediated Changes in Redox Physiology of Mycobacterium Tuberculosis. *Elife* (2017) 6:e25624-54. doi: 10.7554/ELIFE.25624
- Sambaturu N, Mishra M, Chandra N. EpiTracer - An Algorithm for Identifying Epicenters in Condition-Specific Biological Networks. *BMC Genomics* (2016) 17 Suppl 4:543–56. doi: 10.1186/S12864-016-2792-1
- Riva L, Yuan S, Yin X, Martin-Sancho L, Matsunaga N, Pache L, et al. Discovery of SARS-CoV-2 Antiviral Drugs Through Large-Scale Compound Repurposing. *Nature* (2020) 586:113–9. doi: 10.1038/S41586-020-2577-1
- Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Moller R, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* (2020) 181:1036–45.e9. doi: 10.1016/J.CELL.2020.04.026
- Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER Version 14: More Genomes, A New PANTHER GO-Slim and Improvements in Enrichment Analysis Tools. *Nucleic Acids Res* (2019) 47:D419–26. doi: 10.1093/NAR/GKY1038
- Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One* (2011) 6:e21800-09. doi: 10.1371/JOURNAL.PONE.0021800
- Lokugamage KG, Hage A, de Vries M, Valero-Jimenez AM, Schindewolf C, Dittmann M, et al. Type I Interferon Susceptibility Distinguishes SARS-CoV-2 From SARS-CoV. *J Virol* (2020) 94:e01410-20. doi: 10.1128/JVI.01410-20
- Hemmat N, Asadzadeh Z, Ahangar NK, Alemohammad H, Najfzadeh B, Derakhshani A, et al. The Roles of Signaling Pathways in SARS-CoV-2 Infection; Lessons Learned From SARS-CoV and MERS-CoV. *Arch Virol* (2021) 166:675–96. doi: 10.1007/s00705-021-04958-7
- Li W, Qiao J, You Q, Zong S, Peng Q, Liu Y, et al. SARS-CoV-2 Nsp5 Activates NF- κ B Pathway by Upregulating SUMOylation of MAVS. *Front Immunol* (2021) 12:750969. doi: 10.3389/FIMMU.2021.750969

21. Choudhary S, Sharma K, Silakari O. The Interplay Between Inflammatory Pathways and COVID-19: A Critical Review on Pathogenesis and Therapeutic Options. *Microb Pathog* (2021) 150:104673–94. doi: 10.1016/J.MICPATH.2020.104673
22. Olbei M, Hautefort I, Modos D, Treveil A, Poletti M, Gul L, et al. SARS-CoV-2 Causes a Different Cytokine Response Compared to Other Cytokine Storm-Causing Respiratory Viruses in Severely Ill Patients. *Front Immunol* (2021) 12:629193. doi: 10.3389/FIMMU.2021.629193
23. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: Recent Insights Into Emerging Coronaviruses. *Nat Rev Microbiol* (2016) 14:523–34. doi: 10.1038/NRMICRO.2016.81
24. Kircheis R, Haasbach E, Lueftenegger D, Heyken WT, Ocker M, Planz O. NF- κ B Pathway as a Potential Target for Treatment of Critical Stage COVID-19 Patients. *Front Immunol* (2020) 11:598444. doi: 10.3389/FIMMU.2020.598444
25. Chetta M, Rosati A, Marzullo L, Tarsitano M, Bukvic N. A SARS-CoV-2 Host Infection Model Network Based on Genomic Human Transcription Factors (TFs) Depletion. *Heliyon* (2020) 6:e05010-17. doi: 10.1016/J.HELIYON.2020.E05010
26. Purcaru OS, Artene SA, Barcan E, Silosi CA, Stanciu I, Danoiu S, et al. The Interference Between SARS-CoV-2 and Tyrosine Kinase Receptor Signaling in Cancer. *Int J Mol Sci* (2021) 22:4830–42. doi: 10.3390/IJMS22094830
27. Fung TS, Liu DX. Post-Translational Modifications of Coronavirus Proteins: Roles and Function. *Future Virol* (2018) 13:405–30. doi: 10.2217/FVL-2018-0008
