Lung Disease Network Reveals the Impact of Comorbidity on SARS-CoV-2 Infection and Opportunities of Drug Repurposing

ASIM BIKAS DAS (asimbikas@nitw.ac.in)
National Institute of Technology Warangal

Research article

Keywords: COVID-19, SARS-CoV-2, lung disease, comorbidity, disease network

Posted Date: December 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-121999/v1

Lung Disease Network Reveals the Impact of Comorbidity on SARS-CoV-2 Infection and Opportunities of Drug Repurposing

Asim Bikas Das*

Department of Biotechnology, National Institute of Technology Warangal, Warangal 506004, Telangana, India.

*** Correspondence:**
asimbikas@nitw.ac.in

Tel No: +91-8106311048/+91-8332969440

Keywords: COVID-19, SARS-CoV-2, lung disease, comorbidity, disease network

Abstract

Higher mortality of COVID-19 patients with comorbidity is the formidable challenge faced by the health care system. In response to the present crisis, understanding the molecular basis of comorbidity is essential to accelerate the development of drugs. To address this, the genetic association between COVID-19 and various lung disorders was measured and notable molecular resemblance was observed. 141 lung diseases were linked to a neighborhood network of SARS-CoV-2 targets, and 59 lung diseases topologically overlapped with COVID-19 module. This demonstrates the clustering of lung diseases with COVID-19 in the same network vicinity, indicating the potential threat for lung patients upon SARS-CoV-2 infection. Pathobiological similarities between lung diseases and COVID-19, and clinical evidences suggest that shared molecular features probably the reason for comorbidity. Additionally, topological overlap with various lung disorders provides an opportunity to repurpose the drugs used for lung disease to hit the closely associated COVID-19 module. Further analysis showed that the functional protein-protein interaction modules in the lungs, substantially hijacked by SARS-CoV-2, were connected to several lung disorders. The network-based proximity measure identified the FDA approved targets in hijacked protein modules which can be hit by existing drugs to rescue these modules from viral possessions, and can lead to the improvement of clinical conditions.

Introduction

The novel Coronavirus Disease 2019 (COVID-19) cases, caused by SARS-CoV-2 infection, crossed 37,000,000 all over the world as of October 12, 2020. Recent data show that the most affected groups are with two or more pre-existing medical conditions such as hypertension, diabetes, metabolic, cardiovascular, and digestive disorders [1-3]. Moreover, comorbidity (or existence of multiple disorders) causes a higher risk of developing a severe illness, poor prognosis, and higher mortality of COVID-19 patients [4]. A virus causes the disease by hijacking the host cell machinery for its replication. Interactions of the virus with the host perturb the highly organized host cellular networks and re-construct different networks that are favorable for virus replication. The topology of molecular interactions is altered in disease; hence SARS-CoV-2 interaction with healthy human cells will be different from the disease cell, and this could lead to various impacts on humans upon SARS-CoV-2 infection. Human diseases are connected via defects in common genes [5, 6]. Moreover, the similarity in disease phenotypes often indicates underlying genetic connections. Therefore, pre-existing medical conditions can facilitate the appearance of another disease if they share the same or functionally related genes [7, 8]. SARS-CoV-2 has been associated with respiratory tract infection (RTI), and in some cases, it severely damages adult lungs. Here, we demonstrated the underlying molecular link between COVID-19 and lung diseases to understand the basis of comorbidity. In the present work, we have considered a disease in the lung or symptoms in lung or diseases in other tissues or organs affecting lungs as "lung disease". Recent efforts by Gordon et al.[9] identified 26 of the 29 SARS-CoV-2 proteins, which bind to 332 human proteins and hijack the host translational machinery. Here, we constructed a tissue (lungs)-specific neighborhood network of 332 human targets of SARS-CoV-2. The neighborhood network was integrated with lung diseases to build a disease-gene network of the lung. Subsequently, we constructed a lung disease network, which also includes COVID-19. We observed that 141 lung diseases were associated with COVID-19. 49 out of 141 disorders were directly linked to COVID-19, apparently justifying the characteristics of a complex disorder. Further, we observed topological overlaps between 59 lung diseases and COVID-19, indicating a higher chance of comorbidity. This also shows the opportunity to repurpose drugs which are used to treat lung diseases because these drugs can simultaneously hit a lung disease and closely associated COVID-19 module. We also observed that genes in overlapping lung diseases and COVID-19 are coexpressed and involved in similar molecular function and biological processes

compared to the random expectation, representing pathobiological similarities between various lung disorders and COVID-19.

Next, we identified functional protein modules that are maximally perturbed by SARS-CoV-2 and involved in RNA processing, export, and protein synthesis machinery of the cell. Moreover, these protein modules are associated with various lung disorders, indicating the hotspots for comorbidity. Hence, we employed a network-based proximity approach [10] and explored the DrugBank database [11] to find FDA-approved targets to hit these functional protein modules and rescue them from virus possession. The effective use of network-based toolset to identify drugs for COVID-19 treatment has been reported in recent studies [12, 13]. We identified 56 druggable human proteins that are in proximity to the COVID-19 disease module and could be targeted by FDA approved and investigational drugs. The rate of mutation of SARS-CoV-2 is very high, which enables the virus to develop drug resistance [14]. Therefore, identification and targeting the host factors will be an enduring approach instead of targeting viral proteins. In summary, this work presents the risk of different lung disorders at the onset of COVID-19 and drug repurposing opportunities to treat patients with lung disorders.

Results

Construction of SARS-CoV-2 –host interactome in the lung

To depict the SARS-CoV-2 –host interaction network, the protein-protein interaction (PPI) network of the lungs (lung interactome) was obtained from the TissueNet v.2 database [15]. We collected a list of 332 human targets of SARS-CoV-2 from Gordon et al.[9] article and constructed a subnetwork of these 332 proteins from the PPI network of lungs. Out of the 332 viral targets, 323 proteins were present in the subnetwork. The resulting subnetwork, named as SARS-CoV-2 target network (STN), consists of 5050 nodes and 11256 pairwise interactions (Fig.1 a, supplementary table1). Next, we observed that 181 out of 323 viral targets form the largest connected component (LCC) within lung interactome. To measure the statistical significance of LCC, we randomly selected proteins with matching degree and calculated the size of LCC. We repeated the random selection 1000 times and found that the size of random LCC was 136.28±16.05 (Fig.1 b), and z-score = 2.78 (p-value = 5.36×10^{-3}), indicating SARS-CoV-2 –host interaction network did not appear by chance and the target

proteins were located in the same network vicinity [10, 12]. To confirm this, we again computed the dyadicity (D) (a measure of the connectedness of the nodes with the same label, see Methods) among the SARS-CoV-2 targets in the STN to know if they share more or fewer edges than expected in a random configuration of the network. We found D=7.664, indicating high connectedness among SARS-CoV-2 targets. D>1 signifies that SARS-CoV-2 targets form a community like structure to hijack the host cellular machinery. Proteins in a community, if implicated in diseases, then they can exhibit a higher chance of comorbidity than those who are not in the community. This is because proteins in a community frequently interact, coexpress, and are functionally interconnected [16]. Therefore, to understand the link between COVID-19 and other lung diseases, we constructed and analyzed the disease-gene and disease –disease association map linked to STN.

Disease-gene and disease –disease association map of COVID-19 in lungs

To construct a disease association map of STN, we obtained the disease-gene association data from the ORGANizer database [17]. 184 lung diseases, 1957 genes, and 6039 disease-gene pairs were considered for further analysis (see Methods) (Supplementary Table 2). However, we observed 1442 out of 1957 gene are present in lung interactome. To construct the disease-gene association map, we screened the diseases that are associated with proteins (nodes) in STN. A disease and node are then connected if the node is associated with the disorder in the lungs. We observed, 618 proteins, consisting of 36 SARS-CoV-2 targets, were linked to a total of 146 disorders, which includes COVID-19 (Supplementary Table 3). It was observed that the overlap between SARS-CoV2 targets and 1442 genes associated with lung diseases was not statistically significant (Fisher's exact test, p-value= 0.454). A similar observation was reported by Gysi et al. [12] with a group of genes involved in various disease classes. However, the overlap between 5050 nodes in STN and genes in lung diseases was statistically significant (Fisher's exact test, p-value= 2.93×10^{-5}). Fig.2a shows the resulting disease-gene association map of STN, named as the lung disease-gene network (LDGN), consisting of 1814 disease-gene pairs.

The largest connected component within the LDGN consists of 141 lung diseases and 610 genes, indicating that many of the disorders share the common genotype. For example, SARS-CoV-2 targets, FBN1 (degree, k=15), FBLN5 (k =11), COMT (k =9), and neighborhood nodes, OFD1 (k =19), DNAAF2 (k =16) and DNAAF5 (k =16) are linked to multiple disorders (Fig. 2b). Similarly, a disorder in LDGN is also connected with multiple

genes. For instance, ventricular septal defect (k =142), respiratory insufficiency (k =133), congestive heart failure (k =95), apnea (k =63) and hypothyroidism (k =60) (Fig. 2c, Supplementary Fig. 1 and Supplementary Fig. 2).

