Constructing a multiple-layer interactome for SARS-CoV-2 in the context of lung disease: Linking the virus with human genes and co-infecting microbes

Shaoke Lou, Tianxiao Li, Mark Gerstein

Abstract:

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has caused millions of deaths worldwide. Many efforts have focused on unraveling the mechanism of the viral infection to develop effective strategies for treatment and prevention. Previous studies have provided some clarity on the protein-protein interaction linkages occurring during the life cycle of viral infection; however, we lack a complete understanding of the full interactome, comprising human miRNAs and protein-coding genes and co-infecting microbes. To comprehensively determine this, we developed a statistical modeling method using latent Dirichlet allocation (called MLCrosstalk, for multiple-layer crosstalk) to fuse many types of data to construct the full interactome of SARS-CoV-2. Specifically, MLCrosstalk is able to integrate samples with multiple layers of information (e.g., miRNA and microbes), enforce a consistent topic distribution on all data types, and infer individual-level linkages (i.e., differing between patients). We also implement a secondary refinement with network propagation to allow our microbe-gene linkages to address larger network structures (e.g., pathways). Using MLCrosstalk, we generated a list of genes and microbes linked to SARS-CoV-2. Interestingly, we found that two of the identified microbes, *Rothia mucilaginosa* and *Prevotella melaninogenica*, show distinct patterns representing synergistic and antagonistic relationships with the virus, respectively. We also identified several SARS-CoV-2-associated pathways, including the VEGFA-VEGFR2 and immune response pathways, which may provide potential targets for drug design.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused one of deadliest pandemics in human history, infecting more than 175 million people and causing 3.7 million deaths (WHO, Jun 2021). Thus, there is an urgent need to understand the mechanisms governing viral infection and host responses in order to develop effective methods for diagnosis, treatment, and quarantine.

Based on in-depth investigations of the SARS-CoV-2 genome and transcriptome structure (Kim, Lee et al. 2020, Wu, Zhao et al. 2020), researchers have elucidated the SARS-CoV-2 infection pathway (Chen, Malone et al. 2020, Viswanathan, Arya et al. 2020), (Kim, Lee et al. 2020, Shu, Huang et al. 2020, Yin, Mao et al. 2020), (Schoeman and Fielding 2019) and have identified several infection-related pathways within the human host (Ho, Mok et al. 2021). These efforts have revealed key processes in SARS-CoV-2 infection and serve as the cornerstone for further large-scale regulatory network and biosignature studies.

High-throughput methods have also elucidated interactions between SARS-CoV-2 and the host, shedding light on the host protein/virus protein interaction network (Gordon, Hiatt et al. 2020, Gordon, Jang et al. 2020), perturbations in the host gene and cellular networks during the initial stages of SARS-CoV-2 infection (similar to the triggering of cytokine storms) (Li, Guo et al. 2021), and interactions between host proteins and SARS-CoV-2 RNA during active infection (Flynn, Belk et al. 2021). Single-cell RNA sequencing has also provided valuable information regarding biological pathways and biosignatures (Zhang, Wang et al. 2020, Ng, Granados et al.

2021) and has revealed the large-scale cellular and molecular landscape of immune responses during SARS-CoV-2 infection in multiple tissues (Guo, Li et al. 2020, Zhang, Wang et al. 2020, Ren, Wen et al. 2021).

Despite recent achievements, our current understanding remains insufficient to provide an integrative picture of the virus-host interaction. For example, research has shown that microRNAs (miRNAs) play an important role in antiviral immune responses (Mirzaei, Mahdavi et al. 2021) and participate in the host response to SARS-CoV-2 (Arisan, Dart et al. 2020, Jafarinejad-Farsangi, Jazi et al. 2020), and potential miRNA binding sites occur on the SARS-CoV-2 genome (Jafarinejad-Farsangi, Jazi et al. 2020). Moreover, Previous studies have shown that COVID-19 patients usually present with co-infection (Lansbury, Lim et al. 2020). The co-infecting microbes may have similar host responses and pathogenic effects and form co-abundant clusters. Numerous other microbes play an indispensable role in shaping the host immune response, but their effects on SARS-CoV-2 infection remain largely unknown. Also, a recent study reported differential bacterial compositions between COVID-19 patients and healthy controls, with COVID-19 patients having a lower diversity (Xu, Lu et al. 2021). This finding indicates that unknown interactions occur between upper respiratory and gut microbiomes during SARS-CoV-2 infection, suggesting their potential importance in the host response.

Overall, it is challenging to integrate these multiple layers of information, including gene, miRNA, and other microbes' information, and to identify inter-layer associations relevant to the host response in SARS-CoV-2 infection. In this paper, we propose an advanced option, MLCrosstalk, for elucidating host-pathogen interactions. MLCrosstalk incorporates multiple data resources and features and identifies both common and COVID-19-specific host gene-microbiome interactomes in different tissues across human diseases. Using network propagation analysis, we further extend our study to the flow of host responses.

Results

The MLCrosstalk model and validation

Our MLCrosstalk has three major advantages for performing integration analysis of multiple-type data. This approach 1) takes advantage of the Dirichlet distribution of hyperparameter to handle sparse and noisy data, 2) enforces a unitary topic distribution for each patient/sample to facilitate linkage identification between different types of data, and 3) easily extends to multiple data types with missing samples allowed (see Figure 1 for workflow). In our study of COVID-19 datasets, MLCrosstalk extracted dimensionally reduced patterns to infer a comprehensive linkage between host protein-coding genes, noncoding genes (e.g., miRNA), and microbes. MLCrosstalk defined a comprehensive interactome for the gene-microbe-miRNA network. The network then underwent further refinement via network propagation to integrate pathway information and connect host-pathogen interactions with biological relevance.

We first evaluated the trained model by exploring the clustering of topic distributions for the samples. The clustering results indicate that the topic distribution can capture the patterns across disease and tissue groups. The trained model groups most of the COVID-19, healthy, and community-acquired pneumonia (CAP) individuals into distinct

clusters. Using the Kullback-Leibler divergence between topic distributions in comparison to a random background, thus, we found that Topic 9 differs the most from the random background distribution and is the most interesting topic.

We then annotated the functions of the top-weighted genes from Topic 9 using multiple known gene sets, including the Kyoto Encyclopedia of Genes and Genomes (KEGG), WikiPathways, virus-host protein-protein interaction (PPI), and COVID-19-related gene sets. We found that these genes are highly enriched in immune-related pathways and heat-shock response proteins (Fig. 2B). Figure 2C-D shows the top-weighted protein-coding genes, miRNAs, and microbes, with SARS-CoV-2 being one of the strongest contributors for Topic 9.

SARS-Cov-2 links to Microbes

SARS-CoV-2 was the one of most common microbes among the COVID-19 patient samples. We investigated and compared the microbes that may be associated with SARS-CoV-2 based on the similarity from raw abundance, advanced feature weights (topic weights), and target functional linkage correlation. From the raw abundance of microbes, we took the top 100 most abundant microbes, and identified microbe groups based on the abundance.

We then used a different approach to identify COVID-19-associated microbes. The microbe communities identified by raw abundance could not be overlapped with the correlation of reduced signature weights and function (Fig. 3B). We confirmed that *Escherichia coli*, *Enterobacter cloacae* complex, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus* are highly associated with COVID-19, although some, especially those that are Gram-negative, have been found to be commonly hospital-acquired in many contexts (Asmarawati, Rosyid et al. 2021, Baskaran, Lawrence et al. 2021).). While *Rothia M.*, *Fusobacterium periodonticum*, *Prevotella melaninogenica*, and *Haemophilus parainfluenzae*, as well-known pathogens, were also found to link with COVID-19.