28. Mishra S, Bassi G, Nyomba BLG. Inter-Proteomic Posttranslational Modifications of the SARS-CoV-2 and the Host Proteins – A New Frontier. *Exp Biol Med (Maywood)* (2021) 246:749–57. doi: 10.1177/1535370220986785
29. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. *Nature* (2020) 583:459–68. doi: 10.1038/S41586-020-2286-9
30. Harcourt J, Tamin A, Lu X, Kamili S, Sakthivel SK, Murray J, et al. Severe Acute Respiratory Syndrome Coronavirus 2 From Patient With Coronavirus Disease, United States. *Emerg Infect Dis* (2020) 26:1266–73. doi: 10.3201/EID2606.200516
31. Bojkova D, Klann K, Koch B, Widera M, Krause D, Ciesek S, et al. Proteomics of SARS-CoV-2-Infected Host Cells Reveals Therapy Targets. *Nature* (2020) 583:469–72. doi: 10.1038/S41586-020-2332-7
32. Wenzhong L, Hualan L. COVID-19: The CaMKII-Like System of S Protein Drives Membrane Fusion and Induces Syncytial Multinucleated Giant Cells. *Immunol Res* (2021) 69:496–519. doi: 10.1007/S12026-021-09224-1
33. Huang Y, Yang C, feng X, Xu W, wen LS. Structural and Functional Properties of SARS-CoV-2 Spike Protein: Potential Antivirus Drug Development for COVID-19. *Acta Pharmacol Sin* (2020) 41:1141–9. doi: 10.1038/S41401-020-0485-4
34. Bouhaddou M, Memon D, Meyer B, White KM, Rezelj V v., Correa Marrero M, et al. The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell* (2020) 182:685–712.e19. doi: 10.1016/J.CELL.2020.06.034
35. Nanda P, Ghosh A. Genome Scale-Differential Flux Analysis Reveals Deregulation of Lung Cell Metabolism on SARS-CoV-2 Infection. *PLoS Comput Biol* (2021) 17:e1008860-82. doi: 10.1371/JOURNAL.PCBI.1008860
36. Appelberg S, Gupta S, Akusjärvi SS, Ambikan AT, Mikaeloff F, Saccon E, et al. Dysregulation in Akt/mTOR/HIF-1 Signaling Identified by Proteo-Transcriptomics of SARS-CoV-2 Infected Cells. *Emerg Microbes Infect* (2020) 9:1748–60. doi: 10.1080/22221751.2020.1799723
37. Casari I, Manfredi M, Metharom P, Falasca M. Dissecting Lipid Metabolism Alterations in SARS-CoV-2. *Prog Lipid Res* (2021) 82:101092–1100. doi: 10.1016/J.PLIPRES.2021.101092
38. Abu-Farha M, Thanaraj TA, Qaddoumi MG, Hashem A, Abubaker J, Al-Mulla F. The Role of Lipid Metabolism in COVID-19 Virus Infection and as a Drug Target. *Int J Mol Sci* (2020) 21:3544–55. doi: 10.3390/IJMS21103544
39. Müller WEG, Neufurth M, Wang S, Schröder HC, Wang X. Polyphosphate Reverses the Toxicity of the Quasi-Enzyme Bleomycin on Alveolar Endothelial Lung Cells In Vitro. *Cancers (Basel)* (2021) 13:1–20. doi: 10.3390/CANCERS13040750
40. Xu LH, Huang M, Fang SG, Liu DX. Coronavirus Infection Induces DNA Replication Stress Partly Through Interaction of its Nonstructural Protein 13 With the P125 Subunit of DNA Polymerase δ . *J Biol Chem* (2011) 286:39546–59. doi: 10.1074/JBC.M111.242206
41. Cremona CA, Behrens A. ATM Signalling and Cancer. *Oncogene* (2014) 33:3351–60. doi: 10.1038/ONC.2013.275
42. Su M, Chen Y, Qi S, Shi D, Feng L, Sun D. A Mini-Review on Cell Cycle Regulation of Coronavirus Infection. *Front Vet Sci* (2020) 7:586826. doi: 10.3389/FVETS.2020.586826
