

# **Lung biopsy cells transcriptional landscape from COVID-19 patient stratified lung injury in SARS-CoV-2 infection through impaired pulmonary surfactant metabolism**

Abul Bashar Mir Md. Khademul Islam<sup>1\*</sup> and Md. Abdullah-Al-Kamran Khan<sup>2</sup>

<sup>1</sup>Department of Genetic Engineering & Biotechnology, University of Dhaka, Dhaka, Bangladesh.

<sup>2</sup>Department of Mathematics and Natural Sciences, BRAC University, Dhaka, Bangladesh.

## **\* Correspondence:**

Dr. Abul Bashar Mir Md. Khademul Islam  
Associate Professor  
Department of Genetic Engineering and Biotechnology  
University of Dhaka  
Dhaka 1000, Bangladesh.  
Email: khademul@du.ac.bd

## **Keywords**

SARS-CoV-2, COVID-19, Surfactant metabolism, Lung injury, surfactant therapy, respiratory complications

## Abstract

Clinical management of COVID-19 is still complicated due to the lack of therapeutic interventions to reduce the breathing problems, respiratory complications and acute lung injury - which are the major complications of most of the mild to critically affected patients and the molecular mechanisms behind these clinical features are still largely unknown. In this study, we have used the RNA-seq gene expression pattern in the COVID-19 affected lung biopsy cells and compared it with the effects observed in typical cell lines infected with SARS-CoV-2 and SARS-CoV. We performed functional overrepresentation analyses using these differentially expressed genes to signify the processes/pathways which could be deregulated during SARS-CoV-2 infection resulting in the symptomatic impairments observed in COVID-19. Our results showed that the significantly altered processes include inflammatory responses, antiviral cytokine signaling, interferon responses, and interleukin signaling etc. along with downmodulated processes related to lung's functionality like-responses to hypoxia, lung development, respiratory processes, cholesterol biosynthesis and surfactant metabolism. We also found that the viral protein interacting host's proteins involved in similar pathways like: respiratory failure, lung diseases, asthma, and hypoxia responses etc., suggesting viral proteins might be deregulating the processes related to acute lung injury/breathing complications in COVID-19 patients. Protein-protein interaction networks of these processes and map of gene expression of deregulated genes revealed that several viral proteins can directly or indirectly modulate the host genes/proteins of those lung related processes along with several host transcription factors and miRNAs. Surfactant proteins and their regulators SPD, SPC, TTF1 etc. which maintains the stability of the pulmonary tissue are found to be downregulated through viral NSP5, NSP12 that could lead to deficient gaseous exchange by the surface films. Mitochondrial dysfunction owing to the aberration of NDUFA10, NDUFAF5, SAMM50 etc. by NSP12; abnormal thrombosis in lungs through atypical PLAT, EGR1 functions by viral ORF8, NSP12; dulled hypoxia responses due to unusual shift in HIF-1 downstream signaling might be the causative elements behind the acute lung injury in COVID-19 patients. Our study put forward a distinct mechanism of probable virus induced lung damage apart from cytokine storm and advocate the need of further research for alternate therapy in this direction.

## Introduction

The recent Coronavirus Disease (COVID-19) pandemic has affected approximately 3.5 million people across 212 countries and territories and the number active cases are still on the rise (at 30 April, 2020) [1]. So far till the writing of this article around 7% of the infected population has suffered death [1] and the fatality rate is continuously increasing due to the lack of detailed knowledge of the molecular mechanism of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infection and proper targeted therapeutic approaches against it.

SARS-CoV-2 is a single stranded positive sense enveloped RNA virus and belongs to the betacoronavirus genus of coronavirus [2]. It has 11 protein coding genes encompassing its ~29.9Kb genome [3]. About 90% genomic similarity was observed between SARS-CoV-2 and bat derived SARS-like coronavirus, while SARS-CoV-2 genome is ~79% and ~50% similar with that of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome related Coronavirus (MERS-CoV), respectively [2, 4, 5]. Lu et al. (2020) showed the considerable differences between SARS-CoV-2 and SARS-CoV genomes, as in SARS-CoV-2 there has been 380 amino acids substituted, ORF8a deleted, ORF8b elongated, and ORF3b truncated [2]. Though the genomic features are almost similar like those of SARS-CoV, SARS-CoV-2 showed some unique clinical and pathophysiological features like- prolonged incubation period [6], latency inside the host [7] etc. which are making the clinical management of this virus difficult.

Based on the clinical exhibitions of COVID-19, most of the mild to critically affected patients show respiratory complications like- moderate to severe pneumonia, which can further turn into acute respiratory distress syndrome (ARDS), sepsis, and multiple organ dysfunction (MOD) in severely ill patients [8]. Most of these clinical symptoms are associated primarily with respiratory system, specifically with lungs [9] leading to the depleted lung functionality; while the complications with other systems like- cardiovascular system, nervous system etc. were also reported in few patients [10, 11]. Recently cases of pulmonary embolism in the lungs of COVID-19 patients are reported [12].

In SARS-CoV and MERS-CoV infections, increased level of pro-inflammatory cytokines were observed which in turn increased the activation and recruitment of inflammatory cells into the lung tissues, causing the acute lung injury [13]. Similarly, increased levels of many

pro-inflammatory cytokines were also detected in moderate to critically affected COVID-19 patients [14], leading to the respiratory failure from ARDS. However the complex interplays between pro-inflammatory and anti-inflammatory cytokines are yet to be completely illustrated. Apart from the cytokine storm, other factors like- host innate immunity, autoimmunity against the pulmonary epithelial and endothelial cells, host genetics and epigenetic factors also play important role in the pathogenesis of SARS-CoV infection [15, 16]. Moreover, the multifaceted host-virus interactions are also found to be a key player in the pathogenesis of previous coronavirus infections [17].

Previously, transcriptional responses in COVID-19 were experimentally recorded using various *in vivo* or *in vitro* approaches like- using cell lines, animal models and COVID-19 infected lungs with SARS-CoV-2 [18], Nasopharyngeal (NP) swab [19], bronchoalveolar lavage fluid of COVID-19 patients [20] etc. which might not provide the actual image of the host transcriptional responses in lung damage upon the viral infection. How the deregulation in the lung's gene expression profiles is related to the pathogenesis of the SARS-CoV-2 infection and overall pathophysiology of lung failure is not yet devised. Moreover, how the lung's functionality related pathways are modulated by SARS-CoV-2 is still elusive. As a result, designing of more targeted therapeutic approaches for the clinical management of the COVID-19 patients is still very complicated. In this regard, we have analyzed a publicly available transcriptome data from the lung biopsy of a COVID-19 patient and summarized which pathways are particularly modulated during the SARS-CoV-2 infection, and how the virus is playing a role in the deregulation of the biological processes/pathways related to lungs' overall functionality and causing acute lung injury in COVID-19 patients.

## **Materials and Methods**

### **Analysis of microarray expression data**

Microarray expression data on SARS-CoV infected 2B4 cells or uninfected controls for 24 hrs obtained from Gene Expression Omnibus (GEO), accession: GSE17400 (<https://www.ncbi.nlm.nih.gov/geo>) [21]. Raw Affymatrix CEL files were background corrected, normalized using Bioconductor package "affy v1.28.1" using 'rma' algorithm. Quality of microarray experiment (data not shown) was verified by Bioconductor package "arrayQualityMetrics v3.44.0" [22]. Differentially expressed (DE) between two experimental conditions were called using Bioconductor package Limma [23]. Probe annotations were converted to genes using in-house python script basing the Ensembl gene model (Biomart 99)

[24]. The highest absolute expression value was considered for the probes that were annotated to the same gene. We have considered the genes to be differentially expressed, which have FDR [25] p-value  $\leq 0.05$  and Log2 fold change value  $\geq 0.25$  (Supplementary file 1).

### **Analysis of RNA-seq expression data**

Illumina sequenced RNA-seq raw FastQ reads were extracted from GEO database accession: GSE147507 [21]. This data include independent biological triplicates of primary human lung epithelium (NHBE) which were mock treated or infected with SARS-CoV-2 for 24hrs; two technical replicate of post-mortem lung biopsy sample of a COVID-19 deceased patient, along with lung biopsy of two different healthy person. We have checked the raw sequence quality using FastQC program (v0.11.9) [26] and found that the "Per base sequence quality", and "Per sequence quality scores" were high over threshold for all sequences (data not shown). Mapping of reads was done with TopHat (tophat v2.1.1 with Bowtie v2.4.1) [27]. Short reads were uniquely aligned allowing at best two mismatches to the human reference genome from (GRCh38) as downloaded from UCSC database [28]. Sequence matched exactly more than one place with equally quality were discarded to avoid bias [29]. The reads that were not mapped to the genome were utilized to map against the transcriptome (junctions mapping). Ensembl gene model [30] (version 99, as extracted from UCSC) was used for this process. After mapping, we used SubRead package featureCount (v2.21) [31] to calculate absolute read abundance (read count, rc) for each transcript/gene associated to the Ensembl genes. For differential expression (DE) analysis we used DESeq2 (v1.26.0) with R (v3.6.2; 2019-07-05) [32] that uses a model based on the negative binomial distribution. To avoid false positive, we considered only those transcripts where at least 10 reads are annotated in at least one of the samples used in this study and also applied a minimum Log2 fold change of 0.5 for to be differentially apart from adjusted p-value cut-off of  $\leq 0.05$  by FDR. Raw read counts of this experiment is provided in supplementary file 2. To assess the fidelity of the RNA-seq data used in this study and normalization method applied here, we checked the normalized Log2 expression data quality using R/Bioconductor package "arrayQualityMetrics (v3.44.0)" [22]. From this analyses, in our data no outlier was detected by "Distance between arrays", "Boxplots", and "MA plots" methods and replicate samples are clustered together (Supplementary file 3).

### **Retrieval of the host proteins that interact with SARS-CoV and SARS-CoV-2**

We have obtained the list of human proteins that forms high confidence interactions with SARS-CoV and SARS-CoV-2 proteins from conducted previously studies [33-35] and processed their provided proteins name into the associated HGNC official gene symbol.

### **Functional enrichment analysis**

We utilized Gtools (v1.8.4) for enrichment analysis and heatmap generation [36]. We have utilized the Gene Ontology Biological Processes (GOBP) [37], Reactome pathway [38], Bioplanet pathways [39], HumanCyc database [40], DisGeNet [41], KEGG pathway [42] modules, and a custom in house built combined module (Supplementary file 4) for the overrepresentation analysis. Resulting p-values were adjusted for multiple testing using the Benjamin and Hochberg's method of False Discovery Rate (FDR) [25].

### **Mapping of the deregulated genes in cellular pathways**

We have utilized Reactome pathway browser [38] for the mapping of deregulated genes of SARS-CoV-2 infection in different cellular pathways. We then searched and targeted the pathways which are found to be enriched for lung related functionalities.

### **Obtaining the transcription factors which can modulate the differential gene expression**

We have obtained the transcription factors (TFs) which bind to the given differentially expressed genes using a custom TFs module created using ENCODE [43], TRRUST [44], and ChEA [45] database.

### **Obtaining human miRNAs target genes**

We extracted the experimentally validated target genes of human miRNAs from miRTarBase database [46].

### **Extraction of transcription factors modulate human miRNA expression**

We have downloaded the experimentally validated TFs which bind to miRNA promoters and module it from TransmiR (v2.0) database which provides regulatory relations between TFs and miRNAs [47]. For miRNAs come to play role in transcriptional regulation, we have considered those TFs that are expressed (upregulated) itself and that can 'activate' or 'regulate' miRNAs, or in absence of TFs (downregulation) that could otherwise 'suppress' miRNAs.

### **Identification of the host epigenetic factors genes**

We used EpiFactors database [48] to find human genes related to epigenetic activity.

### **Construction of biological networks**

Construction, visualization and analysis of biological networks with differentially expressed genes, their associated transcription factors, associated human miRNAs, and interacting viral proteins were executed in the Cytoscape software (v3.8.0) [49]. We used STRING [50] database to extract highest confidences (0.9) edges only for the protein-protein interactions to reduce any false positive connection.

### **Results**

#### **Antiviral immune responses and organ specific functionalities are deregulated in lungs**

During a viral respiratory infection, many pathways of the host get fine-tuned to battle with the invading pathogen; on the other side, the infecting viruses also try to hijack and modulate host pathways for immune evasion and survival inside the host [51]. These complex interactions lead to the disease complexity from a viral infection, causing several critical pathophysiological conditions in host respiratory system [51]. To explore which particular biological processes/pathways are deregulated in SARS-CoV-2 infection we first identified the deregulated genes in both SARS-CoV and SARS-CoV-2 infections and performed functional enrichment analyses and then did comparative analyses [36].