As shown in Figure 4, these COVID-19-associated microbes have with similar interaction profiles in the COVID-19 patient group (Fig. 4A). but they demonstrate different interacting patterns in the healthy patient group (Fig. 4B). *Rothia M.*, *Fusobacterium periodonticum*, *Prevotella melaninogenica*, and *Haemophilus parainfluenzae* show significant composition changes in COVID-19 patients compared with healthy people in bronchoalveolar lavage fluid (BALF) (Fig. 4C).

Rothia mucilaginosa is a gram-positive coccus that occurs as part of the normal flora of the oropharynx and upper respiratory tract (Baeza Martinez, Zamora Molina et al. 2014, Yang, Liu et al. 2020), has a significantly higher relative abundance in COVID-19 patients and showed a synergic effect with COVID-19. The specific linked genes of *Rothia M.* in COVID-19 patients compared with healthy individuals are highly enriched in the immune response, host-pathogen interaction, and SARS-CoV-2-related gene sets (Fig. 4D). However, when comparing the linked genes of *Rothia M.* between the COVID patient group and the CAP group, the significant genes are more related to cell adhesion. When

comparing COVID-19 with pneumonia, the cell barriers are more important for viral infections (Fig. 4s).

Fusobacterium periodonticum, *Prevotella melaninogenica*, and *Haemophilus parainfluenzae*, show a reduced relative abundance in COVID-19 patients. In particular, the significant high composition of *Prevotella M.* in healthy individuals also results in specific interacting genes, which are shown in Figure 4E. RELA is a subunit of the NF-Kappa B complex, which forms the RELA-RELA complex and is involved in invasion-mediated activation of IL-8 expression and regulating the IFN response during SARS-CoV-2 infection. Another important pathway that shows significant high enrichment is the Notch signaling pathway (including genes PSEN1, NOTCH4, and HDAC2), which is significantly upregulated in the lungs in COVID-19 infections (Rosa, Ahmed et al. 2021). These genes and pathways linked to *Prevotella M.* are also key factors for COVID-19 infection, considering the composition of *Prevotella M.* is significantly decreased in COVID-19 patients. Thus, *Prevotella M.* may potentially have an antagonistic relationship with COVID-19.

SARS-Cov-2 links to Genes & Pathways

We compared the linkages across 10 different SARS-CoV-2 patient samples, which included BALF, bowel, heart, jejunum, kidney, liver, lung, marrow, peripheral blood mononuclear cells, and placenta. We compared host gene-microbe and microbe-miRNA linkages. In addition to common linkages, we identified signature linkages (such as gene-microbe linkages) that are found only in one specific tissue. The genes, microbes, and miRNAs from these signature linkages are clustered in Figure 6, which shows clear tissue-specific patterns, in particular for BALF and lung tissue (Fig. 5s).

We inferred the COVID-19-specific linked genes by comparing their occurrences in the COVID-19 and healthy groups. More genes are highly represented in the COVID-19 group than those in the healthy group. The top-ranked genes and microbes associated with COVID-19 are shown in Figure 5. These genes come from many pathways. Among these genes, cytoplasmic ribosomal protein and VEGFA-VEGFR2 are significant highly associated with COVID-19 (Fig. 5 A,B). Ribosomal RNA is essential for protein synthesis in all living organisms (Schmidt, Lareau et al. 2021) and the VEGFA-VEGFR2 pathway is highly associated with viral entry.

We applied network propagation to aggregate network information. After converging the random walk with restart (RWR), we identified the top-ranked pathway with a high proportion of signature genes and gene-gene connections. The top-ranked pathway network is shown in Figure 5D. Interestingly, we consistently identified the VEGFA-VEGFR2 pathway and immune response pathway. We also identified some cancer-related pathways, which may be because COVID-19 can trigger signaling pathways that respond to cancer.

ACE2 plays important roles in the entry of COVID-19 virus to cells. Our MLCrosstalk method identified genes in the pathway related to viral entry associated with COVID-19, including IFNAR1, IFNAR2, and STAT. Our results also verified that random walk generates stable results that can recall the most biologically relevant linked targets (Fig. 5F, like ACE2, TMPRSS2) after optimization using RWR. A relatively stable number of patients with the same background showed the same gene links with COVID-19.

Microbes & Genes together: How co-infecting microbes can differentially affect gene-gene linkages

As no SARS-CoV-2 virus can be detected in healthy patients, it is difficult to compare the effects of viral infections using SARS-CoV-2 gene linkage information. We have identified two distinct patterns of the co-infection microbes *Rothia M.* and *Prevotella M.*, which have putative synergic and antagonistic effects, respectively. We studied the enrichment of the linked genes and gene-gene connections for a pathway in different individual groups (SARS-CoV-2, healthy, and CAP) from a pathway. We found that the *Rothia M.* linked genes have significant higher representations in the SARS-CoV-2 patient groups (edges in red) in “Nsp1 inhibits translation initiation in the host cell pathway.” By contrast, *Prevotella M.* showed an opposite pattern in which the SARS-CoV-2 patient group had relatively low representation of the gene-gene edges (in green). We found similar distinct patterns between *Rothia* and *Prevotella* for other pathways, especially for the immune response and signaling pathways (such as the type II interferon signaling pathway and Notch signaling pathway).

We extracted the microbe significantly linked genes for the microbes in pathways and linked two pathways together if they shared significant-gene connections according to the known host networks. We then compared the centrality of a gene-gene connection that joined one pathway with another. *Rothia M.* linked genes had more interconnections among SMAD2/3, TGFBR1, P53, and HMOX1-related pathways, whereas the *Prevotella M.* linked genes were highly enriched in the edge connections among the RUNX1, Notch, and AXIN pathways. These findings suggest that *Rothia M.* and *Prevotella M.* co-infect with SARS-Cov-2 in two very different ways.

Discussion

Here, we present MLCrosstalk, which we specifically developed to tackle three major challenges in integrative data mining: heterogeneity and noisiness of data, multiple-type data integration, and personalized linkage identification. Using the SARS-CoV-2 dataset as an example, we demonstrate the capability of our model to capture latent patterns of multiple types of data. Moreover, we show that the sample-specific linkages inferred by MLCrosstalk have strong support of biological evidence.

MLCrosstalk extends latent Dirichlet allocation (LDA) and is capable of handling noisy and missing data. By enforcing a unified topic distribution, MLCrosstalk controls the sparsity of topics and topic components using the hyperparameters and then builds a

latent representation of multiple data types within the same topic domain. Although there are alternative ways to train each type of data with the LDA model separately, a new challenge is finding topic associations between different data types for further linkage inference. In addition, multiple-type data integration can help to identify the biological underpinnings and the comprehensive linkages between different data types.

Many alternative methods can infer the overall association using large cohort datasets. For example, correlation analysis can discover the trends and associations of two features by considering the values of a set of samples. However, these methods cannot give sample-specific associations, which are important, especially for patient-specific or tissue-specific samples. MLCrosstalk infers sample-based linkages by adding the effect of sample topic distribution to adjust the association inference to consider the different host responses from personalized background/conditions.

COVID-19 is one of the most severe public health emergencies in recent years. It is thus imperative to explore the underlying biology and provide insights for the development of treatment strategies. Our MLCrosstalk method can integrate multiple data types and learn the intrinsic latent patterns in an unsupervised manner. From the most informative Topic 9, the biologically relevant genes and pathogens ranked high in the component of Topic 9. MLCrosstalk then inferred the linkages between genes and microbes and integrated pathway information using network propagations. From the results, we identified SARS-CoV-2 co-infection clusters, and *Rothia M.* , and *Prevotella M.* as the representatives of two groups of microbes, with synergic and antagonistic effects, respectively. With the in-depth discovery of COVID-19-related genes, we pinpointed genes in the most enriched pathways, like the VEGFA-VEGFR2 pathway, and constructed functional pathway associations based on the target genes.