The disease-gene association pattern in LDGN indicates the molecular connection of COVID-19 with a wide range of lung disorders. To comprehend the association between COVID-19 and lung diseases, a disease-disease association network (DDAN) was constructed, where two diseases were linked if they shared one associated gene (Fig. 3a). DDAN consists of a total of 141 diseases (nodes) and 1326 links, indicating a higher clustering between diseases. Further it was observed that the degree distribution of DDAN did not follow the scale-free property (Figure 3b). To find the exact topological nature, we measured network transitivity (T_{DDAN}=0.4264) and average path length (L_{DDAN}=*2.0585*) of DDAN. These topological parameters were compared with the equivalent 1000 Erdős−Rényi random graphs. The results show that the average path length of DDAN is significantly less (p-value<0.0001), whereas transitivity is significantly higher (p-value<0.0001) compared to random graphs ($L_{random} = 2.44680$ and $T_{random} = 0.0668$*)* (Figure3c and d). Further, we calculated the small-worldness scalar (*S*) for DDAN as follows:

$$\gamma = \frac{T_{DDAN}}{T_{random}} = 6.383$$

$$\lambda = \frac{L_{DDAN}}{L_{random}} = 0.841$$

$$S = \frac{\gamma}{\lambda} = 7.589$$

A network is said to be a small-world network if *S*>1[18]. Hence, the topology of DDAN represents a small-world property, indicating that any two diseases in DDAN have a high tendency to be interconnected and may cause the overlapping disease pathogenesis.

We observed that 49 diseases in DDAN were directly connected to COVID-19 (Supplementary Fig.3). The Jaccard similarity coefficient was computed based on the number of common genes to identify the extent of molecular overlap between 49 lung diseases and COVID-19. There are several diseases, like respiratory insufficiency, congestive heart failure, respiratory failure, ventricular septal defect, mitral regurgitation, and hyperthyroidism, which are closely associated with COVID-19. It is important to note that although there are molecular connections between COVID-19 and various lung diseases, these overlaps are not

statistically significant (considering only SARS-CoV-2 targets). These molecular connections cannot be ignored to analyze the effect on lung patients upon SARS-CoV-2 infection, however, it provides limited opportunities to comprehend the disease comorbidity.

Topological overlap between disease modules, pathobiological similarities and opportunities for drug repurposing

To obtain a greater understanding of comorbidity, we further measured the network-based separation between two disease modules to understand their degree of overlap. The main advantage of network-based separation measure is that it can predict disease-disease association, even if two diseases do not share genes. If two disease modules overlap with each other, then perturbations to one disease can cause disturbance to another, resulting in similar clinical characteristics [19]. Network-based separation (S_{ab}) (see Methods) between COVID-19 and all lung diseases was measured. 59 out of 184 lung diseases demonstrated overlapping modules ($S_{ab} < 0$) with COVID-19 (Supplementary Table4). The statistical significance of S_{ab} for each disease pair, i.e COVID-19 and each lung disease was evaluated using full randomization model (see Methods). We observed, for all 59 diseases, the z-score was < 0, indicating that these diseases are closely overlapped with COVID-19 than expected by chance. Fig. 4(a-j) shows the top 10 closely overlapping lung disease modules with COVID-19, which are Hemolytic-uremic syndrome ($S_{ab} = -0.2142$), abnormal respiratory motile cilium morphology ($S_{ab} = -0.21138$), obstructive lung disease ($S_{ab} = -0.21022$), pleural effusion ($S_{ab} = -0.18216$), patent foramen ovale ($S_{ab} = -0.1619$), and pulmonary insufficiency ($S_{ab} = -0.15694$). Thus, patients with these disorders are probably more vulnerable to COVID-19 symptoms or vice versa because of overlapping disease modules. Note that abnormal respiratory motile cilium morphology, absent respiratory ciliary axoneme radial spokes, respiratory insufficiency due to defective ciliary clearance are caused by the same set of genes; therefore, we considered only abnormal respiratory motile cilium morphology (ARMCM) in the top 10 list. According to the network-based separation measure, almost 32% of lung diseases have overlapping modules with COVID-19, and 68% of lung diseases are topologically separated. When a distinct set of genes (or proteins) contribute to two or more disease phenotypes that indicate those genes probably coexpress and share molecular functions and biological processes. We observed that the coexpression, molecular functions and biological processes similarity of genes involved in COVID-19 and overlapping lung disease ($S_{ab} < 0$) were significantly (p-

value<0.0001) higher compared to the random control (Fig. 4 k-l). These results indicate pathobiological similarities between COVID-19 and overlapping lung diseases. To further investigate the similarities in clinical features, the results from recent publications were explored. It is important to note that recent reports have raised concerns about lung injuries linked to COVID-19 [20, 21]. Data show a higher percentage of COVID-19 patients in severe conditions were more likely to have the chronic obstructive pulmonary disease (COPD) and impairment of diffusion capacity [4, 22]. Many of lung diseases ($S_{ab} < 0$) (Supplementary Table4) with overlapping module with COVID-19 are linked to these above phenotypes. A closely associated disease with COVID-19, Hemolytic-uremic syndrome ($S_{ab} < 0$) (Fig. 4a) causes the pulmonary haemorrhage which is linked to kidney failure [23] and recent studies show both chronic kidney diseases and chronic pulmonary disease cause the adverse outcomes in patients with COVID-19 [4, 24]. Abnormal respiratory motile cilium (Fig. 4b) or ciliary dyskinesia ($S_{ab} = -0.21138$) causes chronic respiratory tract infections because the improper movement of mucus unable to eliminate fluid, bacteria, and particles from the lungs, which can results in bronchitis (Chronic bronchitis, $S_{ab} = -0.146$) (www.ghr.nlm.nih.gov). The patient with ciliary dyskinesia could be at higher risk of health hazard upon SARS-CoV-2 infection because of a lack of respiratory clearance. Further study shows patient with obstructive lung disease ($S_{ab} = -0.21$) (Fig. 4c) and pulmonary emphysema ($S_{ab} = -0.09$) is at higher risk of pneumothorax when infected with SARS-CoV-2 [25]. Rajendram et al. [26] have predicted that patent foramen ovale (PFO) (Fig. 4e) may be relevant to many patients with COVID-19 because PFO causes pulmonary embolism [27]. Even the pulmonary embolism disease module was overlapped ($S_{ab} = -0.008$) with COVID-19. A clinical study in Wuhan, China [28] reported that almost 5% COVID-19 patients were detected with pleural effusion ($S_{ab} = -0.18$) (Fig. 4d) which is often caused by congestive heart failure and blood clots in lung arteries. Importantly pleural effusion is commonly associated with age-related respiratory problem and also linked cancer [29]. On the other hand, congestive heart failure which causes many lung related disease [30], was also overlapped with COVID-19 disease module (Supplementary Table 4). Moreover meta-analysis by Alqahtani et al. [31] showed that the higher risk of more severity was linked to COPD patients (risk of severity = 63%) compared to patients without COPD (33.4%). These results suggest of the clinical similarities between COVID-19 and overlapping lung disorders. However, these observations are limited, cannot be extrapolated for all overlapping lung disorders without clinical evidences. Besides, a genome-wide association study has shown

the genetic susceptibility locus in chromosome of patients with COVID-19 and respiratory failure [32], and we observed, genes present in this locus (SLC6A20, LZTFL1, CCR9) are also associated with different lung disorders ($S_{ab} < 0$) such as, pulmonary fibrosis, respiratory distress, asthma, nephrotic syndrome.

Clinically characterized lung disease having efficient drugs for treatment and overlapping neighborhoods with COVID-19 have shown the scope for implementing the existing drugs (repurposing) for COVID-19 treatment. When two diseases are localized in the same network vicinity and overlap with each other, then targeting one disease can affect another disease module (Fig. 4n) results in efficient clinical outcomes for both as they have common network neighborhoods [33]. Clinical data from Clinicaltrials.gov database show that some drugs used for lung diseases such as methylprednisolone for tracheal stenosis ($S_{AB} = -0.14915$) and ketamine, budesonide for COPD are in clinical trials for COVID-19 treatment [34]. Therefore, we suggest testing exiting drugs that are presently in use for treating lung disorders on COVID-19 patients for better clinical outcomes. Treating a comorbid patient is challenging, but an accurate clinical picture of patients, the molecular signature of diseases, and drug-target information can improve the present crisis.

Functional protein modules preferentially hijacked by SARS-CoV-2 are linked to a broad range of lung disorders.