We found 3031 (2408 upregulated and 623 downregulated), 142 (91 upregulated and 51 downregulated) and 6714 (2476 upregulated and 4238 downregulated) genes from SARS-CoV infected 2B4 cells, SARS-CoV-2 infected NHBE cells and lung biopsy of COVID-19 patient, respectively (Supplementary Figure 1). We discovered a wide array of genes to be differentially expressed in SARS-CoV-2 infected lung whose expression profiles were observed to be different in SARS-CoV-2 infection in NHBE cells and SARS-CoV infection (Supplementary Figure 1).

As expected, from the enrichment analyses of GOBP, we have observed that biological processes related to antiviral inflammatory responses, viral processes etc. overrepresented in both types of SARS-CoV-2 infections (cell line and primary lung biopsy cells) and SARS-CoV infection (Figure 1A). While several biological processes like- negative regulation of viral replication, immune system process, response to hypoxia, heart development etc. were only enriched for both SARS-CoV-2 infections (Figure 1A), few pivotal processes namely- viral transcription, adaptive immune response, brain development, lung development,

respiratory gaseous exchange by respiratory system were uniquely observed for the deregulated genes from COVID-19 affected lung (Figure 1A).

Similarly, from 'Bioplanet' module, host antiviral immune responses like- various inflammatory cytokine signaling pathways, apoptosis, interferon-I signaling etc. were activated for all of these infections (Figure 1B, 1C). While, HIF-1 signaling, heart development, asthma, type-II interferon signaling etc. pathways were found in both SARS-CoV-2 infections (Figure 1B, Supplementary figure 2), curiously, some pathways like- disease associated with surfactant metabolism, ROS/RNS production, fatty acyl-CoA biosynthesis, ER-phagosome pathway, inflammasomes etc. were only found in SARS-CoV-2 infected lung biopsy (Figure 1C, Supplementary figure 2). To further understand, we used the module 'DisGeNet' for enrichment analysis which revealed that the deregulated genes of SARS-CoV and SARS-CoV-2 infections are also involved in different diseases like- virus diseases, lung diseases, asthma, pneumonitis, hypoxia etc (Figure 1D, Supplementary figure 2). Interestingly, several cholesterol biosynthesis pathways were found to be deregulated only in the lung biopsy sample of COVID-19 patient (Figure 1E, Supplementary Figure 2). As cholesterol in lung plays important roles in maintaining normal lung physiology and protection against many diseases [52], downregulation of these indicates possible association with lung-related comorbidities of COVID-19 patients.

From these enrichment analyses, it is clearly understood that several pathways related to lung's overall functionality are being deregulated by the infecting SARS-CoV-2 virus. While hunting for more definitive clues on which particular processes are being regulated during this infection, we again performed another round of enrichment analysis with our in house combined module (Supplementary file 4) which gather related modules with few genes to a parent term/process which otherwise left out from analysis due to statistical stringency cutoff (module with 10 genes are selectedly) during enrichment analysis. Surprisingly from this enrichment analysis, we have detected several important lung's overall functionality related processes only for the deregulated genes from the lung of COVID-19 patient, namely- lung development, pulmonary surfactant metabolism disease or dysfunction, respiratory processes, regulation of respiratory gaseous exchange etc. along with some antiviral responses (Figure 2, Supplementary figure 3). We have also exported enriched genes list of selected significantly enriched terms in color coded heatmaps. Intriguingly, while equating the expression of these enriched genes, we discovered that the genes of these key lung related processes are



significantly altered in SARS-CoV-2 infected lungs compared to the SARS-CoV-2 infected NHBE cells and SARS-CoV infection (Supplementary figure 4).

### **Genes in lung surfactant metabolism pathway are deregulated in COVID-19 patient's lung**

Pulmonary surfactant proteins play important role in maintaining the surface tension at the air-liquid interface in the alveoli for efficient gaseous exchange, and can also modulate immunological functions of lung's innate immune cells to eliminate pathogens [53]. Moreover, their role in viral infections as well as in impeding inflammatory responses and clearance of apoptotic cells in lungs are previously reported [53]. Our above analysis as indicated that surfactant metabolism could be a target of this virus, we sought to check the routes of deregulations of this pathway in SARS-CoV-2 infected lung and how they are affecting the downstream signaling processes to impede normal lung function through impaired surfactant metabolism. To achieve this goal, we have mapped the differentially expressed genes in this pathway using Reactome pathway browser [38]. We looked deeply into the mechanisms of this pathway to elucidate the probable alterations happening in COVID-19 affected lung.

In normal lung, TTF1-CCDC59 complex can transactivate *SFTPB* and *SFTPC* gene expression which in turn play an important role by regulating the alveolar surface tension [54]. But in COVID-19 affected lung, *TTF1* and *SFTPC* genes are found to be downregulated, whereas *SFTPB* is upregulated (Figure 3). GATA6 transcription factor promotes the transcription of *SFTPA* gene [55] which is involved in immune and inflammatory responses along with lowering the surface tension in the alveoli [56] and both of them are found to be upregulated in SARS-CoV-2 infected lung while the GATA6 antagonist LMCD1 was downregulated (Figure 3). While the complex of CSF2RA and CSF2RB can bind GM-CSF to induce activation of macrophages [57] and degradation of STFPs in the alveolar macrophages [58]. We found *CSF2RA* downregulated and *CSF2RB* upregulated in the lung of the COVID-19 affected patient (Figure 3). Pro-SFTPB and Pro-SFTPC are cleaved NAPSA, CTSH, PGA3-5 for producing active SFTPB and SFTPC [59, 60] but in the COVID-19 affected lung NAPSA was found deregulated (Figure 3) while the rest other enzymes were not found to be significantly differentially expressed (data not shown). Overall, the transcription of the surfactant genes, production of active surfactant proteins, and their turnover might be deregulated in the lung of the COVID-19 patient which

could have result in the severe disease complicacies leading to their death. As these mechanisms are found to be aberrantly modulated upon SARS-CoV-2 infections, virus might be positively facilitating these anomalies.

### **Host proteins those interact with virus are involved in different respiratory functions related pathways and diseases**

In previously reported human coronavirus infections, SARS and MERS coronaviruses are often found to takeover host machineries, suppressing host immune responses and other important biological processes for their continued existence inside the infected cells [61]. We have performed functional enrichment analyses using the previously reported host proteins which can interact with SARS-CoV and SARS-CoV-2 proteins [33-35] to have better understanding in which pathways these proteins are involved and if there any connection with lung degenerative processes – a common complication that we observe in corona infected patient.

Upon the enrichment analyses, as expected, we observed several immune signaling pathways like- interleukin signaling, interferon signaling, apoptosis, inflammasomes etc. for both SARS-CoV and SARS-CoV-2 proteins (Figure 4A, 4C); however, this approach also revealed some vital pathways related to respiratory functions like- HIF-1 signaling, hypoxic and oxygen homeostasis regulation of HIF-1 alpha etc (Figure 4A, 4C). Several respiratory complications related pathways like- lung diseases, asthma, hyperoxia, respiratory failure, pulmonary hypertension etc. were found to be enriched from DisGenNet database module [41] (Figure 4B). All these enlighten that SARS-CoV-2 might be utilizing its proteins in modulating the host lung's normal physiological and immune responses which can be further explored by linking these viral-host protein-protein interactions (PPI) in our previously identified essential lung processes.

### **SARS-CoV-2 proteins and host epigenetic regulators can modulate the functionality of lung and other respiratory processes**

As we have discovered that both deregulated genes in COVID-19 affected lung and SARS-CoV-2 interacting proteins are involved in different important respiratory functionalities, we then produced several functional networks with these deregulated genes, viral protein-host protein interactions, and host's epigenetic regulators involved in those processes to shed insights on viral infection mediated deregulations and their resultant pathophysiological effects in COVID-19. We have mainly targeted four broad biological processes which can

significantly affect the COVID-19 patients upon any dysregulation, namely- response to hypoxia, lung development, respiratory processes, and surfactant metabolism.

Numerous genes in the hypoxic responses and HIF-1 alpha signaling were found to be abruptly regulated in the SARS-CoV-2 infected lung (Supplementary Figure 5). PLAT, a tissue plasminogen activator, have profound roles in maintaining the normal homeostasis of lung and its aberrant regulation can lead to many lung injuries [62]. In the response to hypoxia process, in PPI map of differentially expressed genes of combined GOBP module 'response to hypoxia' and SARS-CoV-2 target genes [33], this PLAT protein was found to be directly targeted by viral protein ORF8 (Figure 5). Moreover, several indirect responses from the viral protein interactions were also revealed (Figure 5). SARS-CoV-2 M protein can target STOM which interacts with *SLC2A1*. *SLC2A1* can also be targeted by host miRNA miR-320a (Figure 5). ORF8 which interacts with OS9 can modulate EGLN1 and EGLN2. KCNMA1 is found to be interacting with host proteins ATP1B1 and PRKACA which interact with viral M and NSP13 proteins, respectively (Figure 5). ORF9c can modulate EDNRA indirectly through F2RL1 (Figure 5). PML functions might be altered through viral N-host MOV10 proteins' interaction. Functions of SLCBA1 can be modulated by NSP13 through PRKACA (Figure 5). NSP7 can regulate ALDH3A1 which can affect CYP1A1 and NR4A2 (Figure 5). NSP5 interacting HDAC2 can curb PML and REST functionalities (Figure 5). NSP12 can modulate the a wide range of hypoxia functions related proteins like-TGFBR2, MECP2, MTHFR, CBFA2T3, EGR1, ANGPTL4 by interacting with the transcription factor TCF12 (Figure 5). NSP12 can deregulate the functions of CFLAR through RIPK1 interactions (Figure 5). Apart from these several host miRNAs can possibly also downregulate the expression of some genes, namely- miR-320a, miR-3188, miR-3661, miR-217, miR-421 and miR-429 (Figure 5). These viral mediated deviations found in the hypoxia responses might be a decisive factor behind the lung injury found in COVID-19 patients [63].

In the lung development network, viral protein NSP13 are found to indirectly target RC3H2, SOX9, GLI3 through CEP350, PRKACA, PRKAR2A proteins (Figure 6). While GLI3 can also be modulated through several viral-host interactions like- ORF10-RBX1, NSP5-HDAC2, M-PSMD8 interaction axis (Figure 6). ORF8 can indirectly modulate ADAMTS2, CHI3L1, NOTCH1, TCF12, FLT4 etc (Figure 6). Transcription factor TCF12 is targeted by viral NSP12 protein which can in turn affect transcriptions of *PDGFRA*, *PDGFRB*, *TGFBR2*, *ID1*, *HES1*, *LTBP3* genes (Figure 6). NSP5 can modulate NOTCH1 through HDAC2 (Figure

6). Moreover, host miRNAs like- miR630 can downregulate TGFBR2; while miR-206, miR320a and miR-375 are downregulated so that SOX9 and WYHAZ are overexpressing (Figure 6). Virus could hamper several growth factor signaling which are crucial for various lung injury repair mechanisms [64, 65].

From the respiratory process network we can delineate that several transcription factors are deregulated which are associated in transcribing ECSIT that is directly targeted by viral ORF9c and indirectly by viral ORF8, NSP7 (Figure 7). Moreover, ECSIT itself is directly targeted by ORF9c (Figure 7). NSP12 can modulate SFTPB, SFTPC, SLC5A3, DUSP10, SAMM50 etc. by targeting TCF12 (Figure 7). Also, in this network we have observed suppressive actions of miRNAs- miR-206, miR-217, miR-375 on NR4A2 and NDST1; as well as upregulation of YWHAZ due to probable downregulation of miR-320a (Figure 7). Virus might be dampening the host immune response in lung by targeting ECSIT [66], as well as could dull the respiratory gaseous exchange in lung by negatively modulating the surfactant proteins [67].

Surfactant metabolism is found to be modulated by not only viral proteins but also through the aberrant host response (Figure 8). Viral proteins NSP12 and NSP5 can target transcription factor TCF12 and epigenetic regulator HDAC2, respectively which in turn are modulating important members of surfactant metabolism process like- TTF1, CCDC59, SFTPB, SFTPC, CSF2RA, CSF2RB, NAPSA, SFTPD, DMBT1 etc (Figure 8). These can also be modulated through CKAP4 which are targeted by viral M, NSP2, NSP9, E, ORF8 proteins (Figure 8). Furthermore, we have observed that viral M and S proteins can interact with the proteins of this process both directly and indirectly (Supplementary figure 6). Host miRNAs miR-421 are found to be downregulating LMCD1; while miR-137, miR-375, miR-429 fail to modulate CCDC59 and GATA6 because of their probable inactivation/suppression by differential expression of its regulating TFs (Figure 8). As *SFTPD* and *SFTPC* are downregulated along with several regulatory partners, their primary function of immunomodulation and efficient air exchange in lung [68, 69] might be seriously hindered by the viral proteins; which could further lead to abnormal pathogenic lung injury [70].