Methods

Data collection and processing

We collected data from CAP, COVID-19, and healthy individuals. The transcriptome data were analyzed using the exceRpt pipeline. Briefly, RNA-seq reads were subjected to quality assessment using FastQC software v.0.10.1 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) both prior to and following 3' adapter clipping. Adapters were removed using FastX v.0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). Identical reads were counted and collapsed to a single entry, and reads containing N's were removed. Clipped, collapsed reads were filtered through the Univec database of common laboratory contaminants and a human ribosomal database before mapping to the human reference genome (hg19) and pre-miRNA sequences using STAR [50]. Reads that did not align were mapped against a ribosomal reference library of bacteria, fungi, and Archaea, compiled by the Ribosome Database Project [51], and then to genomes of bacteria, fungi, plants, and viruses, retrieved from GenBank [51]. In cases where RNA-seq reads aligned equally well to more than one microbe, a "last common ancestor" approach was used, and the read was assigned to the next node up the phylogenetic tree, as performed by similar algorithms [14, 52].

The gene expression, pre-miRNA and exogenous genomic, and rRNA frequency were inferred. The exogenous contents were filtered to remove the potential contamination and to keep only pathogenic microbes. The gene expression of COVID-19, CAP, and healthy individuals were quantile normalized and converted to integers with microbe and miRNA frequency.

MLCrosstalk model

We applied and extended a topic modeling algorithm, which can integrate multiple data types. To make the continuous data work on the topic model, all of the continuous values were converted into integers and scaled to reduce computational intensity.

For any patient group or sample, M (or m) denotes the number of individuals or samples; K (or k) is the number of topics; θ represents the document to topic distribution, or topics; φ denotes the word to topic distribution, or topic component; α, β are the hyper-parameters of the document to topic distribution. The input matrices include gene(g), microbe (b), and (pre)-miRNA (r) abundances, for which each row represents a corresponding sample and each column is a gene, microbe, or miRNA, respectively.

The generative model is shown as below:

Given the topic distribution θ_m for individual m , the topics assigned $z_{m,n}$ are drawn from θ_m for each gene word occurrence from N_1 total genes. Then $w_{m,n}$ are samples from xxx

$$\begin{aligned}
 P(Z, W, \alpha, \beta) &= P(Z^{(g)}, W^{(g)}, Z^{(r)}, W^{(r)}, Z^{(b)}, W^{(b)}; \alpha, \beta) \\
 &= \int_{\theta} \int_{\varphi^{(g)}} \int_{\varphi^{(r)}} \int_{\varphi^{(b)}} P(Z^{(g)}, W^{(g)}, \varphi^{(g)}, Z^{(r)}, W^{(r)}, \varphi^{(r)}, Z^{(b)}, W^{(b)}, \varphi^{(b)}, \theta, \alpha, \beta) d\theta d\varphi^{(g)} d\varphi^{(r)} d\varphi^{(b)} \\
 &= \int \prod_{m=1}^M P(\theta_m; \alpha) \prod_{tg=1}^{N^{(g)}} p(Z_{m,tg}^{(g)} | \theta) \prod_{tr=1}^{N^{(r)}} p(Z_{m,tr}^{(r)} | \theta) \prod_{tb=1}^{N^{(b)}} p(Z_{m,tb}^{(b)} | \theta) d\theta \\
 &\quad \times \int \prod_{k=1}^K P(\varphi_k^{(g)}; \beta) \prod_{m=1}^M \prod_{tg=1}^{N^{(g)}} p(W_{m,tg}^{(g)} | \varphi_{Z_{m,tg}^{(g)}}^{(g)}) d\varphi^{(g)} \\
 &\quad \times \int \prod_{k=1}^K P(\varphi_k^{(r)}; \beta) \prod_{m=1}^M \prod_{tr=1}^{N^{(r)}} p(W_{m,tr}^{(r)} | \varphi_{Z_{m,tr}^{(r)}}^{(r)}) d\varphi^{(r)} \\
 &\quad \times \int \prod_{k=1}^K P(\varphi_k^{(b)}; \beta) \prod_{m=1}^M \prod_{tb=1}^{N^{(b)}} p(W_{m,tb}^{(b)} | \varphi_{Z_{m,tb}^{(b)}}^{(b)}) d\varphi^{(b)} \\
 &= \prod_{m=1}^M \frac{\Delta(n_{m,(.),(.)}^{(g)} + n_{m,(.),(.)}^{(r)} + n_{m,(.),(.)}^{(b)} + \alpha)}{\Delta(\alpha)} \prod_{k=1}^K \frac{\Delta(n_{(.,.),k}^{(g)} + \beta)}{\Delta(\beta)} \prod_{k=1}^K \frac{\Delta(n_{(.,.),k}^{(r)} + \beta)}{\Delta(\beta)} \prod_{k=1}^K \frac{\Delta(n_{(.,.),k}^{(b)} + \beta)}{\Delta(\beta)}
 \end{aligned}$$

I^G, I^B, I^R is the matrix indicator for expression and abundance, where $I = \begin{cases} 1 & \text{if } \text{expr or abundance} > 0 \\ 0 & \text{if } \text{expr or abundance} \leq 0 \end{cases}$ and I is the matrix of #word(gene, microbe or miRNA) by #sample (m).

The linkage $L_{i,j}$ can be defined as

$$L_{i,j;x,y,m} = L(x_i \sim y_j | m) = \frac{\sum_{k=1}^K I_{x,m} I_{y,m} \theta_{m,k} \phi_{i,k}^{(x)} \phi_{j,k}^{(y)}}{\|\phi_{i,\cdot}^{(x)}\| \|\phi_{j,\cdot}^{(y)}\|}, \quad \text{where } x, y \text{ represent gene(G/g), microbe(B/b), and miR (R/r). For example, } \phi_{i,k}^{(g)} \text{ is the topic component of gene } i, \phi_{j,k}^{(b)} \text{ is the topic component of microbe } j, \text{ and the linkage } L_{i,j;m} \\ = (\sum_{k=1}^K I_{m} I_{y,m} \theta_{m,k} \phi_{i,k}^{(g)} \phi_{j,k}^{(b)}) / (\|\phi_{i,k}^{(g)}\| \|\phi_{j,k}^{(b)}\|)$$

To infer a background of $L_{i,j;x,y,m}$, we shuffle the $\phi^{(g)}, \phi^{(b)}$ for each topic k and then calculate the $L_{a,b}^m$ for 1,000 times and use the mean and variance to infer the one-tailed p-value. We then use the FDR adjustment to get a q-value for the inference of linkages for each sample.

Pathway integration and curation

We used the Pathwaycommon v12 all-database version as a base, and then integrated the latest online version of KEGG (July 16, 2021) and Reactome (July 3, 2021) to output all the gene pair lists. We also combined the pathway information from WikiPathways (May 10, 2021) and gene symbols from the HUGO Gene Nomenclature Committee with the gene pair list. Finally, we obtained the gene pair list with pathway information.