Modularity in the network refers to the pattern of connectedness in which nodes are grouped into highly connected subsets [35]. One of the key features in the protein interaction network is that the tightly connected proteins within a community are mostly involved in similar biological functions [36]. Similarly, genes involved in related diseases are shown to be highly connected; moreover, diseases linked to common genes result in the formation of disease modules and comorbidity [37]. We have compared various community detection algorithms, i.e., fast-greedy, walktrap, louvain, leading eigenvector, and spinglass, to identify protein modules in STN [38, 39]. Spinglass showed good partitioning with a higher modularity score compared to other algorithms (see Methods and Supplementary Table 5). Our findings are in agreement with previous studies by Rahiminejad et al. [40], where authors observed good partitioning of the functional protein module using spinglass in eukaryotes. Out of 21 modules, the top four protein modules were selected based on the presence of a large number of SARS-CoV-2 targets (>20) and gene ontology semantic similarity score (>0.2) of biological processes (Supplementary Table 6). A large number of viral targets were considered because these modules are largely hijacked and strongly perturbed upon infection

compared to other functional modules in the network. The modules were named as modules 1, 2, 3, and 4, and each module contains 63, 50, 28, and 23 SARS-CoV-2 target proteins, respectively (Fig. 5). The biological process and pathway enrichment analysis showed that module1, the largest module, is mostly enriched with RNA metabolism, including transcription, mRNA processing, transport, mRNA deadenylation, and surveillance. Presumably, biological processes linked to module1 are hijacked by SARS-CoV-2 in the early stage of infection for the production of its RNA. Notably, the components of module1 are linked to 64 disorders, among which the highly connected are respiratory insufficiency, ventricular septal defect, respiratory distress, pneumonia, and neoplasm of the lung (Fig. 5a, 3rd column, supplementary table 7). It is worth noting that most of the diseases associated with module1are directly connected to COVID-19 (Fig.3b and Fig.S2). On the other hand, hijacking module2 can predominantly affect protein degradation (ERAD pathway, HRD1 complex, regulation of protein catabolic process), transport, folding, and stability (retrograde protein transport, regulation of protein stability, VCP-VIMP-DERL1-DERL2-HRD1-SEL1L complex, regulation of intracellular transport, regulation of vesicle-mediated transport, and protein folding in the endoplasmic reticulum). Module3 and module4 involve several processes, primarily cellular transport, localization, organization, and cell cycle. Modules 2, 3, and 4 were linked to a total of 79, 60, and 32 different disorders, respectively (Supplementary Table 6). The disease association of all four protein modules was significantly higher (p-value<0.0001) than 1000 random gene sets. Besides, we found a broad spectrum of disorders of various classes, such as neoplasms, neurological, and digestive system, were associated with these modules (Supplementary Fig. S4). Gysi et al.[12] predicted that manifestation of SARS-CoV-2 in different human tissues could cause various disorders. Therefore, not only lung-related disorders but comorbidity in other organs can also be a potential threat for COVID-19 patients. To strengthen this observation, the pattern of coexpression of genes in functional modules was analyzed. Genes in the same functional module often show a high coexpression profile; therefore, we calculated Pearson correlation coefficients of pairs of genes using gene expression data of healthy lung tissue from TCGA. The median value of the positive correlation between the genes in all modules was significantly higher (p-value< 0.0001) compared with the random gene set (Fig. 5, fourth column). Therefore, these modules can be identified as coexpression modules that share core transcriptional programs in the lung, indicating that their perturbation can result in a similar disease phenotype. Next, we proceeded to find the targets to hit functional modules by drug repurposing.

Drug repurposing to target the functional modules

We propose targeting functional protein modules, hijacked by SARS-CoV-2, by drug repositioning. There are two main reasons to target these modules. First, the binding of a drug to its target in a module will prevent the replication of the virus. Second, as a module is linked to several lung diseases, targeting a module can improve the severity of comorbidity. We identified 56 approved targets (red color nodes in Fig.6), which can be hit by 144 approved or investigational drugs in the clinical trial, from DrugBank [11] (Supplementary Table 8). The list contains 10 approved drugs, which are at different stages of clinical trials for COVID-19 treatment, including chloroquine targeting Glutathione S-transferase Mu 1(GSTM1) in module3 (Supplementary Fig. S4). However, the efficacy of chloroquine on COVID-19 patients is arguable. We observed the presence of coagulation factor X (F10) in module2, which has been recently implicated as a target due to the potential role of coagulopathy in COVID-19 [41]. To find the effectiveness of targets, we applied network-based proximity measures to calculate the proximity between COVID-19 disease module (network among SARS-CoV-2 targets) and FDA approved targets in the functional module. We used the "closest" (d_c) measure, which represents the closest path length between a target and nearest SARS-CoV-2 target protein. Then we calculated the z-score (z_c) to validate the proximity by comparing the observed target-disease protein distance to the random expectation [10]. We found that all 56 targets were proximal ($z_c < -2$) to the COVID-19 disease module (Supplementary Table9) compared to the random expectation. Next, we found that approved drugs that are in clinical trials for COVID-19 treatment (Supplementary Fig. S5) were also significantly closer to COVID-19 ($z_c < -2$) (Supplementary Table10). Considering the complexity of COVID-19, we suggest aiming multiple locations in the STN. Targeting the different location may help to efficiently rewire the cellular network [42] and can rescue these functional modules from the virus. Drugs such as regorafenib ($z_c = -3.06$), tamoxifen ($z_c = -2.96$) and afanitib ($z_c = -2.53$) have more than one target in different modules (Supplementary Table11, Fig. S4); hence, they can simultaneously intervene in RNA and protein metabolism of the host hijacked by the virus [33]. Targeting these modules may efficiently hinder viral possession and can reduce the growth of the virus. Many of the target proteins suggested do not directly interact with SARS-CoV-2; rather, they are neighborhood nodes, as they are present in the same network vicinity, binding of drugs to these targets may efficiently perturb the network modules as well as viral growth [43].

Importantly, the drugs targeting the functional modules should be tested on SARS-CoV-2 infected cell lines and validated through clinical trials.

Discussion

Currently, there is an urgent need for a speedy drug discovery or vaccine development to stop the infection and rapid transmission of SARS-CoV-2. Most alarming is that the aged COVID-19 patients with comorbidity are at severe health risks worldwide. The present study has shown the risk of SARS-CoV-2 infection on the onset of various lungs related disorders and molecular basis of comorbidity by applying the principle of network biology. COVID-19 appears to be a complex disease because of wide-ranging SARS-CoV-2 targets in the host cell, establishing the molecular connection with various lung-related disorders. The disease-gene, disease-disease association map, and network separation analysis have shown molecular links and clustering of diseases in the same network vicinity, indicating a pathobiological similarity of COVID-19 with various lung disorders. Some of the closely associated diseases with COVID-19 are Hemolytic-uremic syndrome, obstructive lung disease, pleural effusion, and chronic bronchitis. Due to the close association, the pre-existence of these diseases can be the cause of higher mortality of COVID-19 patients. One of the common respiratory problems, asthma, which has overlapping disease module and also directly connected to COVID-19, shows moderate to higher severity upon SARS-CoV-2 infection (www.cdc.gov). These observations will provide a detailed picture of the molecular basis of the severe illness of COVID-19 patients with specific lung disorders and also help us to decipher the patient-specific etiology of COVID-19. Because of multiple molecular connections and overlapping disease modules of COVID-19 with various lung disorders, it is challenging to find specific targets and potential drugs for patients with pre-existing medical conditions. The present crisis cannot wait for long for new drugs to come; therefore, we propose two approaches for drug repositioning. The first approach is the testing of approved drugs developed for lung diseases that have overlapping disease modules with COVID-19. These drugs can simultaneously affect two disease modules that can lead to better clinical conditions. In the second approach, we suggest targeting the host functional protein modules, which are the origin of many lung disorders and are primarily hijacked by SARS-CoV-2. Perturbing these modules by repurposing FDA-approved (or investigational) drugs may rescue the host cellular machinery utilized by the virus for its replication.

Realizing the complexity of SARS-CoV-2 infection, we suggest hitting the multiple targets in the different functional modules to improve clinical outcomes. However, systematic studies to identify drug combinations and their targets are highly recommended to increase clinical efficacy and lower toxicity [33]. Besides, patient-specific high-throughput transcriptomics data, in vitro, or in vivo assays of drug combination and study of pharmacokinetic are essential to establish a proper treatment strategy.