Several epigenetic factors like- HDAC2, DNMT1, CUL2, MOV10, RBX1 and TLE1 which are involved in these processes can be targeted by different viral proteins, through which these abovementioned processes can be significantly altered. Epigenetic modulators are key modulators in balancing the normal lung pathobiology and anomalies of these regulations can

lead to many lung diseases [71]. HDAC2 and DNMT1 have significant roles in chronic obstructive pulmonary disease (COPD) progression [72, 73].

## Discussion

Though there are a wide variety of symptoms and clinical features seen in the COVID-19 patients, almost every mild to critically affected patients showed respiratory and breathing complications which ranges from pneumonia to acute lung injury [74]. Severely ill patients are mostly supported by artificial ventilation as no therapeutic drugs are discovered yet for mitigating these complications due to the missing understandings on the molecular aspects of lung related abnormalities in COVID-19. We have identified some important respiratory and lung related processes in which we got several genes are deregulated upon the progression of SARS-CoV-2 infection. So, we prioritized these pathways to check how viral proteins might be playing a role in these aberrant regulations and searched for potential therapies/drugs for the treatment to lessen the resultant effects of the aberration.

We have compared the deregulated genes from SARS-CoV-2 infected lung with the differentially expressed genes from SARS-CoV-2 infected cell line and SARS-CoV infection, to see how host lung and SARS-CoV-2 are responding upon the infections. We found that a significant variations in the number of total deregulated genes between these. Other report also suggests the deregulation of a huge number of genes when expression analyses are performed using host's infected cells [75]. That means cell lines infected with SARS-CoV-2 might not provide the complete outcomes of the infections. We also checked and verified the quality of our normalized reads to nullify the risk of probable poor sequencing and abnormal read normalization results; all these recommend the suitability of our generated differential expression analysis.

Several SARS-CoV-2 entry receptor (ACE2)/entry associated proteins (TMPRSS2, BSG, CTSL, DPP4) were previously discussed [76]; all of which are readily expressed in the lung, except ACE2 (Supplementary Figure 7A). While analyzing the expression profiling in SARS-CoV-2 infected lung, we discovered a quite unexpected scenario. Astoundingly, in COVID-19 affected lung, *ACE2* was upregulated while the genes of the entry associated proteins were downregulated (Supplementary Figure 7B). More COVID-19 patient data have to be analyzed in concluding this striking feature.

“Cytokine storm” is a much discussed phenomenon in previously reported pathogenic human coronavirus infections which can be lethal for the host for the destruction of its own

respiratory systems [77]. Similar responses can also occur in SARS-CoV-2 infection [78]. We have also observed that several inflammatory and antiviral responses are readily deregulated in the SARS-CoV-2 infected lung which could lead to the abnormalities in overall respiratory functions.

Apart from this, our study also identified dysregulations in several non-cytokine mediated lung function related vital processes, namely- response to hypoxia, lung development, respiratory process, and surfactant metabolism; aberrations in these pathways leading to lung pathobiology in COVID-19. Networks generated combining the viral-host protein interactions information suggest that viral proteins might be actively involved in these deregulations along with several host factors which were also decontrolled due to the viral infections.

Hypoxic conditions are common in respiratory infections due to reduced inhalation of oxygen [79]. Several genes which are involved in hypoxia induced responses are found to be deregulated (Figure 5) in COVID-19 affected lung. ALDH3A1 can protect airway epithelial cells from destruction [80]. Though it is upregulated its functionalities can be impeded through viral proteins (Figure 5). CFLAR functions in shutting down apoptotic responses by interacting with RIPK1 [81] but viral interactions with RIPK1 might prevent it (Figure 5). Virus can induce the transcription of ANGPTL4 by utilizing TCF12 transcription factor (Figure 5) which in turn could cause pulmonary tissue damage [82]. Viral protein could promote the activity of EGLN1 and EGLN2 (Figure 5) to suppress the transcription of HIF-induced genes in hypoxia [83]; on the other hand the regulation through these proteins might be hampered due to viral interactions and the constant overexpression of HIF mediated inflammatory genes could also occur, which could lead to inflammation induced lung damage. Severe hypoxia induced responses can occur through the downregulation of MECP2 [84] in SARS-CoV-2 patients (Figure 5). In hypoxia, REST is induced and can act as a negative regulator of gene expression to maintain a balance between different processes [85] which is here found downregulated and can be targeted through host miRNA miR-421 in COVID-19 (Figure 5). GLUT1 (*SLC2A1* gene) promotes increased glucose transport into hypoxic cells for its prolonged adaptation during this condition [86], but it is found to be downregulated in lung of COVID-19 patients and this can occur through miR-320a (Figure 5). During hypoxia, TGF-beta signaling regulates inflammation and vascular responses [87]; but attenuated TGF $\beta$  expression might lead to disease severity (Figure 5). EGR1 transcription factor can be modulated through viral proteins (Figure 5) in preventing hypoxia induced

EGR1 activation of HIF-1 alpha [88], which could curb the whole hypoxia induced survival responses. Moreover, defective hypoxia response can occur due to the overactivity of EGR1 that can result in anomalous thrombosis [89]. Also, PLAT which is a crucial factor in splitting down clots [90], this function might be directly altered by SARS-CoV-2 ORF8. Many COVID-19 patients are reported to have pulmonary embolism and thrombosis [91], this hypoxic responses could suggest the probable routes of this.

Similarly, several lung development and respiratory process genes/proteins were also found to be deregulated in the COVID-19 patient's lung (Figure 6, 7). SOX9 which is considered an important regulator for the recovery from acute lung injury [92], can be down-modulated by the virus (Figure 6). Though the host miRNAs cannot target the YWHAZ which is a pro-survival protein [93] and also can induce expression surfactant protein A2 [94], it can be indirectly modulated through viral protein NSP7 (Figure 6). Likewise, viral protein ORF8 can indirectly modulate CHI3L1 (Figure 6) which can suppress lung epithelium injury [95]. Virus can block TCF12 and inhibit the expression of PDGFRA and PDGFRB (Figure 6) which could play a role in lung maturation and injury response [96]. GLI3, which exerts essential role in developing the lung, and regulating the innate immune cells [97], is downregulated in COVID-19 lungs (Figure 6). LTBP3 can promote lung alveolarization [98] but can be modulated by SARS-CoV-2 protein NSP12 (Figure 6). Aberration in NOTCH signaling contributes significantly in various lung diseases [99], and in SARS-CoV-2 infection NOTCH1, HES1 is found to be downregulated (Figure 6). CD55, a member of the complement system which plays crucial role in host defense in airway epithelium [100], can be modulated by viral protein NSP12 (Figure 7). CYSLTR1 is upregulated in SARS-CoV-2 infection (Figure 7), which is correlated to COPD [101]. This protein can be modulated by viral ORF9c and ORF3a (Figure 7). ECSIT found to be directly targeted by viral ORF9c protein (Figure 7) to stop the ECSIT mediated antiviral innate immune response [66]. Several mitochondrial genes like- *NDUFA10*, *NDUFAF5*, *SAMM50* are found downregulated in COVID-19 affected lung (Figure 7); as mitochondria plays important role in cellular respiration and lung diseases [102], aberration of these might also lead to lung related complications. DUSP10 can regulate aberrant inflammatory response upon viral infections [103], but in SARS-CoV-2 infected lung this gene is downregulated which might have occurred through the interaction of NSP12-TCF12 interactions (Figure 7).

Pulmonary surfactant proteins are lipoproteins which mainly functions to lower the alveolar surface tension [104] as well as they can elicit some immune stimulatory roles against some

respiratory pathogens [105]. Among the surfactant proteins, SP-A and SP-D mainly evoke immune responses while SP-B and SP-D play roles in maintaining the efficient respiratory gaseous exchanges [69]. Several lung diseases like- asthma, acute respiratory distress syndrome (ARDS), COPD etc. are reported to be associated with aberration in the functionalities of surfactant proteins [70, 106]. In SARS-CoV-2 affected lung, production of the surfactant proteins are found to be deregulated (Figure 8). Viral protein NSP5 can recruit HDAC2 and downregulate the expression of *TTF1* which is needed for the expression of SP-B and SP-C (Figure 8). SP-A and SP-D can be targeted indirectly by several viral proteins (Figure 8) which can lead to the aberrant production of surfactant proteins in lungs, thus might complicating the disease condition. CSF2RA and CSF2RB complex modulate the surfactant recycling, thus maintains the overall balance of surfactant content [58]. We found *CSF2RA* is deregulated in SARS-CoV-2 infected lung which could lead in the aberration of surfactant recycling (Figure 8). While performing the enrichment analysis, we witnessed that several cholesterol biosynthesis pathways are significantly downregulated in COVID-19 patient's lung cells (Figure 1E) which could lead to low accumulation of phospholipids in lungs. As phosphatidylcholine (PC) and phosphatidylglycerol (PG) are the principal phospholipids of surfactant proteins [107], impeded production of lipids will make surfactant proteins non-functional.

Considering all these, lung surfactants might be useful in the treatment of COVID-19 patients, as lung surfactant replacement therapies were previously reported to be successful in other respiratory infections and acute lung injury to reduce the lung damage of the patients [104, 108, 109] and also our analysis found it as an alternative from enrichment of related process deregulated genes (data not shown). Also, other drugs like- respiratory stimulants for COPD [110], sargramostim for treating pulmonary alveolar proteinosis [111], and oseltamivir in curing influenza-related lower respiratory tract complications [112] showed potential improvements in lung's and respiratory system's overall condition. Further *in vitro* and *in vivo* research may be turn out to be useful for the patients who cannot tolerate painful ventilator.

From our results, we can suggest that surfactant proteins production along with other respiratory responses in lung could be deregulated. Along with the antiviral drugs to mitigate the viral responses, drugs which can improve the lung conditions in COVID-19 patients could also be considered as a treatment option for the patients. Our generated model can be useful



for further experiments, as well as more insights will be gathered upon the use of surfactant therapy in laboratory using SARS-CoV-2 infected model organisms.

### **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. The authors declare no conflict of interest.

### **Author's Contribution**

ABMMKI conceived the project, designed the workflow and performed the analyses. Both authors interpreted the results and wrote the manuscript. Both authors read and approved the final manuscript.

### **Funding**

This project was not associated with any internal or external source of funding.

### **Data Availability Statement**

Analyses generated data are deposited as supplementary files.

### **References**

1. Worldometer, *Coronavirus Cases*. 2020. p. 1-22.
2. Lu, R., et al., *Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding*. *The Lancet*, 2020. **395**(10224): p. 565-574.
3. NCBI-Gene, *Gene Links for Nucleotide (Select 1798174254) - Gene - NCBI*. 2020.
4. Jiang, S., L. Du, and Z. Shi, *An emerging coronavirus causing pneumonia outbreak in Wuhan, China: calling for developing therapeutic and prophylactic strategies*. *Emerging Microbes & Infections*, 2020. **9**(1): p. 275-277.
5. Ren, L.-L., et al., *Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study*. *Chinese medical journal*, 2020.
6. Lauer, S.A., et al., *The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application*. *Annals of Internal Medicine*, 2020.
7. Lan, L., et al., *Positive RT-PCR Test Results in Patients Recovered From COVID-19*. *JAMA*, 2020. **323**(15): p. 1502-1503.
8. Huang, C., et al., *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China*. *The Lancet*, 2020. **395**(10223): p. 497-506.