Network propagation

We generated a gene-gene interaction map based on the latest version of several PPI databases (KEGG, Reactome, and WikiPathways), in which each node represents a gene or a protein and each edge represents a gene-gene connection or PPI. Then, we applied the RWR algorithm on the network using the q-value of the microbe-gene linkage q-value as the node value.

$$p^{t+1} = (1 - r)Wp^t + rp^0$$

After RWR converged, we identified the top-ranked significant linked genes based on corrected q-value and pathogen link pathway score as $\sqrt{\frac{\#sigGene * \#sigEdge}{\#Gene * \#Edge}}$

Multiple layer analysis of changes of the host response network

Taking the results from RWR of each sample, we can analyze the linkage shown in the subnetwork for each sample. By summarizing each significant gene linkage, we can plot

a subnetwork heatmap as a layer, where each layer represents a different patient group, and within a layer is the combination of all genes appearing in the RWR results, using different colors on each edge to show the frequency in which it appeared in different patient groups.

We set a cutoff value to filter significant genes in each sample, and then used them to generate a subnetwork of selected pathways for each group. By calculating how many times each edge showed up in the subnetwork after filtering, we obtained the frequency in which it appeared in each group (such as COVID-19, healthy, or CAP). Then, we set a factor to compare the edge frequency between groups. If an edge in the CAP or COVID-19 group was a factor times as large than it was in the healthy group, we marked the edge in the CAP or COVID-19 group in red, and the edge in the healthy group in yellow. If an edge in the CAP or COVID-19 group was a factor times as small than it was in the healthy group, we marked the edge in the CAP or COVID-19 group in green, and the edge in the healthy group in yellow. If a node, which represents a gene, had a related edge that was significantly larger or smaller, we marked this node in black.

Figures

Figure 1. MLCrosstalk workflow. We transform gene expression, microbe abundance, and (pre)miRNA expression data, which are then input into the MLCrosstalk model. After training, we apply network propagation to refine the linkages. Multiple layer comparison and network tracing can identify shared and specific pathways and connections.

Figure 2. Model evaluation and functional analysis. A. Heatmap of the topic distribution across all SARS-CoV-2 samples. B. Functional analysis of the most interesting topic (Topic 9). C–D. The top-weighted protein-coding genes, pre-miRNAs, and microbes for Topic 9.

Figure 3. Microbe co-infection network analysis. A. Heatmap of the most abundant microbes including SARS-CoV-2. B) Co-infection of microbes with COVID-19 based on different metrics: raw abundance scaled, correlation of microbe to topic weight, correlation of interacting gene profile based on FDR, correlation of interacting gene profile based on RWR.

Figure 4. Microbe functional associations in the COVID-19 patient group. A) Heatmap of SARS-CoV-2-microbe clusters in the COVID-19 patient group. B) Heatmap of co-infection microbes in healthy patients. C) Composition of top associated microbes between COVID-19 and healthy patients. D) Functional enrichment of Rothia M. COVID-19-specific linked genes. E) Functional enrichment of Prevotella M. healthy-specific genes. F) Rothia-associated miRNA.

Figure 5. COVID-19 links. A) UMAP analysis of gene enrichment. B) Top 127 genes with highlighted enriched pathways. C) Top-ranked pathways from network propagation. D) COVID-19-associated miRNA.

Figure 5s. Linkage comparison across different tissues. A. Gene clusters in tissue-specific gene-microbe linkages. B. Microbe clusters in tissue-specific gene-microbe linkages. C. miRNA clusters in tissue-specific miRNA-microbe linkages. D. Microbe clusters in tissue-specific miRNA-microbe linkages.

Figure 6. Multiple-layer comparison of co-infecting microbes potentially affect gene-gene linkages differentially. A,C) Rothia M. shows increased gene-gene linkages in the Nsp1 pathway and the type II interferon signaling pathway in COVID-19 patients; B,D) Prevotella M. shows decreased gene-gene linkages in the Nsp1 pathway and the Notch signaling pathway in COVID-19 patients; E) Rothia M. shows increased pathway-pathway connections in COVID-19 patients; F) Prevotella M. shows decreased pathway-pathway connections in COVID-19 patients

References

Arisan, E. D., et al. (2020). "The Prediction of miRNAs in SARS-CoV-2 Genomes: hsa-miR Databases Identify 7 Key miRs Linked to Host Responses and Virus Pathogenicity-Related KEGG Pathways Significant for Comorbidities." *Viruses* **12**(6).

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is a member of the betacoronavirus family, which causes COVID-19 disease. SARS-CoV-2 pathogenicity in humans leads to increased mortality rates due to alterations of significant pathways, including some resulting in exacerbated inflammatory responses linked to the "cytokine storm" and extensive lung pathology, as well as being linked to a number of comorbidities. Our current study compared five SARS-CoV-2 sequences from different geographical regions to those from SARS, MERS and two cold viruses, OC43 and 229E, to identify the presence of miR-like sequences. We identified seven key miRs, which highlight considerable differences between the SARS-CoV-2 sequences, compared with the other viruses. The level of conservation between the five SARS-CoV-2 sequences was identical but poor compared with the other sequences, with SARS showing the highest degree of conservation. This decrease in similarity could result in reduced levels of transcriptional control, as well as a change in the physiological effect of the virus and associated host-pathogen responses. MERS and the milder symptom viruses showed greater differences and even significant sequence gaps. This divergence away from the SARS-CoV-2 sequences broadly mirrors the phylogenetic relationships obtained from the whole-genome alignments. Therefore, patterns of mutation, occurring during sequence divergence from the longer established human viruses to the more recent ones, may have led to the emergence of sequence motifs that can be related directly to the pathogenicity of SARS-CoV-2. Importantly, we identified 7 key-microRNAs (miRs 8066, 5197, 3611, 3934-3p, 1307-3p, 3691-3p, 1468-5p) with significant links to KEGG pathways linked to viral pathogenicity and host responses. According to Bioproject data (PRJNA615032), SARS-CoV-2 mediated transcriptomic alterations were similar to the target pathways of the selected 7 miRs identified in our study. This mechanism could have considerable significance in determining the symptom spectrum of future potential pandemics. KEGG pathway analysis revealed a number of critical pathways linked to the seven identified miRs that may provide insight into the interplay between the virus and comorbidities. Based on our reported findings, miRNAs may constitute potential and effective therapeutic approaches in COVID-19 and its pathological consequences.

Asmarawati, T. P., et al. (2021). "The clinical impact of bacterial co-infection among moderate, severe and critically ill COVID-19 patients in the second referral hospital in Surabaya." *F1000Res* **10**: 113.

Background: Data on the prevalence of bacterial co-infections among COVID-19 patients are limited, especially in our country, Indonesia. We aimed to assess the rate of bacterial co-infections in hospitalized COVID-19 patients and report the most common microorganisms involved and the antibiotic use in these patients. Methods: This study is a retrospective cohort study, among COVID-19 adult patients admitted to Universitas Airlangga Hospital Surabaya from 14 March-30 September 2020. The bacterial infection is defined based on clinical assessment, laboratory parameters, and microbiology

results. Results: A total of 218 patients with moderate to critical illness and confirmed COVID-19 were included in this study. Bacterial infection was confirmed in 43 patients (19.7%). COVID-19 patients with bacterial infections had longer hospital length of stay (17.6 +/- 6.62 vs 13.31 +/- 7.12), a higher proportion of respiratory failure, intensive care treatment, and ventilator use. COVID-19 patients with bacterial infection had a worse prognosis than those without bacterial infection ($p < 0.04$). The empirical antibiotic was given to 75.2% of the patients. Gram-negative bacteria were commonly found as causative agents in this study ($n = 39$; 70.37%). Conclusion: COVID-19 patients with bacterial infection have a longer length of stay and worse outcomes. Healthcare-associated infections during intensive care treatment for COVID-19 patients must be carefully prevented.