43. Hasan MZ, Islam S, Matsumoto K, Kawai T. SARS-CoV-2 Infection Initiates Interleukin-17-Enriched Transcriptional Response in Different Cells From Multiple Organs. *Sci Rep* (2021) 11:16814–25. doi: 10.1038/S41598-021-96110-3
44. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: An R Package for Measuring Semantic Similarity Among GO Terms and Gene Products. *Bioinformatics* (2010) 26:976–8. doi: 10.1093/BIOINFORMATICS/BTQ064
45. Demasi M. COVID-19 and Metabolic Syndrome: Could Diet be the Key? *BMJ Evid Based Med* (2021) 26:1–2. doi: 10.1136/BMJEBM-2020-111451
46. Pei R, Feng J, Zhang Y, Sun H, Li L, Yang X, et al. Host Metabolism Dysregulation and Cell Tropism Identification in Human Airway and Alveolar Organoids Upon SARS-CoV-2 Infection. *Protein Cell* (2021) 12:717–33. doi: 10.1007/S13238-020-00811-W
47. Overmyer KA, Shishkova E, Miller IJ, Balnis J, Bernstein MN, Peters-Clarke TM, et al. Large-Scale Multi-Omic Analysis of COVID-19 Severity. *Cell Syst* (2021) 12:23–40.e7. doi: 10.1016/J.CELLS.2020.10.003
48. Policard M, Jain S, Rego S, Dakshnamurthy S. Immune Characterization and Profiles of SARS-CoV-2 Infected Patients Reveals Potential Host Therapeutic Targets and SARS-CoV-2 Oncogenesis Mechanism. *bioRxiv* (2021). doi: 10.1101/2021.02.17.431721
49. Jyotsana N, King MR. The Impact of COVID-19 on Cancer Risk and Treatment. *Cell Mol Bioeng* (2020) 13:285–91. doi: 10.1007/S12195-020-00630-3
50. Zong Z, Wei Y, Ren J, Zhang L, Zhou F. The Intersection of COVID-19 and Cancer: Signaling Pathways and Treatment Implications. *Mol Cancer* (2021) 20:76–95. doi: 10.1186/S12943-021-01363-1
51. Deb S, Arrighi S. Potential Effects of COVID-19 on Cytochrome P450-Mediated Drug Metabolism and Disposition in Infected Patients. *Eur J Drug Metab Pharmacokinet* (2021) 46:185–203. doi: 10.1007/S13318-020-00668-8
52. Mortaz E, Tabarsi P, Varahram M, Folkerts G, Adcock IM. The Immune Response and Immunopathology of COVID-19. *Front Immunol* (2020) 11:2037. doi: 10.3389/FIMMU.2020.02037
53. Li L, Wo J, Shao J, Zhu H, Wu N, Li M, et al. SARS-Coronavirus Replicates in Mononuclear Cells of Peripheral Blood (PBMCs) From SARS Patients. *J Clin Virol* (2003) 28:239–44. doi: 10.1016/S1386-6532(03)00195-1
54. Moustafa A, Khalel RS, Aziz RK. Traces of SARS-CoV-2 RNA in Peripheral Blood Cells of Patients With COVID-19. *OMICS* (2021) 25:475–83. doi: 10.1089/OMI.2021.0068
55. Zheng S, Fan J, Yu F, Feng B, Lou B, Zou Q, et al. Viral Load Dynamics and Disease Severity in Patients Infected With SARS-CoV-2 in Zhejiang Province, China, January-March 2020: Retrospective Cohort Study. *BMJ* (2020) 369:m1443–51. doi: 10.1136/BMJ.M1443
56. Kalfaoglu B, Almeida-Santos J, Tye CA, Satou Y, Ono M. T-Cell Dysregulation in COVID-19. *Biochem Biophys Res Commun* (2021) 538:204–10. doi: 10.1016/J.BBRC.2020.10.079
57. Sadanandam A, Bopp T, Dixit S, Knapp DJHF, Emperumal CP, Vergidis P, et al. A Blood Transcriptome-Based Analysis of Disease Progression, Immune Regulation, and Symptoms in Coronavirus-Infected Patients. *Cell Death Discovery* (2020) 6:141–55. doi: 10.1038/S41420-020-00376-X
58. Zhou Y, Liao X, Song X, He M, Xiao F, Jin X, et al. Severe Adaptive Immune Suppression May Be Why Patients With Severe COVID-19 Cannot Be Discharged From the ICU Even After Negative Viral Tests. *Front Immunol* (2021) 12:755579. doi: 10.3389/FIMMU.2021.755579