9. Galiatsatos, P. *What Coronavirus Does to the Lungs*. 2020 [cited 2020 24 April]; Available from: <https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/what-coronavirus-does-to-the-lungs>.
10. Mao, L., et al., *Neurologic Manifestations of Hospitalized Patients With Coronavirus Disease 2019 in Wuhan, China*. JAMA Neurology, 2020.
11. Zheng, Y.-Y., et al., *COVID-19 and the cardiovascular system*. Nature Reviews Cardiology, 2020. **17**(5): p. 259-260.
12. Rotzinger, D.C., et al., *Pulmonary embolism in patients with COVID-19: Time to change the paradigm of computed tomography*. Thrombosis research, 2020. **190**: p. 58-59.
13. Ye, Q., B. Wang, and J. Mao, *The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19*. Journal of Infection, 2020.
14. Mehta, P., et al., *COVID-19: consider cytokine storm syndromes and immunosuppression*. The Lancet, 2020. **395**(10229): p. 1033-1034.
15. Gu, J. and C. Korteweg, *Pathology and pathogenesis of severe acute respiratory syndrome*. The American journal of pathology, 2007. **170**(4): p. 1136-1147.
16. Schäfer, A. and R.S. Baric, *Epigenetic Landscape during Coronavirus Infection*. Pathogens (Basel, Switzerland), 2017. **6**(1): p. 8.
17. Fung, S.-Y., et al., *A tug-of-war between severe acute respiratory syndrome coronavirus 2 and host antiviral defence: lessons from other pathogenic viruses*. Emerging microbes & infections, 2020. **9**(1): p. 558-570.
18. Blanco-Melo, D., et al., *Imbalanced host response to SARS-CoV-2 drives development of COVID-19*. Cell, 2020.
19. Butler, D.J., et al., *Host, Viral, and Environmental Transcriptome Profiles of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)*. bioRxiv, 2020: p. 2020.04.20.048066.
20. Xiong, Y., et al., *Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients*. Emerging Microbes & Infections, 2020. **9**(1): p. 761-770.
21. Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets—update*. Nucleic Acids Research, 2012. **41**(D1): p. D991-D995.
22. Kauffmann, A., R. Gentleman, and W. Huber, *arrayQualityMetrics—a bioconductor package for quality assessment of microarray data*. Bioinformatics, 2009. **25**(3): p. 415-416.
23. Smyth, G.K., *Limma: linear models for microarray data*. 2005, Springer. p. 397-420.
24. Flicek, P., et al., *Ensembl 2008*. Nucleic acids research, 2007. **36**(suppl\_1): p. D707-D714.
25. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*. Journal of the Royal statistical society: series B (Methodological), 1995. **57**(1): p. 289-300.
26. Andrews, S., *FastQC: a quality control tool for high throughput sequence data*. 2010, Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
27. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq*. Bioinformatics, 2009. **25**(9): p. 1105-11.
28. Lander, E.S., et al., *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**(6822): p. 860-921.
29. Hansen, K.D., S.E. Brenner, and S. Dudoit, *Biases in Illumina transcriptome sequencing caused by random hexamer priming*. Nucleic Acids Res, 2010. **38**(12): p. e131.

30. Hubbard, T.J.P., et al., *Ensembl 2007*. Nucleic acids research, 2007. **35**(suppl\_1): p. D610-D617.
31. Liao, Y., G.K. Smyth, and W. Shi, *The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote*. Nucleic Acids Res, 2013. **41**(10): p. e108.
32. Anders, S. and W. Huber, *Differential expression analysis for sequence count data*. Genome Biology, 2010. **11**(10): p. R106-R106.
33. Gordon, D.E., et al., *A SARS-CoV-2 protein interaction map reveals targets for drug repurposing*. Nature, 2020.
34. Pfefferle, S., et al., *The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors*. PLoS pathogens, 2011. **7**(10): p. e1002331-e1002331.
35. Srinivasan, S., et al., *Structural Genomics of SARS-CoV-2 Indicates Evolutionary Conserved Functional Regions of Viral Proteins*. Viruses, 2020. **12**(4): p. 360.
36. Perez-Llamas, C. and N. Lopez-Bigas, *Gitools: analysis and visualisation of genomic data using interactive heat-maps*. PloS one, 2011. **6**(5).
37. Ashburner, M., et al., *Gene ontology: tool for the unification of biology*. Nature genetics, 2000. **25**(1): p. 25-29.
38. Jassal, B., et al., *The reactome pathway knowledgebase*. Nucleic Acids Res, 2020. **48**(D1): p. D498-d503.
39. Huang, R., et al., *The NCATS BioPlanet – An Integrated Platform for Exploring the Universe of Cellular Signaling Pathways for Toxicology, Systems Biology, and Chemical Genomics*. Frontiers in Pharmacology, 2019. **10**(445).
40. Romero, P., et al., *Computational prediction of human metabolic pathways from the complete human genome*. Genome Biology, 2004. **6**(1): p. R2.
41. Pinero, J., et al., *The DisGeNET knowledge platform for disease genomics: 2019 update*. Nucleic Acids Res, 2020. **48**(D1): p. D845-d855.
42. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes*. Nucleic acids research, 2000. **28**(1): p. 27-30.
43. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome*. Nature, 2012. **489**(7414): p. 57-74.
44. Han, H., et al., *TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions*. Nucleic Acids Res, 2018. **46**(D1): p. D380-d386.
45. Lachmann, A., et al., *ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments*. Bioinformatics (Oxford, England), 2010. **26**(19): p. 2438-2444.
46. Huang, H.-Y., et al., *miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database*. Nucleic Acids Research, 2019. **48**(D1): p. D148-D154.
47. Tong, Z., et al., *TransmiR v2.0: an updated transcription factor-microRNA regulation database*. Nucleic Acids Res, 2019. **47**(D1): p. D253-d258.
48. Medvedeva, Y.A., et al., *EpiFactors: a comprehensive database of human epigenetic factors and complexes*. Database, 2015. **2015**.
49. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. Genome Res, 2003. **13**(11): p. 2498-504.
50. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets*. Nucleic acids research, 2019. **47**(D1): p. D607-D613.
51. Manjarrez-Zavala, M.E., et al., *Pathogenesis of viral respiratory infection*. Respiratory Disease and Infection: A New Insight, 2013: p. 1.

52. Fessler, M.B., *A New Frontier in Immunometabolism. Cholesterol in Lung Health and Disease*. Annals of the American Thoracic Society, 2017. **14**(Supplement\_5): p. S399-S405.
53. Glasser, J.R. and R.K. Mallampalli, *Surfactant and its role in the pathobiology of pulmonary infection*. Microbes and infection, 2012. **14**(1): p. 17-25.
54. Yang, M.C., et al., *The TTF-1/TAP26 complex differentially modulates surfactant protein-B (SP-B) and -C (SP-C) promoters in lung cells*. Biochem Biophys Res Commun, 2006. **344**(2): p. 484-90.
55. Bruno, M.D., et al., *GATA-6 activates transcription of surfactant protein A*. J Biol Chem, 2000. **275**(2): p. 1043-9.
56. Silveyra, P. and J. Floros, *Genetic complexity of the human surfactant-associated proteins SP-A1 and SP-A2*. Gene, 2013. **531**(2): p. 126-32.
57. Martinez, F.O. and S. Gordon, *The M1 and M2 paradigm of macrophage activation: time for reassessment*. F1000prime reports, 2014. **6**: p. 13-13.
58. Ikegami, M., *Surfactant catabolism*. Respirology, 2006. **11 Suppl**: p. S24-7.
59. Johansson, J., H. Jornvall, and T. Curstedt, *Human surfactant polypeptide SP-B. Disulfide bridges, C-terminal end, and peptide analysis of the airway form*. FEBS Lett, 1992. **301**(2): p. 165-7.
60. Johansson, J., et al., *Hydrophobic 3.7 kDa surfactant polypeptide: structural characterization of the human and bovine forms*. FEBS Lett, 1988. **232**(1): p. 61-4.
61. Fung, T.S. and D.X. Liu, *Human Coronavirus: Host-Pathogen Interaction*. Annual Review of Microbiology, 2019. **73**(1): p. 529-557.
62. Sisson, T.H. and R.H. Simon, *The plasminogen activation system in lung disease*. Curr Drug Targets, 2007. **8**(9): p. 1016-29.
63. Shimoda, L.A. and G.L. Semenza, *HIF and the lung: role of hypoxia-inducible factors in pulmonary development and disease*. American journal of respiratory and critical care medicine, 2011. **183**(2): p. 152-156.
64. Gouveia, L., C. Betsholtz, and J. Andrae, *Exploring the effect of PDGF-A deletion in the adult lung: implications in homeostasis and injury*.
65. Olajuyin, A.M., X. Zhang, and H.-L. Ji, *Alveolar type 2 progenitor cells for lung injury repair*. Cell Death Discovery, 2019. **5**(1): p. 63.
66. Lei, C.Q., et al., *ECSIT bridges RIG-I-like receptors to VISA in signaling events of innate antiviral responses*. J Innate Immun, 2015. **7**(2): p. 153-64.
67. Veldhuizen, E.J.A. and H.P. Haagsman, *Role of pulmonary surfactant components in surface film formation and dynamics*. Biochimica et Biophysica Acta (BBA) - Biomembranes, 2000. **1467**(2): p. 255-270.
68. Mulugeta, S. and M.F. Beers, *Surfactant protein C: its unique properties and emerging immunomodulatory role in the lung*. Microbes Infect, 2006. **8**(8): p. 2317-23.
69. Nayak, A., et al., *An Insight into the Diverse Roles of Surfactant Proteins, SP-A and SP-D in Innate and Adaptive Immunity*. Frontiers in Immunology, 2012. **3**(131).
70. Sorensen, G.L., *Surfactant Protein D in Respiratory and Non-Respiratory Diseases*. Frontiers in medicine, 2018. **5**: p. 18-18.
71. Mortaz, E., et al., *Epigenetics and chromatin remodeling play a role in lung disease*. Tanaffos, 2011. **10**(4): p. 7-16.
72. Barnes, P.J., *Role of HDAC2 in the pathophysiology of COPD*. Annu Rev Physiol, 2009. **71**: p. 451-64.
73. He, X., et al., *DNA Methyltransferase 1(DNMT1) Promotes Reactive Oxygen Substances(ROS)-mediated Lung Endothelial Cell Apoptosis in Chronic Obstructive Pulmonary Disease*, in A28. ADVANCES IN COPD AND ASTHMA. p. A1200-A1200.

74. Cascella, M., et al., *Features, evaluation and treatment coronavirus (COVID-19)*, in *StatPearls [Internet]*. 2020, StatPearls Publishing.
75. Butler, D.J., et al., *Host, Viral, and Environmental Transcriptome Profiles of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)*. bioRxiv, 2020: p. 2020.04.20.048066.
76. Sungnak, W., et al., *SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes*. Nature Medicine, 2020.
77. Channappanavar, R. and S. Perlman, *Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology*. Semin Immunopathol, 2017. **39**(5): p. 529-539.
78. Pedersen, S.F. and Y.C. Ho, *SARS-CoV-2: a storm is raging*. J Clin Invest, 2020. **130**(5): p. 2202-2205.
79. Schaible, B., K. Schaffer, and C.T. Taylor, *Hypoxia, innate immunity and infection in the lung*. Respir Physiol Neurobiol, 2010. **174**(3): p. 235-43.
80. Jang, J.-H., et al., *Aldehyde dehydrogenase 3A1 protects airway epithelial cells from cigarette smoke-induced DNA damage and cytotoxicity*. Free radical biology & medicine, 2014. **68**: p. 80-86.
81. Faiz, A., et al., *Cigarette smoke exposure decreases CFLAR expression in the bronchial epithelium, augmenting susceptibility for lung epithelial cell death and DAMP release*. Scientific Reports, 2018. **8**(1): p. 12426.
82. Li, L., et al., *Angiopoietin-like 4 Increases Pulmonary Tissue Leakiness and Damage during Influenza Pneumonia*. Cell Rep, 2015. **10**(5): p. 654-663.
83. To, K.K. and L.E. Huang, *Suppression of hypoxia-inducible factor 1 $\alpha$  (HIF-1 $\alpha$ ) transcriptional activity by the HIF prolyl hydroxylase EGLN1*. Journal of Biological Chemistry, 2005. **280**(45): p. 38102-38107.
84. Kron, M., et al., *Altered responses of MeCP2-deficient mouse brain stem to severe hypoxia*. J Neurophysiol, 2011. **105**(6): p. 3067-79.
85. Cavadas, M.A.S., et al., *REST is a hypoxia-responsive transcriptional repressor*. Scientific Reports, 2016. **6**(1): p. 31355.
86. Zhang, J.Z., A. Behrooz, and F. Ismail-Beigi, *Regulation of glucose transport by hypoxia*. Am J Kidney Dis, 1999. **34**(1): p. 189-202.
87. Akman, H.O., et al., *Response to hypoxia involves transforming growth factor-beta2 and Smad proteins in human endothelial cells*. Blood, 2001. **98**(12): p. 3324-31.
88. Sperandio, S., et al., *The transcription factor Egr1 regulates the HIF-1 $\alpha$  gene during hypoxia*. Mol Carcinog, 2009. **48**(1): p. 38-44.
89. Yan, S.-F., et al., *Hypoxia/Hypoxemia-Induced Activation of the Procoagulant Pathways and the Pathogenesis of Ischemia-Associated Thrombosis*. Arteriosclerosis, Thrombosis, and Vascular Biology, 1999. **19**(9): p. 2029-2035.
90. Niedermeyer, J., E. Meissner, and H. Fabel, *Thrombolytic therapy in pulmonary embolism. Indications and therapeutic strategies*. Z Gesamte Inn Med, 1993. **48**(6-7): p. 332-43.
91. Sulemane, S., et al., *Acute pulmonary embolism in conjunction with intramural right ventricular thrombus in a SARS-CoV-2-positive patient*. European Heart Journal - Cardiovascular Imaging, 2020.
92. Li, L., et al., *Sox9 Activation is Essential for the Recovery of Lung Function after Acute Lung Injury*. Cellular Physiology and Biochemistry, 2015. **37**(3): p. 1113-1122.
93. Kasinski, A., et al., *Transcriptional Regulation of YWHAZ, the Gene Encoding 14-3-3 $\zeta$* . PLOS ONE, 2014. **9**(4): p. e93480.