Baeza Martinez, C., et al. (2014). "Rothia mucilaginosa pneumonia in an immunocompetent patient." *Arch Bronconeumol* **50**(11): 493-495.

Rothia mucilaginosa is a gram-positive coccus that occurs as part of the normal flora of the oropharynx and upper respiratory tract. Lower respiratory tract infections caused by this organism are rare and usually occur in immunocompromised patients. This is the case of an immunocompetent 47-year-old woman with right upper lobe pneumonia in which *R. mucilaginosa* was isolated in sputum and bronchial aspirate. Infections caused by this agent in the last four years in our hospital were reviewed. The most common predisposing factor was COPD with bronchiectasis. *R. mucilaginosa* was identified as the causative agent for pneumonia in only two cases, of which one was our case and the other was a patient with lung cancer.

Baskaran, V., et al. (2021). "Co-infection in critically ill patients with COVID-19: an observational cohort study from England." *J Med Microbiol* **70**(4).

Introduction. During previous viral pandemics, reported co-infection rates and implicated pathogens have varied. In the 1918 influenza pandemic, a large proportion of severe illness and death was complicated by bacterial co-infection, predominantly *Streptococcus pneumoniae* and *Staphylococcus aureus*. **Gap statement.** A better understanding of the incidence of co-infection in patients with COVID-19 infection and the pathogens involved is necessary for effective antimicrobial stewardship. **Aim.** To describe the incidence and nature of co-infection in critically ill adults with COVID-19 infection in England. **Methodology.** A retrospective cohort study of adults with COVID-19 admitted to seven intensive care units (ICUs) in England up to 18 May 2020, was performed. Patients with completed ICU stays were included. The proportion and type of organisms were determined at <48 and >48 h following hospital admission, corresponding to community and hospital-acquired co-infections. **Results.** Of 254 patients studied (median age 59 years (IQR 49-69); 64.6 % male), 139 clinically significant organisms were identified from 83 (32.7 %) patients. Bacterial co-infections/co-colonisation were identified within 48 h of admission in 14 (5.5 %) patients; the commonest pathogens were *Staphylococcus aureus* (four patients) and *Streptococcus pneumoniae* (two patients). The proportion of pathogens detected increased with duration of ICU stay, consisting largely of Gram-negative bacteria, particularly *Klebsiella*

pneumoniae and Escherichia coli. The co-infection/ co-colonisation rate >48 h after admission was 27/1000 person-days (95 % CI 21.3-34.1). Patients with co-infections/ co-colonisation were more likely to die in ICU (crude OR 1.78,95 % CI 1.03-3.08, P=0.04) compared to those without co-infections/ co-colonisation. Conclusion. We found limited evidence for community-acquired bacterial co-infection in hospitalised adults with COVID-19, but a high rate of Gram-negative infection acquired during ICU stay.

Chen, J., et al. (2020). "Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex." *Cell* **182**(6): 1560-1573 e1513.

SARS-CoV-2 is the causative agent of the 2019-2020 pandemic. The SARS-CoV-2 genome is replicated and transcribed by the RNA-dependent RNA polymerase holoenzyme (subunits nsp7/nsp82/nsp12) along with a cast of accessory factors. One of these factors is the nsp13 helicase. Both the holo-RdRp and nsp13 are essential for viral replication and are targets for treating the disease COVID-19. Here we present cryoelectron microscopic structures of the SARS-CoV-2 holo-RdRp with an RNA template product in complex with two molecules of the nsp13 helicase. The Nidovirales order-specific N-terminal domains of each nsp13 interact with the N-terminal extension of each copy of nsp8. One nsp13 also contacts the nsp12 thumb. The structure places the nucleic acid-binding ATPase domains of the helicase directly in front of the replicating-transcribing holo-RdRp, constraining models for nsp13 function. We also observe ADP-Mg(2+) bound in the nsp12 N-terminal nidovirus RdRp-associated nucleotidyltransferase domain, detailing a new pocket for anti-viral therapy development.

Flynn, R. A., et al. (2021). Discovery and functional interrogation of SARS-CoV-2 RNA-host protein interactions.

Graphical abstract Highlights d ChIRP-MS of SARS-CoV-2 RNA identifies viral RNA-host protein interaction networks d Comparative analysis identifies SARS-specific and multi-viral RNA-protein complexes d SARS-CoV-2 interactome-focused CRISPR screens reveal a broad antiviral response d Host mitochondria serve as a general organelle platform for anti-SARS-CoV-2 immunity

Gordon, D. E., et al. (2020). "Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms." *Science* **370**(6521).

The COVID-19 pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a grave threat to public health and the global economy. SARS-CoV-2 is closely related to the more lethal but less transmissible coronaviruses SARS-CoV-1 and Middle East respiratory syndrome coronavirus (MERS-CoV). Here, we have carried out comparative viral-human protein-protein interaction and viral protein localization analyses for all three viruses. Subsequent functional genetic screening identified host factors that functionally impinge on coronavirus proliferation, including Tom70, a mitochondrial chaperone protein that interacts with both SARS-CoV-1 and SARS-CoV-2 ORF9b, an interaction we structurally characterized using cryo-electron microscopy. Combining genetically validated host factors with both COVID-19 patient genetic data

and medical billing records identified molecular mechanisms and potential drug treatments that merit further molecular and clinical study.

Gordon, D. E., et al. (2020). "A SARS-CoV-2 protein interaction map reveals targets for drug repurposing." *Nature* **583**(7816): 459-468.

A newly described coronavirus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is the causative agent of coronavirus disease 2019 (COVID-19), has infected over 2.3 million people, led to the death of more than 160,000 individuals and caused worldwide social and economic disruption(1,2). There are no antiviral drugs with proven clinical efficacy for the treatment of COVID-19, nor are there any vaccines that prevent infection with SARS-CoV-2, and efforts to develop drugs and vaccines are hampered by the limited knowledge of the molecular details of how SARS-CoV-2 infects cells. Here we cloned, tagged and expressed 26 of the 29 SARS-CoV-2 proteins in human cells and identified the human proteins that physically associated with each of the SARS-CoV-2 proteins using affinity-purification mass spectrometry, identifying 332 high-confidence protein-protein interactions between SARS-CoV-2 and human proteins. Among these, we identify 66 druggable human proteins or host factors targeted by 69 compounds (of which, 29 drugs are approved by the US Food and Drug Administration, 12 are in clinical trials and 28 are preclinical compounds). We screened a subset of these in multiple viral assays and found two sets of pharmacological agents that displayed antiviral activity: inhibitors of mRNA translation and predicted regulators of the sigma-1 and sigma-2 receptors. Further studies of these host-factor-targeting agents, including their combination with drugs that directly target viral enzymes, could lead to a therapeutic regimen to treat COVID-19.

Guo, C., et al. (2020). "Single-cell analysis of two severe COVID-19 patients reveals a monocyte-associated and tocilizumab-responding cytokine storm." *Nat Commun* **11**(1): 3924.

Several studies show that the immunosuppressive drugs targeting the interleukin-6 (IL-6) receptor, including tocilizumab, ameliorate lethal inflammatory responses in COVID-19 patients infected with SARS-CoV-2. Here, by employing single-cell analysis of the immune cell composition of two severe-stage COVID-19 patients prior to and following tocilizumab-induced remission, we identify a monocyte subpopulation that contributes to the inflammatory cytokine storms. Furthermore, although tocilizumab treatment attenuates the inflammation, immune cells, including plasma B cells and CD8(+) T cells, still exhibit robust humoral and cellular antiviral immune responses. Thus, in addition to providing a high-dimensional dataset on the immune cell distribution at multiple stages of the COVID-19, our work also provides insights into the therapeutic effects of tocilizumab, and identifies potential target cell populations for treating COVID-19-related cytokine storms.