Materials and Methods

Construction of a lung-specific PPI network of SARS-CoV-2 targets

Human lung tissue-specific interactome data were retrieved from the TissueNet v.2 database. To generate tissue-specific PPIs, TissueNet v.2 synergizes between large-scale data of human PPIs and tissue-specific expression profiles. PPIs from four major PPI databases, BioGrid, IntAct, MINT, and DIP, were obtained and consolidated. Then it integrated resulting PPIs with RNA-sequencing profiles of the Genotype-Tissue Expression consortium (GTEx). We downloaded 168296 lung-specific interactions from TissueNet v.2 to construct SARS-CoV-2 targets interactome. Next, we obtained a list of 332 human proteins targeted by SARS-CoV-2 [9] and built a subnetwork called the SARS-CoV-2 target network (STN). The 9 SARS-CoV-2 targets (AATF, CEP43, CISD3, MTARC1, NUP62, SRP19, THTPA, TIMM10B, and TRIM59) did not have any interaction in the lung.

Construction of a lung-specific disease-gene and disease-disease network

The disease-gene association data in the lungs or effecting lungs were retrieved from the Gene ORGANizer (geneorganizer.huji.ac.il) [17]. The Gene ORGANizer is a phenotype-based curated database that links human genes to the body parts they affect. Phenotypes that were classified by Human Phenotype Ontology (HPO) were considered with certain modifications. Disease-gene pairs that were not included but matched with the HPO phenotype were manually included. Aspirin-induced asthma and asthma were considered as asthma. Pulmonary emphysema, sarcoidosis, and silicosis, and their associated genes were also added to the list. Finally, 6040 disease-gene pairs, which include a total of 184 various lung diseases, were mapped to STN. 618 out of 5050 nodes of STN were linked to 145 lung diseases, and 36 out of 618 genes were the direct targets of SARS-CoV-2. These 36 genes

were connected to COVID-19 as a new diseases-genes pair. Finally, a lung disease-gene association network, consisting of 1815 disease-gene pairs, including COVID-19, was constructed. The disease-disease association network was derived from the lung disease-gene association network; two diseases were connected if they shared one common gene. disgenet2r package [44] was used to study the association of disease classes with functional protein modules.

Network-based separation measure between diseases

To identify the overlapping disease modules, a "separation" measure, S_{ab} was calculated between COVID-19 (a) and lung disease (b) using the following formula:

$$\boldsymbol{S_{ab}} = < \boldsymbol{d_{ab}} > - \frac{< \boldsymbol{d_{aa}} > - < \boldsymbol{d_{bb}} >}{\mathbf{2}}$$

$\boldsymbol{S_{ab}}$ compares the shortest distances between proteins connected to each disease, $< \boldsymbol{d_{aa}} >$ and $< \boldsymbol{d_{bb}} >$, to the shortest distances $< \boldsymbol{d_{ab}} >$between a-b protein pairs. Positive $\boldsymbol{S_{ab}}$ shows that the two disease modules are separated on the lung interactome, whereas negative values indicate overlapping modules. The statistical significance of module overlap between COVID-19 and lung disease was evaluated using the full randomization model. The same number of protein associated with two diseases was randomly sampled for 1000 times and corresponding S_{ab}^{ran} between the two gene set was calculated. Next, z-score was calculated as follows

$$\text{z-score} = \frac{S_{ab} - m}{\sigma}$$

where m and σ indicates the mean value and standard deviation of 1000 S_{ab}^{ran}. Here, z-score < 0 indicates that the two diseases are closely overlapped than expected by chance [19].

Community detection

We applied fast-greedy, walktrap, louvain, leading eigenvector, and spinglass on STN as an undirected, unweighted network. These community detection algorithms segregate the nodes into higher-density modules. Each of these algorithms optimizes an objective function, i.e, modularity. Communities separated by spinglass were selected for subsequent analysis based on the modularity score and community size. Spinglass uses a random number generator to find the communities. Therefore we ran Spinglass 10 times with different seed values. We

compared the rand statistics between each run, and results showed that the community structures are highly similar (>0.7) among each other [39, 40].

Network-based proximity measure

The network proximity between drug targets (A) and SARS-CoV-2 targets in host (B) was measured by closest method (d_c).

$$d_c = \frac{1}{\|A\| + \|B\|}\left(\sum_{a\in A} min_{b\in B}\ d(a,b) + \sum_{b\in B} min_{b\in A} d(a,b)\right)$$

$d(a, b)$ represents the shortest distance between genes a and b in the lung interactome. The statistical significance of proximity was evaluated by z-score (z_c). The z_c was calculated by comparing the observed distance to a reference distance distribution. To compute reference distance distribution, the sets of proteins of similar size and degree as the drug targets and disease proteins were randomly selected for 1000 time from lung interactome. The mean and standard deviation of distance distribution was calculated to compute the z_c[10, 33].

Process and pathway enrichment analysis and gene ontology (GO) semantic similarity

Pathway and process enrichment analysis were performed using the Metascape [45]. GO Biological Processes, KEGG Pathway, and Reactome were used as ontology sources. GO semantic similarity between genes was measured by Wang *et al.*[46] method using GOSemSim package in R. Considering two genes G1 and G2 annotated by GO term sets GO1 = [go11, go12, …, go1m] and GO2 = [go21, go22, …, go2n] respectively their semantic similarity score of Wang's method is defined as

$$\mathrm{Sim}(\mathrm{G1}, \mathrm{G2}) = \frac{\sum_{1\le i\le m} \mathrm{Sim}\left(go_{1i,} GO_2\right) + \sum_{1\le j\le n} \mathrm{Sim}\left(go_{2j,} GO_1\right)}{m+n}$$

Correlation analysis

TCGA gene expression datasets of healthy human lung tissues were downloaded from the UCSC Xena project (https://xenabrowser.net/datapages/)[47]. $\log_2$(RPKM +1) (RPKM: Reads Per Kilobase Million) transformed data of adjacent healthy tissue of 59 lung

adenocarcinoma patients were retrieved, and Pearson correlation coefficient was computed to measure the coexpression levels using the Hmisc package in R.

Computation of topological parameters

Largest connected component (LCC), dyadicity, and Jaccard similarity coefficient were measured using igraph package in R. Dyadicity (D) measures the number of the same label edges divided by the expected number of same label edges, and D> 1 indicates higher connectedness between the nodes with the same label. The Jaccard similarity coefficient of two nodes is the number of common neighbors divided by the number of nodes that are neighbors of at least one of the two nodes being considered.

Tools for data analysis, plotting and statistical test

R packages tidyverse and stringr were used for data analysis, and plotting of graphs was done using ggplot2. Networks were visualized using Gephi. The statistical significance between the groups was analyzed using a non-parametric Mann-Whitney test in R. The statistical significance of the overlap between gene lists was analyzed using Fisher's exact test.

Data Availability

The datasets generated after analysis are available from the corresponding author on reasonable request.

Acknowledgments

I thank Dr. Urmila Saxena (National Institute of Technology Warangal, Telangana) for critically reading the manuscript and her constructive comments and Dr. Subir Bhattacharjee (Purulia Government Medical College & Hospital, West Bengal) for discussions on lung disease. I also thank the National Institute of Technology Warangal for providing the facilities.

Author Contributions

ABD conceived and designed the study, performed the experiment, and wrote the manuscript.

Conflict of interest

The author declares he has no competing interests

Ethical approval

This article does not contain any studies with human participants or animals performed by the author.

Reference:

1. Mao R, Liang J, Shen J, Ghosh S, Zhu LR, Yang H, Wu KC, Chen MH: **Implications of COVID-19 for patients with pre-existing digestive diseases**. *Lancet Gastroenterol Hepatol* 2020, **5**(5):426-428.
2. Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, Bi Z, Zhao Y: **Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China**. *Clin Res Cardiol* 2020, **109**(5):531-538.
3. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC *et al*: **Clinical Characteristics of Coronavirus Disease 2019 in China**. *N Engl J Med* 2020, **382**(18):1708-1720.
4. Guan WJ, Liang WH, Zhao Y, Liang HR, Chen ZS, Li YM, Liu XQ, Chen RC, Tang CL, Wang T *et al*: **Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis**. *Eur Respir J* 2020, **55**(5):2000547.
5. Das AB: **Disease association of human tumor suppressor genes**. *Mol Genet Genomics* 2019, **294**(4):931-940.
6. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL: **The human disease network**. *Proc Natl Acad Sci U S A* 2007, **104**(21):8685-8690.
7. Zheng C, Xu R: **Large-scale mining disease comorbidity relationships from post-market drug adverse events surveillance data**. *BMC Bioinformatics* 2018, **19**(Suppl 17):500.
8. Zhou X, Menche J, Barabasi AL, Sharma A: **Human symptoms-disease network**. *Nat Commun* 2014, **5**:4212.
9. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, O'Meara MJ, Rezelj VV, Guo JZ, Swaney DL *et al*: **A SARS-CoV-2 protein interaction map reveals targets for drug repurposing**. *Nature* 2020, **583**(7816):459-468.
10. Guney E, Menche J, Vidal M, Barabasi AL: **Network-based in silico drug efficacy screening**. *Nat Commun* 2016, **7**:10331.
11. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z *et al*: **DrugBank 5.0: a major update to the DrugBank database for 2018**. *Nucleic Acids Res* 2018, **46**(D1):D1074-D1082.
12. Gysi MD, Valle ID, Zitnik M, Ameli A, Gan X, Varol O, Sanchez H, Baron RM, Ghiassian D, Loscalzo J *et al*: **Network Medicine Framework for Identifying Drug Repurposing Opportunities for COVID-19** *arXiv:200407229* 2020.
13. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F: **Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2**. *Cell Discov* 2020, **6**:14.

14. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, Masciovecchio C, Angeletti S, Ciccozzi M, Gallo RC *et al*: **Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant**. *J Transl Med* 2020, **18**(1):179.
15. Basha O, Barshir R, Sharon M, Lerman E, Kirson BF, Hekselman I, Yeger-Lotem E: **The TissueNet v.2 database: A quantitative view of protein-protein interactions across human tissues**. *Nucleic Acids Res* 2017, **45**(D1):D427-D431.
16. Wang Q, Liu W, Ning S, Ye J, Huang T, Li Y, Wang P, Shi H, Li X: **Community of protein complexes impacts disease association**. *Eur J Hum Genet* 2012, **20**(11):1162-1167.
17. Gokhman D, Kelman G, Amartely A, Gershon G, Tsur S, Carmel L: **Gene ORGANizer: linking genes to the organs they affect**. *Nucleic Acids Res* 2017, **45**(W1):W138-W145.
18. Bassett DS, Bullmore ET: **Small-World Brain Networks Revisited**. *Neuroscientist* 2017, **23**(5):499-516.
19. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabasi AL: **Disease networks. Uncovering disease-disease relationships through the incomplete interactome**. *Science* 2015, **347**(6224):1257601.
20. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X *et al*: **Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study**. *Lancet* 2020, **395**(10229):1054-1062.
21. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, Liu S, Zhao P, Liu H, Zhu L *et al*: **Pathological findings of COVID-19 associated with acute respiratory distress syndrome**. *Lancet Respir Med* 2020, **8**(4):420-422.
22. Mo X, Jian W, Su Z, Chen M, Peng H, Peng P, Lei C, Chen R, Zhong N, Li S: **Abnormal pulmonary function in COVID-19 patients at time of hospital discharge**. *Eur Respir J* 2020, **55**(6):2001217.
23. Derebail VK, Parikh P, Jennette JC, Kshirsagar AV: **A rare cause of the pulmonary-renal syndrome: a case of atypical haemolytic-uraemic syndrome complicated by pulmonary haemorrhage**. *NDT Plus* 2008, **1**(6):417-419.
24. Lushina N, Kuo JS, Shaikh HA: **Pulmonary, Cerebral, and Renal Thromboembolic Disease Associated with COVID-19 Infection**. *Radiology* 2020, **296**(3):E181-E183.
25. Poggiali E, Vercelli A, Iannicelli T, Tinelli V, Celoni L, Magnacavallo A: **COVID-19, Chronic Obstructive Pulmonary Disease and Pneumothorax: A Frightening Triad**. *Eur J Case Rep Intern Med* 2020, **7**(7):001742.
26. Rajendram R, Kharal GA, Mahmood N, Puri R, Kharal M: **Rethinking the respiratory paradigm of COVID-19: a 'hole' in the argument**. *Intensive Care Med* 2020, **46**(7):1496-1497.
27. Ozgul U, Golbasi Z, Gulel O, Yildirim N: **Paradoxical and pulmonary embolism due to a thrombus entrapped in a patent foramen ovale**. *Tex Heart Inst J* 2006, **33**(1):78-80.

28. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, Fan Y, Zheng C: **Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study**. *Lancet Infect Dis* 2020, **20**(4):425-434.
29. Sugiura S, Ando Y, Minami H, Ando M, Sakai S, Shimokata K: **Prognostic value of pleural effusion in patients with non-small cell lung cancer**. *Clin Cancer Res* 1997, **3**(1):47-50.
30. Gehlbach BK, Geppert E: **The pulmonary manifestations of left heart failure**. *Chest* 2004, **125**(2):669-682.
31. Alqahtani JS, Oyelade T, Aldhahir AM, Alghamdi SM, Almehmadi M, Alqahtani AS, Quaderi S, Mandal S, Hurst JR: **Prevalence, Severity and Mortality associated with COPD and Smoking in patients with COVID-19: A Rapid Systematic Review and Meta-Analysis**. *PLoS One* 2020, **15**(5):e0233147.
32. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, Fernandez J, Prati D, Baselli G, Asselta R *et al*: **Genomewide Association Study of Severe Covid-19 with Respiratory Failure**. *N Engl J Med* 2020, **383**(16):1522-1534.
33. Cheng F, Kovacs IA, Barabasi AL: **Network-based prediction of drug combinations**. *Nat Commun* 2019, **10**(1):1197.
34. Rappaport N, Twik M, Plaschkes I, Nudel R, Iny Stein T, Levitt J, Gershoni M, Morrey CP, Safran M, Lancet D: **MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search**. *Nucleic Acids Res* 2017, **45**(D1):D877-D887.
35. Wagner GP, Pavlicev M, Cheverud JM: **The road to modularity**. *Nat Rev Genet* 2007, **8**(12):921-931.
36. Tripathi S, Moutari S, Dehmer M, Emmert-Streib F: **Comparison of module detection algorithms in protein networks and investigation of the biological meaning of predicted modules**. *BMC Bioinformatics* 2016, **17**:129.
37. Barabasi AL, Gulbahce N, Loscalzo J: **Network medicine: a network-based approach to human disease**. *Nat Rev Genet* 2011, **12**(1):56-68.
38. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E: **Fast unfolding of communities in large networks**. *Journal of Statistical Mechanics* 2008, **2008**(10):P10008.
39. Reichardt J, Bornholdt S: **Statistical mechanics of community detection**. *Phys Rev E Stat Nonlinear Soft Matter Phys* 2006, **74**(016110).
40. Rahiminejad S, Maurya MR, Subramaniam S: **Topological and functional comparison of community detection algorithms in biological networks**. *BMC Bioinformatics* 2019, **20**(1):212.
41. Kollias A, Kyriakoulis KG, Dimakakos E, Poulakou G, Stergiou GS, Syrigos K: **Thromboembolic risk and anticoagulant therapy in COVID-19 patients: emerging evidence and call for action**. *Br J Haematol* 2020, **189**(5):846-847.
42. Lee MJ, Ye AS, Gardino AK, Heijink AM, Sorger PK, MacBeath G, Yaffe MB: **Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks**. *Cell* 2012, **149**(4):780-794.
43. Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M: **Drug-target network**. *Nat Biotechnol* 2007, **25**(10):1119-1126.

44. Pinero J, Bravo A, Queralt-Rosinach N, Gutierrez-Sacristan A, Deu-Pons J, Centeno E, Garcia-Garcia J, Sanz F, Furlong LI: **DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants**. *Nucleic Acids Res* 2017, **45**(D1):D833-D839.
45. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK: **Metascape provides a biologist-oriented resource for the analysis of systems-level datasets**. *Nat Commun* 2019, **10**(1):1523.
46. 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**(7):976-978.
47. Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN *et al*: **Visualizing and interpreting cancer genomics data via the Xena platform**. *Nat Biotechnol* 2020, **38**(6):675-678.

Figure legends

Fig. 1 a) Neighbourhood interaction network of SARS-CoV-2 targets (STN) in the lung. The size of the node is proportional to its degree. b) SARS-CoV-2 targets form a large connected component (LCC) of size 181 in lung interactome. The size of the LCC is significantly larger than the random expectation.

Fig. 2 Disease-gene association network. a) Lung disease –gene network (LDGN), including COVID19 (yellow node). The network shows the SARS-CoV-2 targets (red) and neighborhood genes (green). b) & c) Dot plot shows the highly connected diseases (k>20) and genes in LDGN, respectively.

Fig. 3 Disease-disease association network (DDAN) a) Shows DDAN, including COVID19, red nodes represent the diseases that are directly direct linked to COVID19. b) Scatter plot shows the degree distribution of DDAN, which does not follow the scale-free property. c) The average path length between the diseases in DDNA and distribution of average path length of 1000 random networks (green). d) Transitivity of DDNA and distribution of transitivity of 1000 random networks (pink).