94. Noutsios, G.T., et al., *14-3-3 isoforms bind directly exon B of the 5'-UTR of human surfactant protein A2 mRNA*. American journal of physiology. Lung cellular and molecular physiology, 2015. **309**(2): p. L147-L157.
95. Zhou, Y., et al., *Chitinase 3-like 1 suppresses injury and promotes fibroproliferative responses in Mammalian lung fibrosis*. Science translational medicine, 2014. **6**(240): p. 240ra76-240ra76.
96. Li, R., et al., *Pdgfra marks a cellular lineage with distinct contributions to myofibroblasts in lung maturation and injury response*. eLife, 2018. **7**: p. e36865.
97. Matissek, S.J. and S.F. Elswa, *GLI3: a mediator of genetic diseases, development and cancer*. Cell Communication and Signaling, 2020. **18**(1): p. 54.
98. Saito, A., M. Horie, and T. Nagase, *TGF- $\beta$  Signaling in Lung Health and Disease*. International journal of molecular sciences, 2018. **19**(8): p. 2460.
99. Xu, K., N. Moghal, and S.E. Egan, *Notch Signaling in Lung Development and Disease*, in *Notch Signaling in Embryology and Cancer*, J. Reichrath and S. Reichrath, Editors. 2012, Springer US: New York, NY. p. 89-98.
100. Kulkarni, H.S., et al., *The complement system in the airway epithelium: An overlooked host defense mechanism and therapeutic target?* Journal of Allergy and Clinical Immunology, 2018. **141**(5): p. 1582-1586.e1.
101. Zhu, J., et al., *Cysteinyl leukotriene 1 receptor expression associated with bronchial inflammation in severe exacerbations of COPD*. Chest, 2012. **142**(2): p. 347-357.
102. Cloonan, S.M. and A.M.K. Choi, *Mitochondria in lung disease*. The Journal of clinical investigation, 2016. **126**(3): p. 809-820.
103. Manley, G.C.A., et al., *DUSP10 Negatively Regulates the Inflammatory Response to Rhinovirus through Interleukin-1 $\beta$  Signaling*. Journal of virology, 2019. **93**(2): p. e01659-18.
104. Walther, F.J., L.M. Gordon, and A.J. Waring, *Advances in synthetic lung surfactant protein technology*. Expert Review of Respiratory Medicine, 2019. **13**(6): p. 499-501.
105. Wright, J.R., *Immunoregulatory functions of surfactant proteins*. Nature Reviews Immunology, 2005. **5**(1): p. 58-68.
106. Han, S. and R.K. Mallampalli, *The Role of Surfactant in Lung Disease and Host Defense against Pulmonary Infections*. Annals of the American Thoracic Society, 2015. **12**(5): p. 765-774.
107. Yamano, G., et al., *ABCA3 is a lamellar body membrane protein in human lung alveolar type II cells*. FEBS Lett, 2001. **508**(2): p. 221-5.
108. Skokic, F., et al., *Surfactant Replacement Therapy in Influenza A H1N1*. The Pediatric Infectious Disease Journal, 2010. **29**(4): p. 387.
109. Czyzewski, A.M., et al., *Effective in vivo treatment of acute lung injury with helical, amphipathic peptoid mimics of pulmonary surfactant proteins*. Scientific Reports, 2018. **8**(1): p. 6795.
110. Bales, M.J. and E.M. Timpe, *Respiratory stimulant use in chronic obstructive pulmonary disease*. Ann Pharmacother, 2004. **38**(10): p. 1722-5.
111. Tazawa, R., et al., *Inhaled GM-CSF for Pulmonary Alveolar Proteinosis*. New England Journal of Medicine, 2019. **381**(10): p. 923-932.
112. Kaiser, L., et al., *Impact of Oseltamivir Treatment on Influenza-Related Lower Respiratory Tract Complications and Hospitalizations*. Archives of Internal Medicine, 2003. **163**(14): p. 1667-1672.

## Figure legends

**Figure 1:** Enrichment analysis and comparison between deregulated genes in SARS-CoV, SARS-CoV-2 (NHBE cells) and SARS-CoV-2 (Lung biopsy) infections using **A.** GOBP module, **B.** Bioplanet pathway module, **C.** Reactome pathway module, **D.** DisGeNet module, **E.** HumanCyc module. Selected significant terms are represented in heatmap. Significance of enrichment in terms of adjusted p-value ( $< 0.05$ ) is represented in color coded P-value scale for all heatmaps. Color towards red indicates higher significance and color towards yellow indicates less significance, while grey means non-significant. All enriched terms and detail statistics are in Supplementary file 6.

**Figure 2:** Enrichment analysis and comparison between deregulated genes and the genes of some selected processes in SARS-CoV, SARS-CoV-2 (NHBE cells) and SARS-CoV-2 (Lung biopsy) infections using combined module. Selected significant terms are represented in heatmap in upper panel. Color schemes are similar as Figure 1. All enriched terms and detail statistics are in Supplementary file 6. Lower panel heatmaps presents enriched genes for some selected terms from upper panel enrichment analysis. For individual processes, blue means presence (significantly differentially expressed gene) while grey means absence (not significantly differentially expressed genes for this module for this experimental condition).

**Figure 3:** Schematic representation of lung surfactant metabolism pathway from Reactome pathway database. Color towards yellow indicates upregulation and while blue indicates downregulation.

**Figure 4:** Enrichment analysis and comparison between host proteins interacting SARS-CoV proteins (Pfefferle et al.), SARS-CoV-2 proteins (Srinivasan et al.) and SARS-CoV-2 proteins (Gordon et al.) and the genes of some selected processes using **A.** Bioplanet pathway module, **B.** DisGeNet module, **C.** KEGG pathway module. Selected significant terms are represented in heatmap. All enriched terms and detail statistics are in Supplementary file 6. Color schemes are similar as Figure 1 and Figure 2.

**Figure 5:** Network representing the interactions between genes in response to hypoxia process (combined module genes) along with SARS-CoV-2 proteins (Gordon et al.), and host miRNAs. Hexagon, ellipse, rounded rectangle, octagon represents viral proteins, process related genes, proteins that interacts viral proteins and host miRNAs, respectively. Blunted arrow indicates suppression by miRNAs, dotted arrow pointed with open half-circle indicates downregulated miRNAs failing to modulate host gene, arrowed line pointed with open half-

circle indicates targets of viral proteins, and arrowed line pointed with open diamond indicates transcription factors of a gene.

**Figure 6:** Network representing the interactions between genes in lung development process (combined module genes) along with SARS-CoV-2 proteins (Gordon et al.), and host miRNAs. Legends are similar as Figure 5.

**Figure 7:** Network representing the interactions between genes in respiratory processes (combined module genes) along with SARS-CoV-2 proteins (Gordon et al.), and host miRNAs. Legends are similar as Figure 5.

**Figure 8:** Network representing the interactions between genes in surfactant metabolism along with SARS-CoV-2 proteins (Gordon et al.), and host miRNAs. Legends are similar as Figure 5.

### Supplementary Figure legends

**Supplementary Figure 1:** Venn diagrams for comparing the differences between **A.** Deregulated genes in SARS-CoV, SARS-CoV-2 (NHBE cells), and SARS-CoV-2 (lung biopsy) infections, **B.** upregulated genes in SARS-CoV, downregulated genes in SARS-CoV-2 (NHBE cells), and upregulated genes in SARS-CoV-2 (lung biopsy) infections, **C.** upregulated genes in SARS-CoV, upregulated genes in SARS-CoV-2 (NHBE cells), and downregulated genes in SARS-CoV-2 (lung biopsy) infections, **D.** upregulated genes in SARS-CoV, downregulated genes in SARS-CoV-2 (NHBE cells), and downregulated genes in SARS-CoV-2 (lung biopsy) infections, **E.** upregulated genes in SARS-CoV, upregulated genes in SARS-CoV-2 (NHBE cells), and upregulated genes in SARS-CoV-2 (lung biopsy) infections, **F.** downregulated genes in SARS-CoV, upregulated genes in SARS-CoV-2 (NHBE cells), and upregulated genes in SARS-CoV-2 (lung biopsy) infections, **G.** downregulated genes in SARS-CoV, downregulated genes in SARS-CoV-2 (NHBE cells), and upregulated genes in SARS-CoV-2 (lung biopsy) infections, **H.** downregulated genes in SARS-CoV, upregulated genes in SARS-CoV-2 (NHBE cells), and downregulated genes in SARS-CoV-2 (lung biopsy) infections, **I.** downregulated genes in SARS-CoV, SARS-CoV-2 (NHBE cells), and SARS-CoV-2 (lung biopsy) infections.

**Supplementary Figure 2:** Deregulated genes of selected terms from Figure 1 in SARS-CoV, SARS-CoV-2 (NHBE cells) and SARS-CoV-2 (Lung biopsy) infections. Genes of selected



significant terms are represented here. For individual processes, blue means presence (differentially expressed gene of the module term) while grey means absence (not differentially expressed in the experimental condition in that module term). Processes in orange, green, purple, red, cyan color background represent Bioplanet, HumanCyc, GOBP, Reactome, DisGeNet enriched terms, respectively.

**Supplementary Figure 3:** Deregulated genes of additional terms from combined module based enrichment in SARS-CoV, SARS-CoV-2 (NHBE cells) and SARS-CoV-2 (Lung biopsy) infections. Genes of selected significant terms are represented here. For individual processes, blue means presence while grey means absence (color code as in Supplementary Figure 2).

**Supplementary Figure 4:** Expression profiles of genes in different lung associated processes obtained from enrichment analysis with combined module. Color towards red indicates more differential upregulation while color towards green indicates more differential downregulation; yellow color suggesting less expression change compared to control.

**Supplementary Figure 5:** Network representing the interactions between genes in surfactant metabolism along with SARS-CoV-2 proteins (Srinivasan et al.), and host miRNAs. Legends are similar as Figure 5.

**Supplementary Figure 7:** Expression profiles of SARS-CoV-2 receptor/associated proteins for entry into the target cell in **A**. Different organs of normal human, **B**. SARS-CoV-2, SARS-CoV infections. For FPKM scale, color towards blue means higher expression while color towards grey indicates lower level of expression. For  $\text{Log}_2$  fold change scale, color towards red indicates upregulation while color towards green indicates downregulation; yellow color suggesting no significant expression change.