Ho, J. S. Y., et al. (2021). "TOP1 inhibition therapy protects against SARS-CoV-2-induced lethal inflammation." *Cell* **184**(10): 2618-2632 e2617.

The ongoing pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is currently affecting millions of lives worldwide. Large retrospective

studies indicate that an elevated level of inflammatory cytokines and pro-inflammatory factors are associated with both increased disease severity and mortality. Here, using multidimensional epigenetic, transcriptional, in vitro, and in vivo analyses, we report that topoisomerase 1 (TOP1) inhibition suppresses lethal inflammation induced by SARS-CoV-2. Therapeutic treatment with two doses of topotecan (TPT), an FDA-approved TOP1 inhibitor, suppresses infection-induced inflammation in hamsters. TPT treatment as late as 4 days post-infection reduces morbidity and rescues mortality in a transgenic mouse model. These results support the potential of TOP1 inhibition as an effective host-directed therapy against severe SARS-CoV-2 infection. TPT and its derivatives are inexpensive clinical-grade inhibitors available in most countries. Clinical trials are needed to evaluate the efficacy of repurposing TOP1 inhibitors for severe coronavirus disease 2019 (COVID-19) in humans.

Jafarinejad-Farsangi, S., et al. (2020). "High affinity of host human microRNAs to SARS-CoV-2 genome: An in silico analysis." *Noncoding RNA Res* 5(4): 222-231.

Background: Coronavirus disease 2019 (COVID-19) caused by a novel betacoronavirus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has attracted top health concerns worldwide within a few months after its appearance. Since viruses are highly dependent on the host small RNAs (microRNAs) for their replication and propagation, in this study, top miRNAs targeting SARS-CoV-2 genome and top miRNAs targeting differentially expressed genes (DEGs) in lungs of patients infected with SARS-CoV-2, were predicted. Methods: All human mature miRNA sequences were acquired from miRBase database. MiRanda tool was used to predict the potential human miRNA binding sites on the SARS-CoV-2 genome. EdgeR identified differentially expressed genes (DEGs) in response to SARS-CoV-2 infection from GEO147507 data. Gene Set Enrichment Analysis (GSEA) and DEGs annotation analysis were performed using ToppGene and Metascape tools. Results: 160 miRNAs with a perfect matching in the seed region were identified. Among them, there was 15 miRNAs with more than three binding sites and 12 miRNAs with a free energy binding of -29 kCal/Mol. MiR-29 family had the most binding sites (11 sites) on the SARS-CoV-2 genome. MiR-21 occupied four binding sites and was among the top miRNAs that targeted up-regulated DEGs. In addition to miR-21, miR-16, let-7b, let-7e, and miR-146a were the top miRNAs targeting DEGs. Conclusion: Collectively, more experimental studies especially miRNA-based studies are needed to explore detailed molecular mechanisms of SARS-CoV-2 infection. Moreover, the role of DEGs including STAT1, CCND1, CXCL-10, and MAPKAPK2 in SARS-CoV-2 should be investigated to identify the similarities and differences between SARS-CoV-2 and other respiratory viruses.

Kim, D., et al. (2020). "The Architecture of SARS-CoV-2 Transcriptome." *Cell* 181(4): 914-921 e910.

SARS-CoV-2 is a betacoronavirus responsible for the COVID-19 pandemic. Although the SARS-CoV-2 genome was reported recently, its transcriptomic architecture is unknown. Utilizing two complementary sequencing techniques, we present a high-resolution map of the SARS-CoV-2 transcriptome and epitranscriptome. DNA nanoball sequencing

shows that the transcriptome is highly complex owing to numerous discontinuous transcription events. In addition to the canonical genomic and 9 subgenomic RNAs, SARS-CoV-2 produces transcripts encoding unknown ORFs with fusion, deletion, and/or frameshift. Using nanopore direct RNA sequencing, we further find at least 41 RNA modification sites on viral transcripts, with the most frequent motif, AAGAA. Modified RNAs have shorter poly(A) tails than unmodified RNAs, suggesting a link between the modification and the 3' tail. Functional investigation of the unknown transcripts and RNA modifications discovered in this study will open new directions to our understanding of the life cycle and pathogenicity of SARS-CoV-2.

Lansbury, L., et al. (2020). "Co-infections in people with COVID-19: a systematic review and meta-analysis." *J Infect* **81**(2): 266-275.

OBJECTIVES: In previous influenza pandemics, bacterial co-infections have been a major cause of mortality. We aimed to evaluate the burden of co-infections in patients with COVID-19. **METHODS:** We systematically searched Embase, Medline, Cochrane Library, LILACS and CINAHL for eligible studies published from 1 January 2020 to 17 April 2020. We included patients of all ages, in all settings. The main outcome was the proportion of patients with a bacterial, fungal or viral co-infection. **RESULTS:** Thirty studies including 3834 patients were included. Overall, 7% of hospitalised COVID-19 patients had a bacterial co-infection (95% CI 3-12%, n=2183, I(2)=92.2%). A higher proportion of ICU patients had bacterial co-infections than patients in mixed ward/ICU settings (14%, 95% CI 5-26, I(2)=74.7% versus 4%, 95% CI 1-9, I(2)= 91.7%). The commonest bacteria were *Mycoplasma pneumonia*, *Pseudomonas aeruginosa* and *Haemophilus influenzae*. The pooled proportion with a viral co-infection was 3% (95% CI 1-6, n=1014, I(2)=62.3%), with Respiratory Syncytial Virus and influenza A the commonest. Three studies reported fungal co-infections. **CONCLUSIONS:** A low proportion of COVID-19 patients have a bacterial co-infection; less than in previous influenza pandemics. These findings do not support the routine use of antibiotics in the management of confirmed COVID-19 infection.

Li, J., et al. (2021). "Virus-Host Interactome and Proteomic Survey Reveal Potential Virulence Factors Influencing SARS-CoV-2 Pathogenesis." *Med* **2**: 99-112.e117.

Summary Background The ongoing coronavirus disease (COVID-19) pandemic caused by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) is a global public health concern due to relatively easy person-to-person transmission and the current lack of effective antiviral therapy. However, the exact molecular mechanisms of SARS-CoV-2 pathogenesis remain largely unknown. **Methods** Genome wide screening was used to establish intra-viral and viral-host interactomes. Quantitative proteomics was used to investigate peripheral blood mononuclear cell (PBMC) proteome signature in COVID-19. **Findings** We elucidated 286 host proteins targeted by SARS-CoV-2 and more than 350 host proteins that are significantly perturbed in COVID-19 derived PBMCs. This signature in severe COVID-19 PBMCs reveals significant upregulation of cellular proteins related to neutrophil activation and blood coagulation, as well as downregulation of proteins mediating T cell receptor signaling. From the interactome, we further identified that

non-structural protein 10 interacts with NF-kappa-B-repressing factor (NKRF) to facilitate interleukin-8 (IL-8) induction, which potentially contributes to IL-8-mediated chemotaxis of neutrophils and the overexuberant host inflammatory response observed in COVID-19 patients

Conclusions Our study not only presents a systematic examination of SARS-CoV-2-induced perturbation of host targets and cellular networks but also reveals insights into the mechanisms by which SARS-CoV-2 triggers cytokine storms, representing a powerful resource in the pursuit for therapeutic intervention

Mirzaei, R., et al. (2021). "The emerging role of microRNAs in the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection." *Int Immunopharmacol* **90**: 107204.