Fig.4 Network-based separation (S_{ab}) and pathobiological similarities. a-j show observed S_{ab}, z-score (red arrow) and distribution of S_{ab}^{ran} of top 10 overlapping lung disease with COVID-19 (here, ARMCM indicates abnormal respiratory motile cilium morphology). k box plot represents the pairwise correlation between genes is significantly(p-value<0.0001) higher than the random gene sets. l,m box plots show the distribution of functional

similarities (MF) and GO processes (BP) between the genes involved in lung disease and COVID-19. The GO processes and functional similarity between the genes are significantly high ($p < 0.0001$) compared to the random gene sets (note, in Fig. k,l, and m 1-10 indicates disease in a similar sequence as it is mentioned in Fig. a to j). n shows the strategy of drug repurposing to target the COVID19 module.

Fig. 5 Community detection in STN and functional protein module. a,b,c &d show the modules 1 to 4, pathway and process enrichment analysis of each module, their disease associations, and positive correlation between genes in each module in healthy lung tissue. The pairwise correlation between genes in each module is significantly (***, p-value<0.0001) higher than the random gene sets.

Fig. 6 Targetable protein in functional modules: The red nodes in each module indicate the FDA-approved targets.

Figures

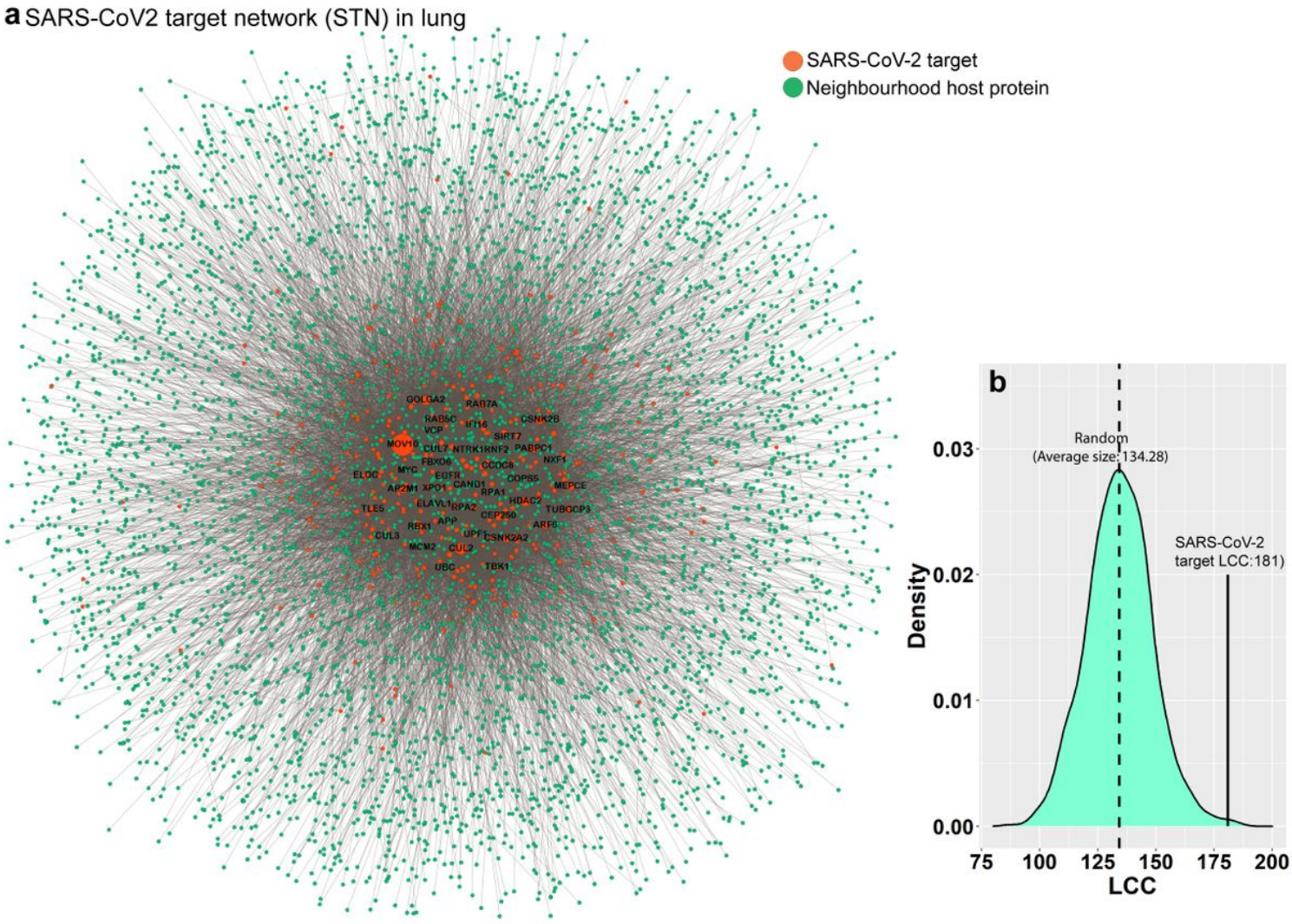

Figure 1

a) Neighbourhood interaction network of SARS-CoV-2 targets (STN) in the lung. The size of the node is proportional to its degree. b) SARS-CoV-2 targets form a large connected component (LCC) of size 181 in lung interactome. The size of the LCC is significantly larger than the random expectation.

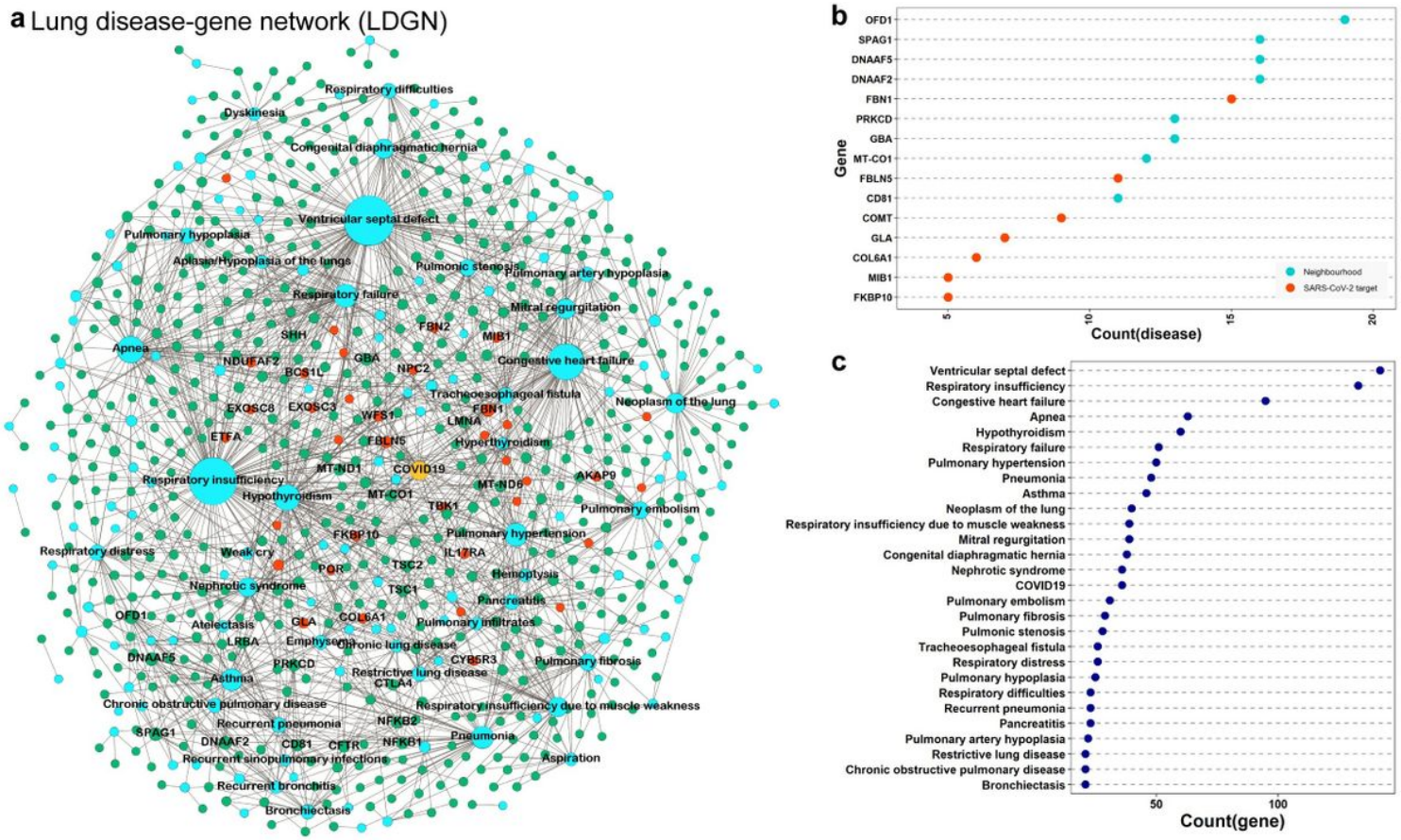

Figure 2

Disease-gene association network. a) Lung disease –gene network (LDGN), including COVID19 (yellow node). The network shows the SARS-CoV-2 targets (red) and neighborhood genes (green). b) & c) Dot plot shows the highly connected diseases (k>20) and genes in LDGN, respectively.