### List of supplementary files

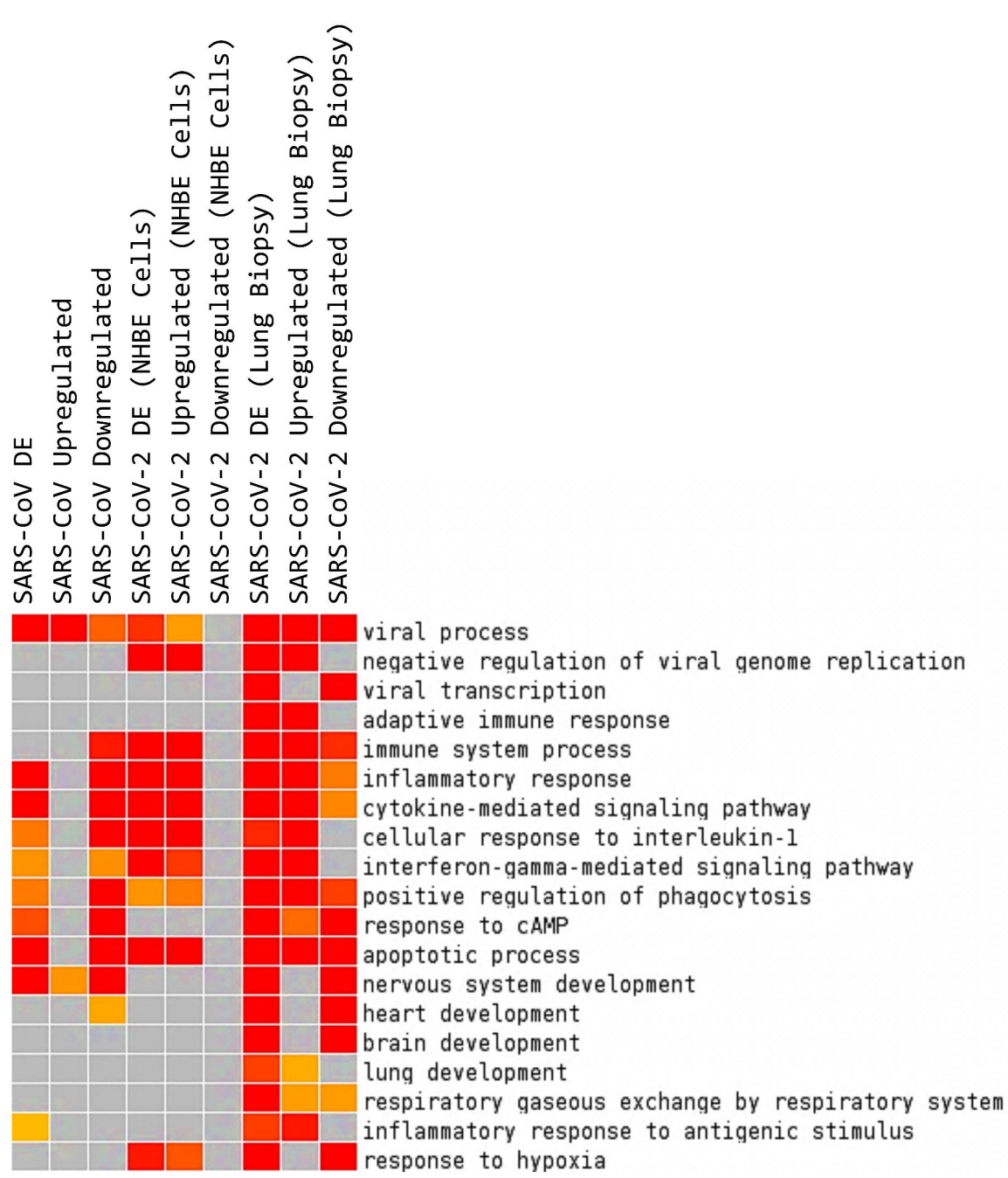
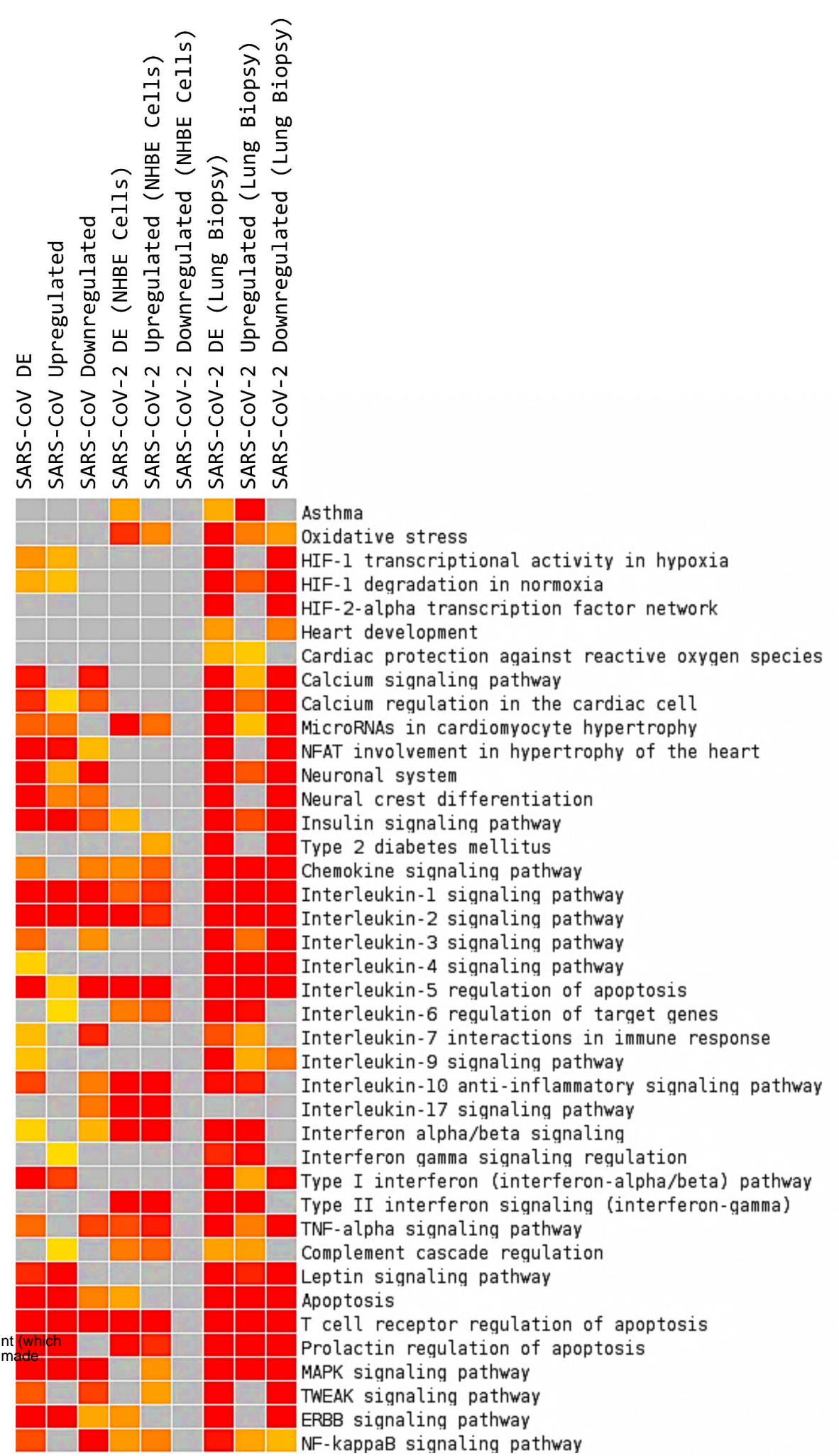
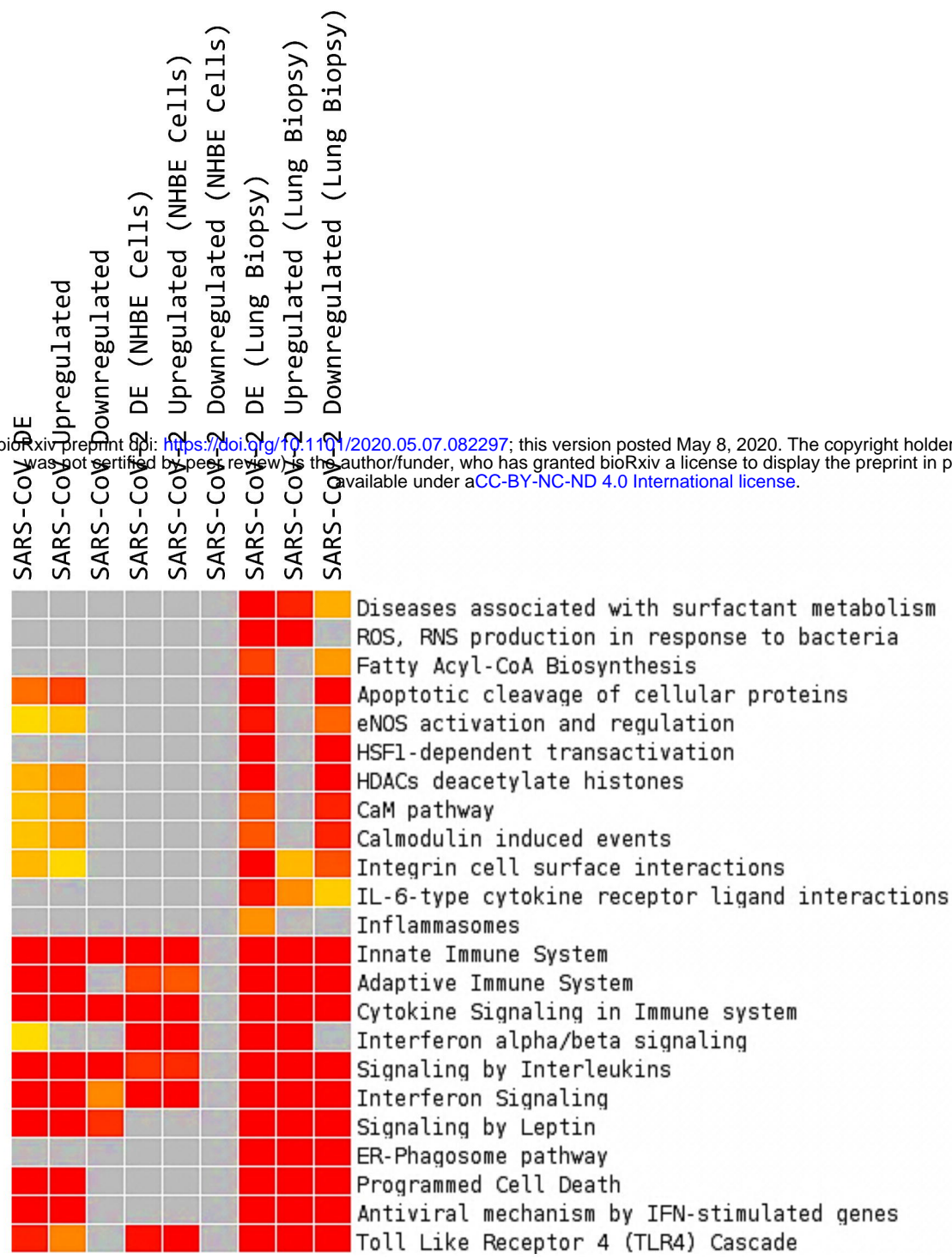
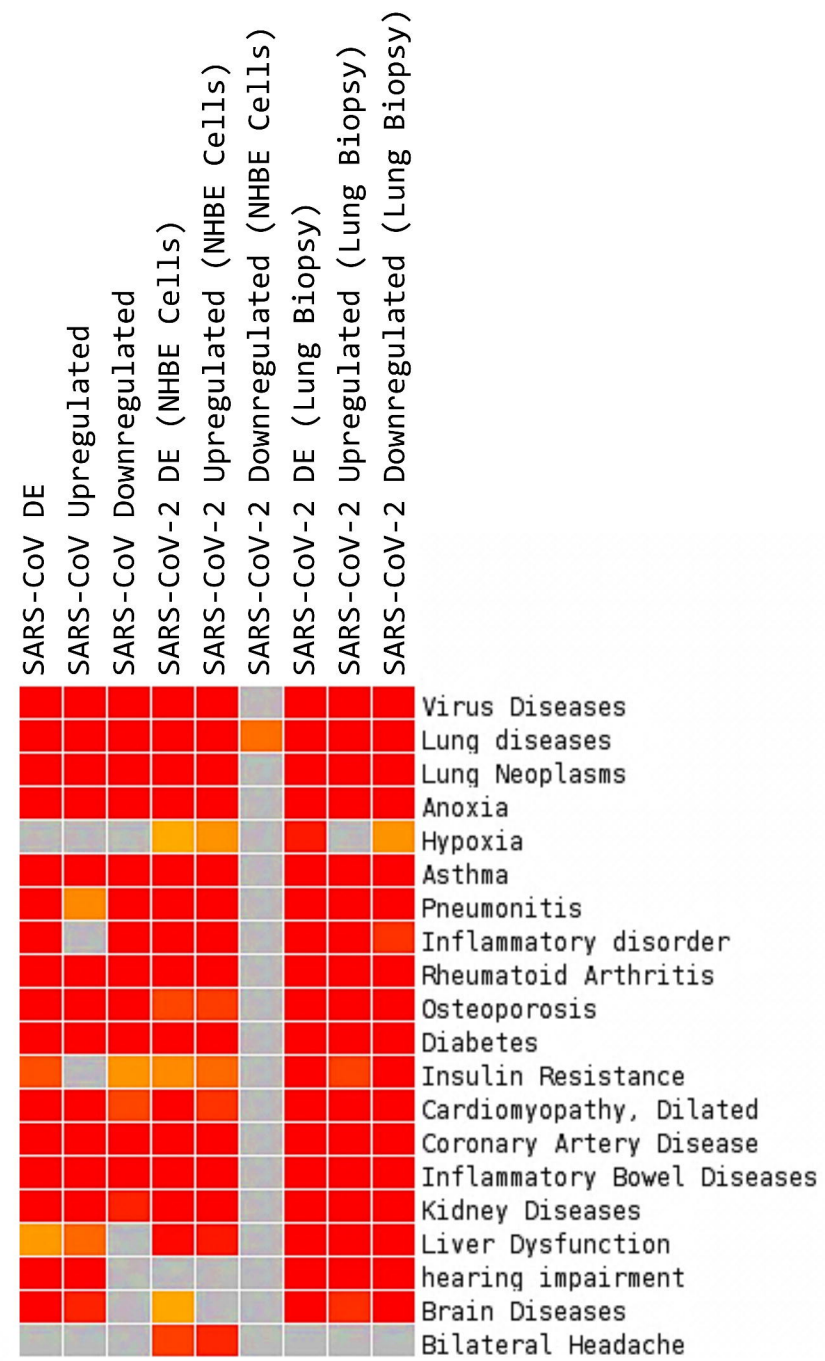
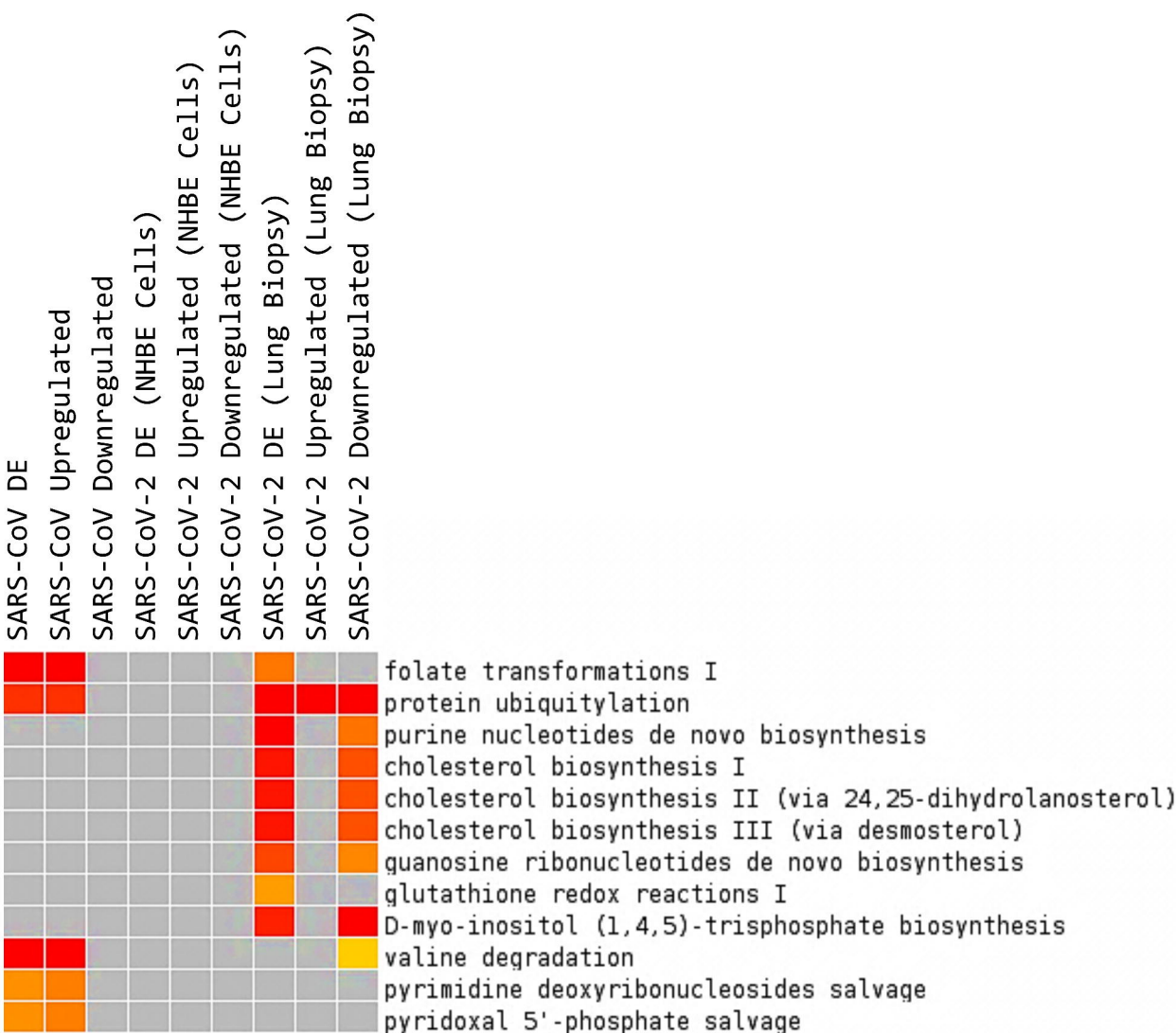
**Supplementary file 1:**  $\text{Log}_2$  Normalized expression data of control and SARS-CoV infection (GEO accession: GSE17400).