The novel coronavirus disease 2019 (COVID-19) pandemic has imposed significant public health problems for the human populations worldwide after the 1918 influenza A virus (IVA) (H1N1) pandemic. Although numerous efforts have been made to unravel the mechanisms underlying the coronavirus, a notable gap remains in our perception of the COVID-19 pathogenesis. The innate and adaptive immune systems have a pivotal role in the fate of viral infections, such as COVID-19 pandemic. MicroRNAs (miRNAs) are known as short noncoding RNA molecules and appear as indispensable governors of almost any cellular means. Several lines of evidence demonstrate that miRNAs participate in essential mechanisms of cell biology, regulation of the immune system, and the onset and progression of numerous types of disorders. The immune responses to viral respiratory infections (VRIs), including influenza virus (IV), respiratory syncytial virus (RSV), and rhinovirus (RV), are correlated with the ectopic expression of miRNAs. Alterations of the miRNA expression in epithelial cells may contribute to the pathogenesis of chronic and acute airway infections. Hence, analyzing the role of these types of nucleotides in antiviral immune responses and the characterization of miRNA target genes might contribute to understanding the mechanisms of the interplay between the host and viruses, and in the future, potentially result in discovering therapeutic strategies for the prevention and treatment of acute COVID-19 infection. In this article, we present a general review of current studies concerning the function of miRNAs in different VRIs, particularly in coronavirus infection, and address all available therapeutic prospects to mitigate the burden of viral infections.

Ng, D. L., et al. (2021). "A diagnostic host response biosignature for COVID-19 from RNA profiling of nasal swabs and blood." *Sci Adv* **7**(6).

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes coronavirus disease-19 (COVID-19), has emerged as the cause of a global pandemic. We used RNA sequencing to analyze 286 nasopharyngeal (NP) swab and 53 whole-blood (WB) samples from 333 patients with COVID-19 and controls. Overall, a muted immune response was observed in COVID-19 relative to other infections (influenza, other seasonal coronaviruses, and bacterial sepsis), with paradoxical down-regulation of several key differentially expressed genes. Hospitalized patients and outpatients exhibited up-regulation of interferon-associated pathways, although heightened and more robust inflammatory responses were observed in hospitalized patients with more clinically severe illness. Two-layer machine learning-based host classifiers consisting of

complete (>1000 genes), medium (<100), and small (<20) gene biomarker panels identified COVID-19 disease with 85.1-86.5% accuracy when benchmarked using an independent test set. SARS-CoV-2 infection has a distinct biosignature that differs between NP swabs and WB and can be leveraged for COVID-19 diagnosis.

Ren, X., et al. (2021). "COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas." *Cell* **184**(7): 1895-1913 e1819.

A dysfunctional immune response in coronavirus disease 2019 (COVID-19) patients is a recurrent theme impacting symptoms and mortality, yet a detailed understanding of pertinent immune cells is not complete. We applied single-cell RNA sequencing to 284 samples from 196 COVID-19 patients and controls and created a comprehensive immune landscape with 1.46 million cells. The large dataset enabled us to identify that different peripheral immune subtype changes are associated with distinct clinical features, including age, sex, severity, and disease stages of COVID-19. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA was found in diverse epithelial and immune cell types, accompanied by dramatic transcriptomic changes within virus-positive cells. Systemic upregulation of S100A8/A9, mainly by megakaryocytes and monocytes in the peripheral blood, may contribute to the cytokine storms frequently observed in severe patients. Our data provide a rich resource for understanding the pathogenesis of and developing effective therapeutic strategies for COVID-19.

Rosa, B. A., et al. (2021). "IFN signaling and neutrophil degranulation transcriptional signatures are induced during SARS-CoV-2 infection." *Commun Biol* **4**(1): 290.

SARS-CoV-2 virus has infected more than 92 million people worldwide resulting in the Coronavirus disease 2019 (COVID-19). Using a rhesus macaque model of SARS-CoV-2 infection, we have characterized the transcriptional signatures induced in the lungs of juvenile and old macaques following infection. Genes associated with Interferon (IFN) signaling, neutrophil degranulation and innate immune pathways are significantly induced in macaque infected lungs, while pathways associated with collagen formation are downregulated, as also seen in lungs of macaques with tuberculosis. In COVID-19, increasing age is a significant risk factor for poor prognosis and increased mortality. Type I IFN and Notch signaling pathways are significantly upregulated in lungs of juvenile infected macaques when compared with old infected macaques. These results are corroborated with increased peripheral neutrophil counts and neutrophil lymphocyte ratio in older individuals with COVID-19 disease. Together, our transcriptomic studies have delineated disease pathways that improve our understanding of the immunopathogenesis of COVID-19.

Schmidt, N., et al. (2021). "The SARS-CoV-2 RNA-protein interactome in infected human cells." *Nat Microbiol* **6**(3): 339-353.

Characterizing the interactions that SARS-CoV-2 viral RNAs make with host cell proteins during infection can improve our understanding of viral RNA functions and the host innate immune response. Using RNA antisense purification and mass spectrometry, we identified up to 104 human proteins that directly and specifically bind to SARS-CoV-2

RNAs in infected human cells. We integrated the SARS-CoV-2 RNA interactome with changes in proteome abundance induced by viral infection and linked interactome proteins to cellular pathways relevant to SARS-CoV-2 infections. We demonstrated by genetic perturbation that cellular nucleic acid-binding protein (CNBP) and La-related protein 1 (LARP1), two of the most strongly enriched viral RNA binders, restrict SARS-CoV-2 replication in infected cells and provide a global map of their direct RNA contact sites. Pharmacological inhibition of three other RNA interactome members, PPIA, ATP1A1, and the ARP2/3 complex, reduced viral replication in two human cell lines. The identification of host dependency factors and defence strategies as presented in this work will improve the design of targeted therapeutics against SARS-CoV-2.

Schoeman, D. and B. C. Fielding (2019). "Coronavirus envelope protein: current knowledge." *Virology* **16**(1): 69.

BACKGROUND: Coronaviruses (CoVs) primarily cause enzootic infections in birds and mammals but, in the last few decades, have shown to be capable of infecting humans as well. The outbreak of severe acute respiratory syndrome (SARS) in 2003 and, more recently, Middle-East respiratory syndrome (MERS) has demonstrated the lethality of CoVs when they cross the species barrier and infect humans. A renewed interest in coronaviral research has led to the discovery of several novel human CoVs and since then much progress has been made in understanding the CoV life cycle. The CoV envelope (E) protein is a small, integral membrane protein involved in several aspects of the virus' life cycle, such as assembly, budding, envelope formation, and pathogenesis. Recent studies have expanded on its structural motifs and topology, its functions as an ion-channelling viroporin, and its interactions with both other CoV proteins and host cell proteins. **MAIN BODY:** This review aims to establish the current knowledge on CoV E by highlighting the recent progress that has been made and comparing it to previous knowledge. It also compares E to other viral proteins of a similar nature to speculate the relevance of these new findings. Good progress has been made but much still remains unknown and this review has identified some gaps in the current knowledge and made suggestions for consideration in future research. **CONCLUSIONS:** The most progress has been made on SARS-CoV E, highlighting specific structural requirements for its functions in the CoV life cycle as well as mechanisms behind its pathogenesis. Data shows that E is involved in critical aspects of the viral life cycle and that CoVs lacking E make promising vaccine candidates. The high mortality rate of certain CoVs, along with their ease of transmission, underpins the need for more research into CoV molecular biology which can aid in the production of effective anti-coronaviral agents for both human CoVs and enzootic CoVs.

Shu, T., et al. (2020). "SARS-Coronavirus-2 Nsp13 Possesses NTPase and RNA Helicase Activities That Can Be Inhibited by Bismuth Salts." *Virology* **35**(3): 321-329.