59. Ding J, Lugo-Martinez J, Yuan Y, Huang AJ, Suder EL, et al. Reconstructed Signaling and Regulatory Networks Identify Potential Drugs for SARS-CoV-2 Infection. *bioRxiv* (2021). doi: 10.1101/2020.06.01.127589
60. Gupta A, Chander Chiang K. Prostaglandin D 2 as a Mediator of Lymphopenia and a Therapeutic Target in COVID-19 Disease. *Med Hypotheses* (2020) 143:10122–26. doi: 10.1016/J.MEHY.2020.110122

61. Maekawa T, Ashihara E, Kimura S. The Bcr-Abl Tyrosine Kinase Inhibitor Imatinib and Promising New Agents Against Philadelphia Chromosome-Positive Leukemias. *Int J Clin Oncol* (2007) 12:327–40. doi: 10.1007/S10147-007-0699-1
62. Bath PMW, Blecic S, Bogousslavsky J, Boysen G, Davis S, Diez-Tejedor E, et al. Tirilazad Mesylate in Acute Ischemic Stroke: A Systematic Review. *Tirilazad Int Steering Committee Stroke* (2000) 31:2257–65. doi: 10.1161/01.STR.31.9.2257
63. Cagno V, Magliocco G, Tapparel C, Daali Y. The Tyrosine Kinase Inhibitor Nilotinib Inhibits SARS-CoV-2 *In Vitro*. *Basic Clin Pharmacol Toxicol* (2021) 128:621–4. doi: 10.1111/BCPT.13537
64. Chandel V, Sharma PP, Raj S, Choudhari R, Rathi B, Kumar D. Structure-Based Drug Repurposing for Targeting Nsp9 Replicase and Spike Proteins of Severe Acute Respiratory Syndrome Coronavirus 2. *J Biomol Struct Dyn* (2022) 40:249–62. doi: 10.1080/07391102.2020.1811773
65. da Silva Santos PS, da Fonseca Orcina B, Machado RRG, Vilhena FV, da Costa Alves LM, Zangrando MSR, et al. Beneficial Effects of a Mouthwash Containing an Antiviral Phthalocyanine Derivative on the Length of Hospital Stay for COVID-19: Randomised Trial. *Sci Rep* (2021) 11:19937–47. doi: 10.1038/S41598-021-99013-5
66. Bohan D, van Ert H, Ruggio N, Rogers KJ, Badreddine M, Briseno JAA, et al. Phosphatidylserine Receptors Enhance SARS-CoV-2 Infection. *PLoS Pathog* (2021) 17:e1009743–75. doi: 10.1371/JOURNAL.PPAT.1009743
67. Scaltriti M, Elkabets M, Baselga J. Molecular Pathways: AXL, a Membrane Receptor Mediator of Resistance to Therapy. *Clin Cancer Res* (2016) 22:1313–7. doi: 10.1158/1078-0432.CCR-15-1458
68. Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD, et al. GeneMANIA Update 2018. *Nucleic Acids Res* (2018) 46(W1):W60–4. doi: 10.1093/NAR/GKY311
69. Thair SA, He YD, Hasin-Brumshtein Y, Sakaram S, Pandya R, Toh J, et al. Transcriptomic Similarities and Differences in Host Response Between SARS-CoV-2 and Other Viral Infections. *iScience* (2020) 24:101947–84. doi: 10.1016/J.ISCI.2020.101947
70. Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering Regulatory and Signalling Circuits in Molecular Interaction Networks. *Bioinformatics* (2002) 18 Suppl 1:S233–40. doi: 10.1093/BIOINFORMATICS/18.SUPPL_1.S233
71. Mitra K, Carvunis AR, Ramesh SK, Ideker T. Integrative Approaches for Finding Modular Structure in Biological Networks. *Nat Rev Genet* (2013) 14:719–32. doi: 10.1038/NRG3552
72. Frieman M, Coleman C, Daugherty SC, Rasko D, Shefchek K, Sengamalay N, et al. *GEO Accession Viewer* (2019). Available at: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56192> (Accessed March 1, 2022).