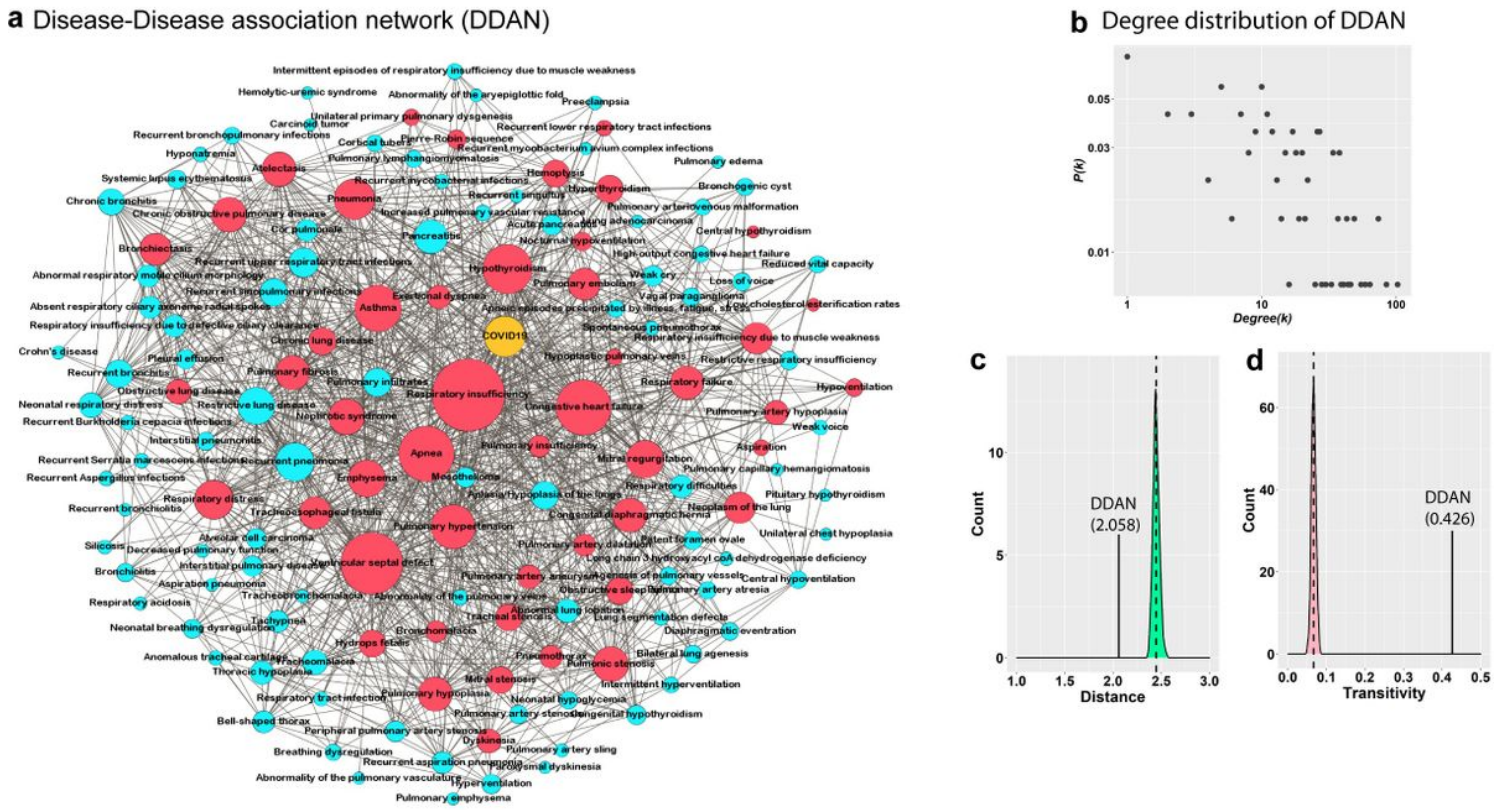

Figure 3

Disease-disease association network (DDAN) a) Shows DDAN, including COVID19, red nodes represent the diseases that are directly direct linked to COVID19. b) Scatter plot shows the degree distribution of DDAN, which does not follow the scale-free property. c) The average path length between the diseases in DDNA and distribution of average path length of 1000 random networks (green). d) Transitivity of DDNA and distribution of transitivity of 1000 random networks (pink).

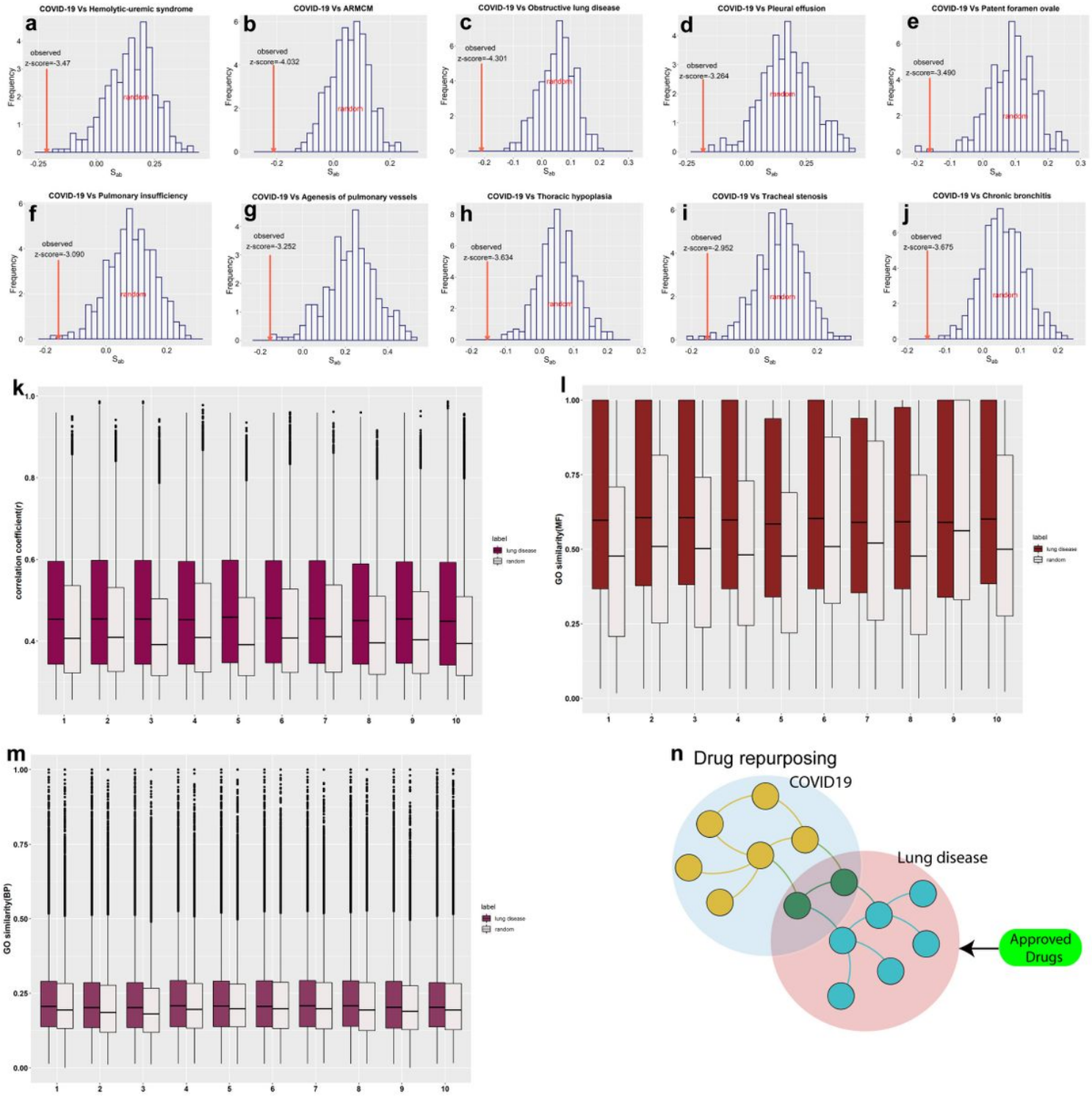

Figure 4

Network-based separation (Sab) and pathobiological similarities. a-j show observed Sab, z-score (red arrow) and distribution of Sabran of top 10 overlapping lung disease with COVID-19 (here, ARMCM indicates abnormal respiratory motile cilium morphology). k box plot represents the pairwise correlation between genes is significantly(p-value<0.0001) higher than the random gene sets. l,m box plots show the distribution of functional similarities (MF) and GO processes (BP) between the genes involved in lung disease and COVID-19. The GO processes and functional similarity between the genes are significantly

high ($p < 0.0001$) compared to the random gene sets (note, in Fig. k,l, and m 1-10 indicates disease in a similar sequence as it is mentioned in Fig. a to j). n shows the strategy of drug repurposing to target the COVID19 module.