**Supplementary file 2:** Raw read counts expression of RNA-seq data (GEO accession: GSE147507) of control and SARS-CoV-2 infection.

**Supplementary file 3:** Quality analysis results of processed normalized RNA-seq data using “arrayQualityMetrics”.

**Supplementary file 4:** List of the combined modules along with the genes and their associated original term and sources.

**Supplementary file 5:** List of host proteins which interact with SARS-CoV, SARS-CoV-2 proteins.

**A.****B.****C.****D.****E.**

P-value for enrichment

0.0

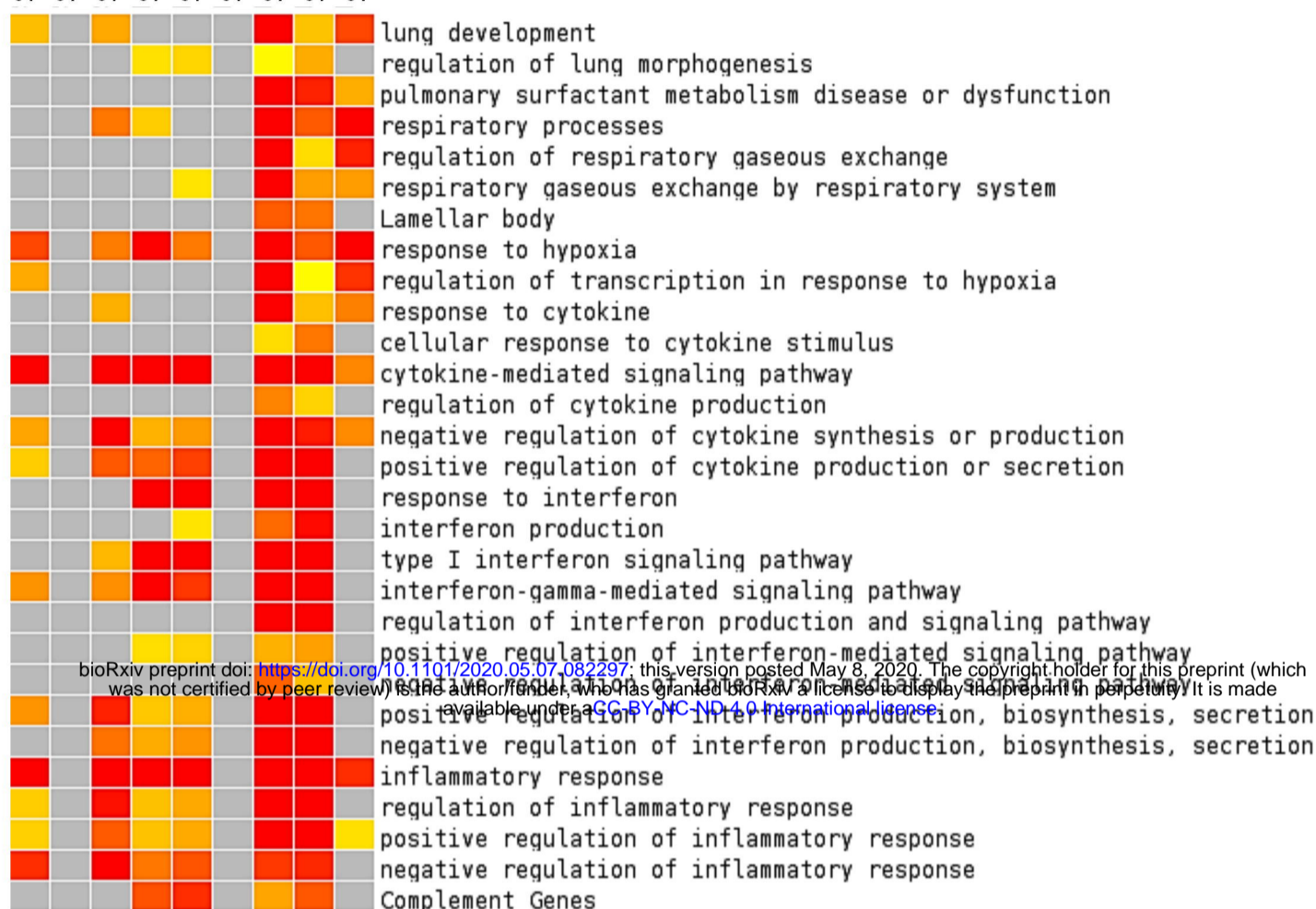
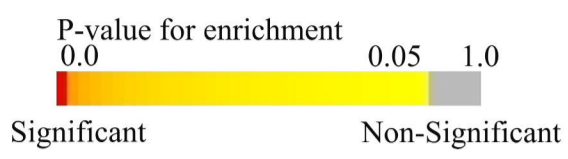
0.05

1.0

Significant

Non-Significant

SARS-CoV DE  
 SARS-CoV Upregulated  
 SARS-CoV Downregulated  
 SARS-CoV-2 DE (NHBE Cells)  
 SARS-CoV-2 Upregulated (NHBE Cells)  
 SARS-CoV-2 Downregulated (NHBE Cells)  
 SARS-CoV-2 DE (Lung Biopsy)  
 SARS-CoV-2 Upregulated (Lung Biopsy)  
 SARS-CoV-2 Downregulated (Lung Biopsy)



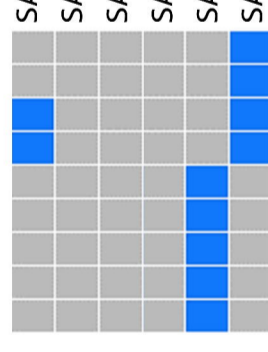
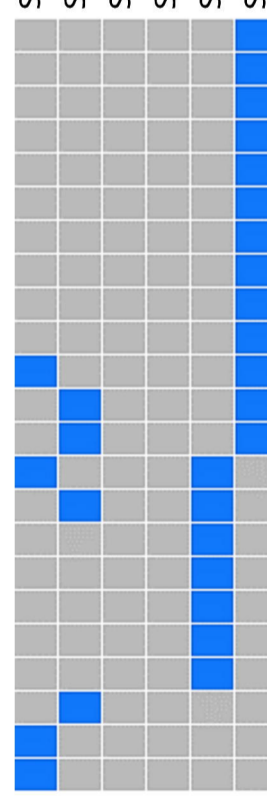
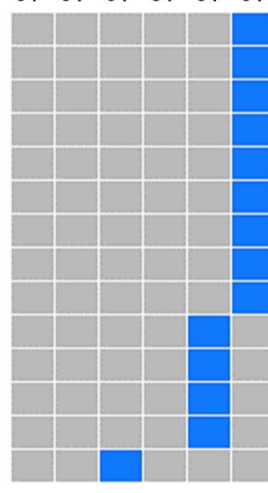
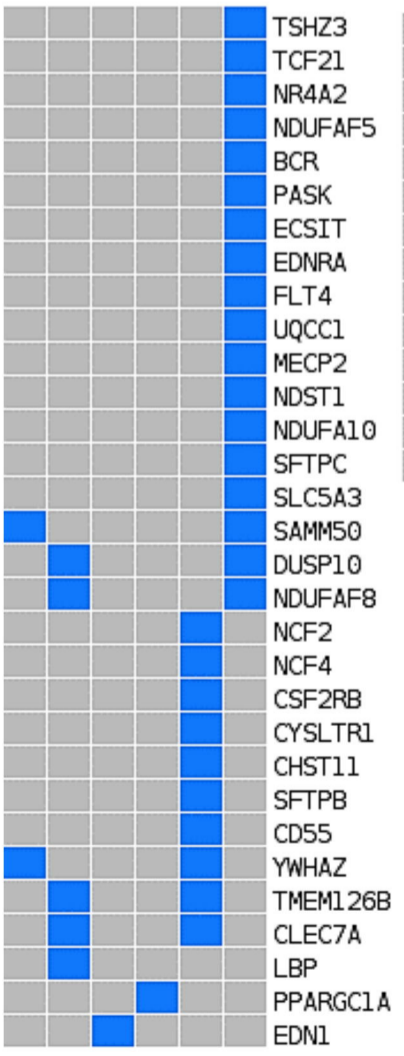
bioRxiv preprint doi: <https://doi.org/10.1101/2020.05.07.082297>; this version posted May 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SARS-CoV Upregulated  
 SARS-CoV Downregulated  
 SARS-CoV-2 Upregulated (NHBE Cells)  
 SARS-CoV-2 Downregulated (NHBE Cells)  
 SARS-CoV-2 Upregulated (Lung Biopsy)  
 SARS-CoV-2 Downregulated (Lung Biopsy)

SARS-CoV Upregulated  
 SARS-CoV Downregulated  
 SARS-CoV-2 Upregulated (NHBE Cells)  
 SARS-CoV-2 Downregulated (NHBE Cells)  
 SARS-CoV-2 Upregulated (Lung Biopsy)  
 SARS-CoV-2 Downregulated (Lung Biopsy)