73. *Sequence Read Archive*. NCBI/NLM/NIH. Available at: <https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?cmd=show&f=software&m=software&s=software> (Accessed March 1, 2022).
74. GitHub. *FelixKrueger/TrimGalore: A Wrapper Around Cutadapt and FastQC to Consistently Apply Adapter and Quality Trimming to FastQ Files, With Extra Functionality for RRBS Data*. Available at: <https://github.com/FelixKrueger/TrimGalore> (Accessed March 1, 2022).
75. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression. *Nat Methods* (2017) 14:417–9. doi: 10.1038/NMETH.4197
76. Zpyrych-Walczak J, Szabelska A, Handschuh L, Górczak K, Klamecka K, Figlerowicz M, et al. The Impact of Normalization Methods on RNA-Seq Data Analysis. *BioMed Res Int* (2015) 2015:621690–700. doi: 10.1155/2015/621690
77. *Loess Function - RDocumentation*. Available at: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/loess> (Accessed March 1, 2022).
78. Kuznetsova I, Lugmayr A, Siira SJ, Rackham O, Filipovska A. CirGO: An Alternative Circular Way of Visualising Gene Ontology Terms. *BMC Bioinf* (2019) 20:84–91. doi: 10.1186/S12859-019-2671-2
79. *Create Elegant Data Visualisations Using the Grammar of Graphics • Ggplot2*. Available at: <https://ggplot2.tidyverse.org/> (Accessed March 1, 2022).
80. Khan A, Mathelier A. Intervene: A Tool for Intersection and Visualization of Multiple Gene or Genomic Region Sets. *BMC Bioinf* (2017) 18:287–95. doi: 10.1186/S12859-017-1708-7
81. CRAN. *Package Pheatmap*. Available at: <https://cran.r-project.org/web/packages/pheatmap/index.html> (Accessed March 1, 2022).
82. Trott O, Olson AJ. AutoDock Vina: Improving the Speed and Accuracy of Docking With a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* (2010) 31:455–61. doi: 10.1002/JCC.21334
83. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res* (2000) 28:235–42. doi: 10.1093/NAR/28.1.235
84. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An Open Chemical Toolbox. *J Cheminform* (2011) 3:33–47. doi: 10.1186/1758-2946-3-33
85. Krivák R, Hoksza D. P2Rank: Machine Learning Based Tool for Rapid and Accurate Prediction of Ligand Binding Sites From Protein Structure. *J Cheminform* (2018) 10:39–51. doi: 10.1186/S13321-018-0285-8
86. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: A Free Tool to Discover Chemistry for Biology. *J Chem Inf Model* (2012) 52:1757–68. doi: 10.1021/ci3001277
87. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J Comput Chem* (2004) 25:1605–12. doi: 10.1002/JCC.20084

Author Disclaimer: The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Agrawal, Sambaturu, Olgun and Hannehalli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.