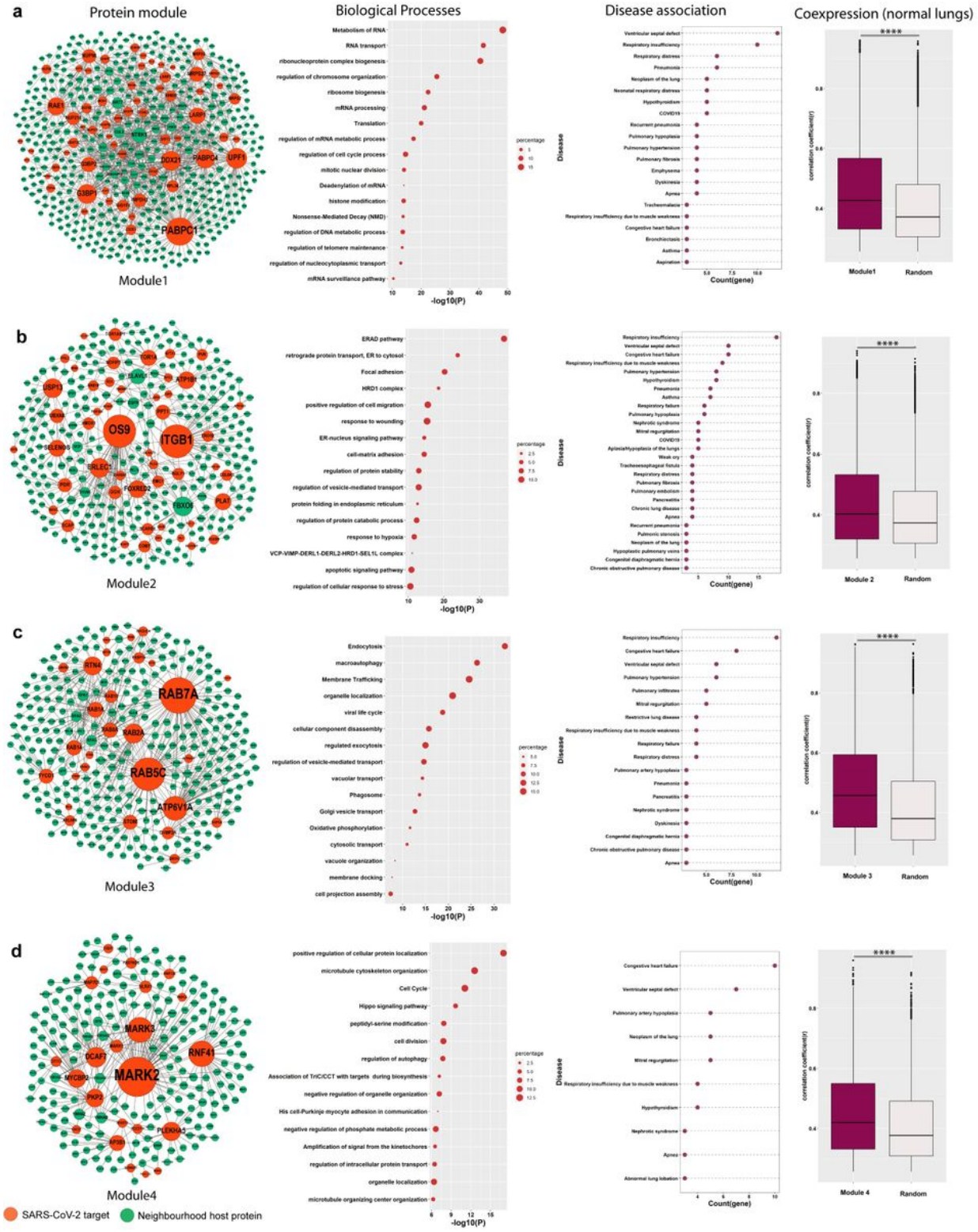

Figure 5

Community detection in STN and functional protein module. a,b,c &d show the modules 1 to 4, pathway and process enrichment analysis of each module, their disease associations, and positive correlation

between genes in each module in healthy lung tissue. The pairwise correlation between genes in each module is significantly (***, p-value<0.0001) higher than the random gene sets.

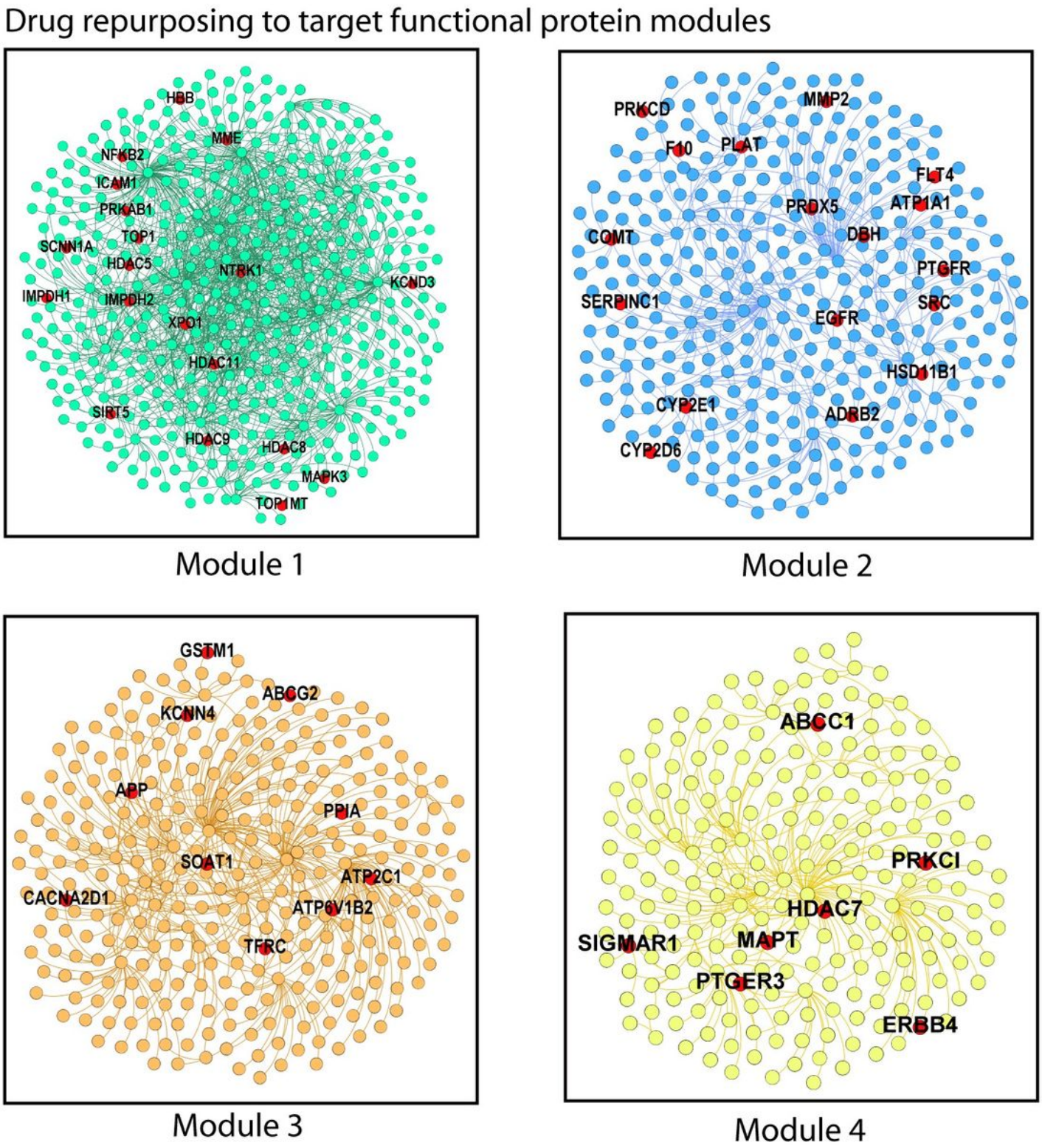

Figure 6

Targetable protein in functional modules: The red nodes in each module indicate the FDA-approved targets.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.pdf
- SupplementaryTable1.xlsx
- SupplementaryTable23and4.xlsx
- SupplementaryTable56and7.xlsx
- SupplementaryTable8910and11.xlsx