SARS-CoV Upregulated  
 SARS-CoV Downregulated  
 SARS-CoV-2 Upregulated (NHBE Cells)  
 SARS-CoV-2 Downregulated (NHBE Cells)  
 SARS-CoV-2 Upregulated (Lung Biopsy)  
 SARS-CoV-2 Downregulated (Lung Biopsy)

SARS-CoV Upregulated  
 SARS-CoV Downregulated  
 SARS-CoV-2 Upregulated (NHBE Cells)  
 SARS-CoV-2 Downregulated (NHBE Cells)  
 SARS-CoV-2 Upregulated (Lung Biopsy)  
 SARS-CoV-2 Downregulated (Lung Biopsy)

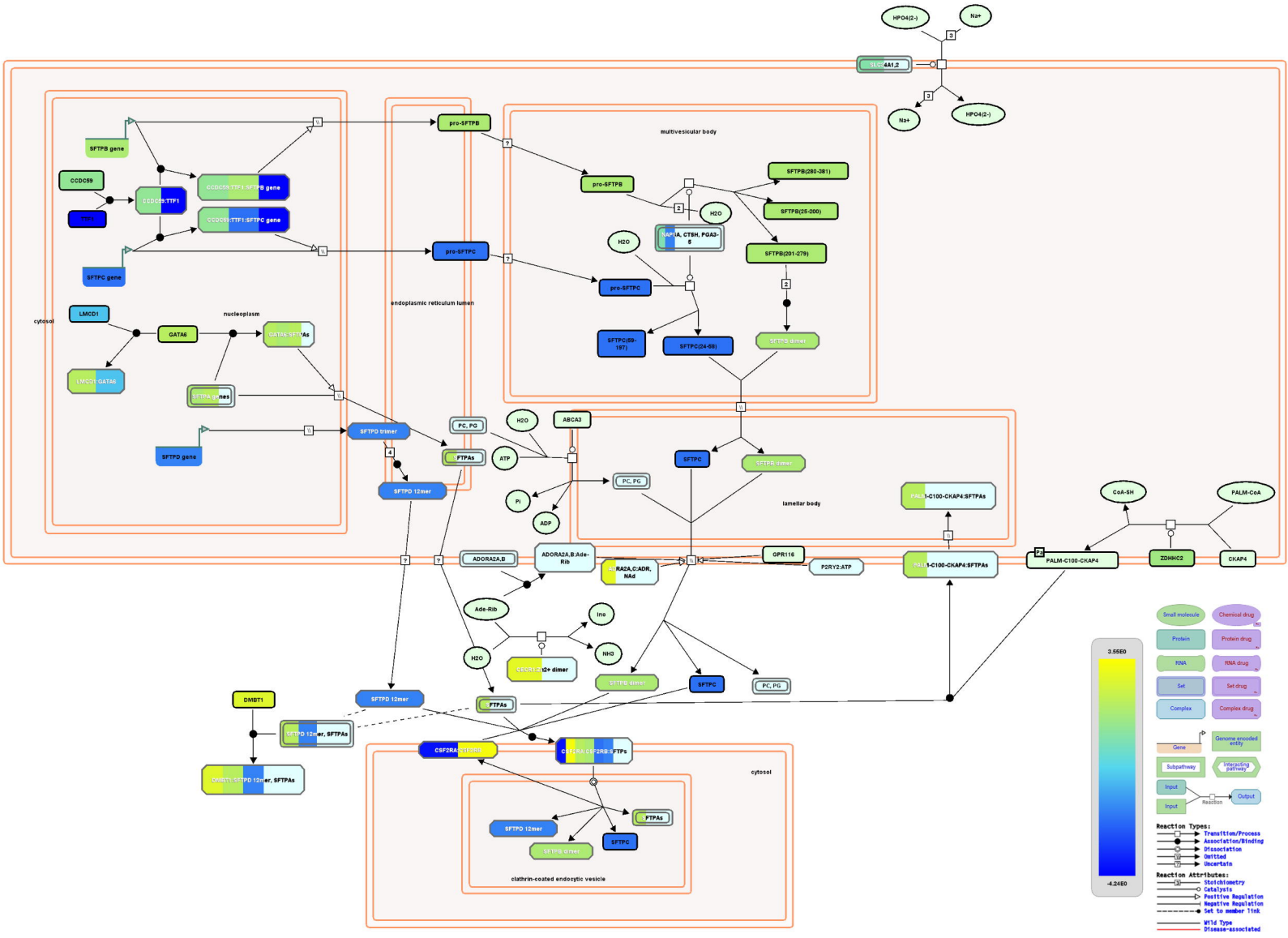


Pulmonary surfactant metabolism disease or dysfunction

Present Absent

Respiratory processes

Lung Development



**Small molecule** (purple oval)

**Protein** (blue rectangle)

**RNA** (green rectangle)

**Set** (yellow rectangle)

**Complex** (orange rectangle)

**Gene** (green rounded rectangle)

**Subpathway** (blue rounded rectangle)

**Input** (blue arrow)

**Output** (blue arrow)

**Reaction** (square)

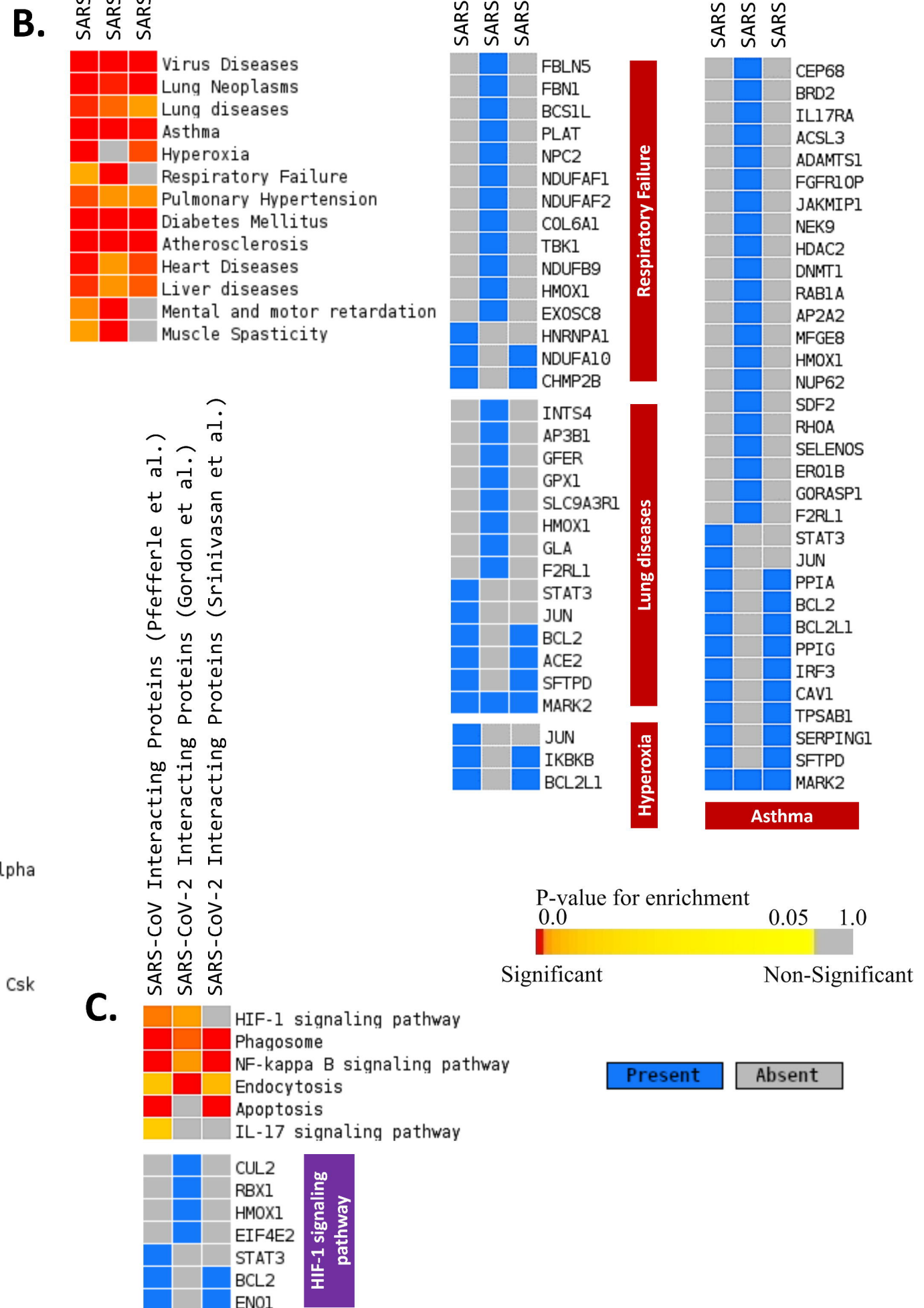
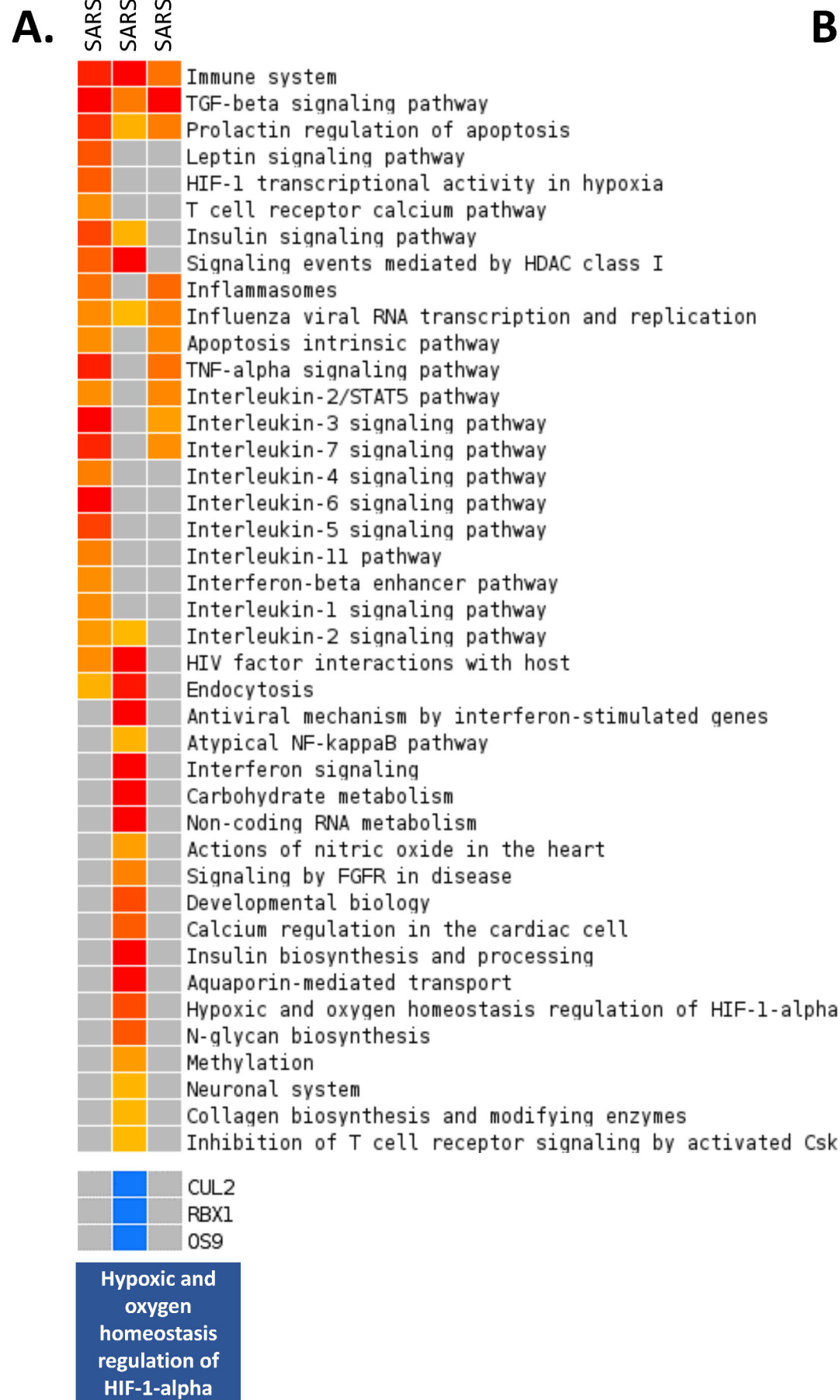
**Reaction Types:**

- Transition/Process
- Association/Binding
- Dissociation
- Outflow
- Inflow

**Reaction Attributes:**

- Stoichiometry
- Catalysis
- Positive Regulation
- Negative Regulation
- Set to number 1
- WT1 Type
- Disease-associated





P-value for enrichment  
0.0 0.05 1.0

Significant Non-Significant

Present Absent