The ongoing outbreak of Coronavirus Disease 2019 (COVID-19) has become a global public health emergency. SARS-coronavirus-2 (SARS-CoV-2), the causative pathogen of COVID-19, is a positive-sense single-stranded RNA virus belonging to the family Coronaviridae. For RNA viruses, virus-encoded RNA helicases have long been recognized

to play pivotal roles during viral life cycles by facilitating the correct folding and replication of viral RNAs. Here, our studies show that SARS-CoV-2-encoded nonstructural protein 13 (nsp13) possesses the nucleoside triphosphate hydrolase (NTPase) and RNA helicase activities that can hydrolyze all types of NTPs and unwind RNA helices independently of the presence of NTP, and further characterize the biochemical characteristics of these two enzymatic activities associated with SARS-CoV-2 nsp13. Moreover, we found that some bismuth salts could effectively inhibit both the NTPase and RNA helicase activities of SARS-CoV-2 nsp13 in a dose-dependent manner. Thus, our findings demonstrate the NTPase and helicase activities of SARS-CoV-2 nsp13, which may play an important role in SARS-CoV-2 replication and serve as a target for antivirals.

Viswanathan, T., et al. (2020). "Structural basis of RNA cap modification by SARS-CoV-2." Nat Commun **11**(1): 3718.

The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), the causative agent of COVID-19 illness, has caused millions of infections worldwide. In SARS coronaviruses, the non-structural protein 16 (nsp16), in conjunction with nsp10, methylates the 5'-end of virally encoded mRNAs to mimic cellular mRNAs, thus protecting the virus from host innate immune restriction. We report here the high-resolution structure of a ternary complex of SARS-CoV-2 nsp16 and nsp10 in the presence of cognate RNA substrate analogue and methyl donor, S-adenosyl methionine (SAM). The nsp16/nsp10 heterodimer is captured in the act of 2'-O methylation of the ribose sugar of the first nucleotide of SARS-CoV-2 mRNA. We observe large conformational changes associated with substrate binding as the enzyme transitions from a binary to a ternary state. This induced fit model provides mechanistic insights into the 2'-O methylation of the viral mRNA cap. We also discover a distant (25 Å) ligand-binding site unique to SARS-CoV-2, which can alternatively be targeted, in addition to RNA cap and SAM pockets, for antiviral development.

Wu, F., et al. (2020). "A new coronavirus associated with human respiratory disease in China." Nature **579**(7798): 265-269.

Emerging infectious diseases, such as severe acute respiratory syndrome (SARS) and Zika virus disease, present a major threat to public health(1-3). Despite intense research efforts, how, when and where new diseases appear are still a source of considerable uncertainty. A severe respiratory disease was recently reported in Wuhan, Hubei province, China. As of 25 January 2020, at least 1,975 cases had been reported since the first patient was hospitalized on 12 December 2019. Epidemiological investigations have suggested that the outbreak was associated with a seafood market in Wuhan. Here we study a single patient who was a worker at the market and who was admitted to the Central Hospital of Wuhan on 26 December 2019 while experiencing a severe respiratory syndrome that included fever, dizziness and a cough. Metagenomic RNA sequencing(4) of a sample of bronchoalveolar lavage fluid from the patient identified a new RNA virus strain from the family Coronaviridae, which is designated here 'WH-Human 1' coronavirus (and has also been referred to as '2019-nCoV'). Phylogenetic

analysis of the complete viral genome (29,903 nucleotides) revealed that the virus was most closely related (89.1% nucleotide similarity) to a group of SARS-like coronaviruses (genus Betacoronavirus, subgenus Sarbecovirus) that had previously been found in bats in China(5). This outbreak highlights the ongoing ability of viral spill-over from animals to cause severe disease in humans.

Xu, R., et al. (2021). "Temporal association between human upper respiratory and gut bacterial microbiomes during the course of COVID-19 in adults." *Commun Biol* **4**(1): 240.

SARS-CoV-2 is the cause of COVID-19. It infects multiple organs including the respiratory tract and gut. Dynamic changes of regional microbiomes in infected adults are largely unknown. Here, we performed longitudinal analyses of throat and anal swabs from 35 COVID-19 and 19 healthy adult controls, as well as 10 non-COVID-19 patients with other diseases, by 16 S rRNA gene sequencing. The results showed a partitioning of the patients into 3-4 categories based on microbial community types (I-IV) in both sites. The bacterial diversity was lower in COVID-19 patients than healthy controls and decreased gradually from community type I to III/IV. Although the dynamic change of microbiome was complex during COVID-19, a synchronous restoration of both the upper respiratory and gut microbiomes from early dysbiosis towards late more diverse status was observed in 6/8 mild COVID-19 adult patients. These findings reveal previously unknown interactions between upper respiratory and gut microbiomes during COVID-19.

Yang, L., et al. (2020). "Severe Pneumonia Advanced to Lung Abscess and Empyema Due to *Rothia Mucilaginosa* in an Immunocompetent Patient." *Am J Med Sci* **359**(1): 54-56.

Yin, W., et al. (2020). "Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir." *Science* **368**(6498): 1499-1504.

The pandemic of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has become a global crisis. Replication of SARS-CoV-2 requires the viral RNA-dependent RNA polymerase (RdRp) enzyme, a target of the antiviral drug remdesivir. Here we report the cryo-electron microscopy structure of the SARS-CoV-2 RdRp, both in the apo form at 2.8-angstrom resolution and in complex with a 50-base template-primer RNA and remdesivir at 2.5-angstrom resolution. The complex structure reveals that the partial double-stranded RNA template is inserted into the central channel of the RdRp, where remdesivir is covalently incorporated into the primer strand at the first replicated base pair, and terminates chain elongation. Our structures provide insights into the mechanism of viral RNA replication and a rational template for drug design to combat the viral infection.

Zhang, J. Y., et al. (2020). "Single-cell landscape of immunological responses in patients with COVID-19." *Nat Immunol* **21**(9): 1107-1118.

In coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, the relationship between disease severity and the host immune response is not fully understood. Here we performed single-cell RNA sequencing in peripheral blood samples of 5 healthy donors and 13 patients with COVID-

19, including moderate, severe and convalescent cases. Through determining the transcriptional profiles of immune cells, coupled with assembled T cell receptor and B cell receptor sequences, we analyzed the functional properties of immune cells. Most cell types in patients with COVID-19 showed a strong interferon-alpha response and an overall acute inflammatory response. Moreover, intensive expansion of highly cytotoxic effector T cell subsets, such as CD4(+) effector-GNLY (granulysin), CD8(+) effector-GNLY and NKT CD160, was associated with convalescence in moderate patients. In severe patients, the immune landscape featured a deranged interferon response, profound immune exhaustion with skewed T cell receptor repertoire and broad T cell expansion. These findings illustrate the dynamic nature of immune responses during disease progression.

A

B

C

A

B

B

C

D

E

- Host-pathogen interaction of human coronaviruses - autophagy
- SARS-CoV-2 altering angiogenesis via NRP1
- SARS coronavirus and innate immunity
- Type I interferon induction and signaling during SARS-CoV-2 infection

F

- Significant higher than healthy
- Significant lower than healthy
- Healthy reference
- No changes

WP5027:nsp1 from SARS-CoV-2 inhibits translation initiation in the host cell

CAP

Healthy

SARS-CoV-2

B

WP5027:nsp1 from SARS-CoV-2 inhibits translation initiation in the host cell

WP619:Type II interferon signaling

CAP

Healthy

SARS-CoV-2

WP61:Notch Signaling Pathway

F