Indian Journal of Biochemistry & Biophysics
Vol. 58, February 2021, pp. 7-12

A Review

Host response to SARS-CoV-2: Insight from transcriptomic studies

Saroj Kant Mohapatra* & Samanwoy Mukhopadhyay

Department of Computational Biology, National Institute of Biomedical Genomics, Kalyani-741 251, West Bengal, India

Received 01 October 2020; *revised 05 October 2020*

In the recent months, a number of transcriptomic studies have generated high-resolution data on the genes and pathways that are dysregulated in the patients infected with SARS-CoV-2 and enriched our understanding of the disease biology of this novel viral infection. The cumulative evidence collected from these data are considered in this article. Three motifs emerge with potential for future research and clinical translation. First, instead of a broad cytokine storm, one needs to interrogate the disease in terms of timing of specific cytokine up-regulation. Second, there is a subpopulation of immature or developing neutrophils in the patients with severe COVID-19 illness. This needs to be probed further for mechanistic insight and possible drug targets. Third, complement and coagulation cascades are significantly dysregulated in COVID-19, leading to the common clinical observation of a hypercoagulable state being associated with poor outcome. Interactions of these pathways with other immune-inflammatory pathways are important areas of future research. Finally, with rapid advances in relevant technologies in medicine (clinical transcriptomics, systems biology and artificial intelligence), we envisage deployment of these platforms in the clinical laboratory which shall benefit timely management of critical infectious illnesses such as COVID-19 and sepsis.

Keywords: Complement and coagulation, COVID-19, Neutrophil, scRNA seq, Systems biology, Transcriptomics

Introduction

The emergence of the corona virus disease 2019 (COVID-19) pandemic has made the scientific community scramble for answers. In addition to the virulence of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), variation in the host response has been considered an important contributor to outcome of the infection[1–3]. In this context, transcriptomic profiling has emerged as a powerful tool to dissect the host perturbation induced by an infection, *i.e.*, genome-scale disease biology of COVID-19. In this approach, one measures the level of gene expression (*i.e.*, transcript abundance) of the entire host genome, and performs unbiased analysis that enables discovery of pathways and processes not visible by candidate gene approach. Here, we discuss some recent studies illuminating the human host response to SARS-CoV-2, followed by the biological and clinical implication of these discoveries. The findings from these studies complement those from other important areas of research, such as, vaccine development and novel therapeutics.

Multiple studies on the COVID-19 transcriptome have generated plethora of data and provided insights into different aspects of the disease biology. List of the questions that have been answered by the recent transcriptome studies are presented in (Box 1). Most of the studies have performed transcriptome profiling by RNA sequencing (RNA seq) that generates the expression data of the entire human genome. This enables unbiased and deep analysis and exploration of differential and co-expression structures in the data space. A complementary approach takes into account those genes prioritized for an infectious disease (such as immune-related pathway genes), using a different technology[1].

Single-cell RNA-sequencing (scRNA seq) is becoming more common and has been performed by many authors[2,4]. This technique, *i.e.*, scRNA seq, differs from the more familiar bulk RNA sequencing in some important ways. Both techniques start with a sample of the heterogeneous tissue (consisting of multiple cell types). However, in bulk RNA seq method, total mRNA, or equivalently complementary

*Correspondence:
Phone: +91-33-66422111
E-mail: skm1@nibmg.ac.in

Abbreviations: ARDS, Acute respiratory distress syndrome; cDNA, Complementary DNA; COVID-19, coronavirus disease-2019; HPIV3, Human parainfluenza virus-3; IAV, Influenza A virus; MERS-CoV, Middle East respiratory syndrome-related coronavirus; PBMC, Peripheral blood mononuclear cells; RNA seq, RNA sequencing; RSV, Respiratory syncytial virus; SARS-CoV, Severe acute respiratory syndrome coronavirus; scRNA seq, Single-cell RNA-sequencing

Box 1 — Different questions that are answered by transcriptomic studies reported

1. What is the difference in blood transcripts of healthy subjects and those infected with SARS-CoV-2?[1,2,4,5]

2. What is a blood-based signature of severity and poor outcome (*e.g*., severity, death)?[1]

3. How do the gene expression change over time during the course of illness?[1]

4. How do different tissues other than blood (lungs, nasopharynx) respond to infection?[3,5,7,8]

5. How do different immune cells (leukocytes, monocytes, lymphocytes) respond to infection with SARS-CoV-2?[1, 2, 4, 8]

6. Is there a cytokine storm in COVID-19?[1-5,7]

*Some studies address multiple questions.

DNA (cDNA) is extracted from the sample followed by sequencing of the RNA molecules, leading to an estimate of the average transcript abundance in the sample. On the other hand, in the scRNA seq method, barcoding is done for single-cell origin of the cDNAs. Consequently, after sequencing of cDNA, methods for dimensionality reduction (such as clustering or principal components analysis) are performed to identify cell types associated with higher or lower levels of gene expression. While bulk RNA seq can identify differences in average transcript abundance between two samples, scRNA seq can identify differences in gene expression levels between specific cells of two samples. Also, scRNA seq can detect transcriptional differences between specific cell types (especially, if these are rare cell types in the given tissue), even when bulk RNA seq shows no difference in average gene expression levels.

Generation of high-dimensional data by scRNA seq for each subject offers the opportunity to explore the cellular context of the association between specific transcripts and the clinico-pathological phenotype of interest. For COVID-19, identification of the immune cells associated with severity or outcome is motivated by the quest for diagnostic/prognostic biomarkers (for stratification of the patients) and immunomodulatory therapy (either boosting appropriate and suppressing harmful cellular processes). Some key results from these studies are listed in the next section.

One of the first studies was done by Blanco-Melo and colleagues who compared the host response to SARS-CoV-2 versus other respiratory viruses, such as Middle East respiratory syndrome-related coronavirus (MERS-CoV), severe acute respiratory syndrome related coronavirus 1 (SARS-CoV-1), human parainfluenza virus 3 (HPIV3), respiratory syncytial virus (RSV) and influenza A virus (IAV), using cell and animal models in addition to transcriptional and serum profiling of patients[5]. Host response to an infection was shown to be very different between SARS-CoV-2 and other respiratory viruses. In the patients infected with SARS-CoV-2, there was an inappropriate response to the infection including low type I and III interferon signalling but high levels of pro-inflammatory cytokines/chemokines such as interleukin 6 (IL-6), interleukin 1 receptor type 1 (IL1RA), monocyte-recruiting C-C motif chemokine ligand 2 (CCL2), C-C motif chemokine ligand 8 (CCL8) and neutrophil chemoattractant C-X-C motif chemokine ligand 2 (CXCL2), C-X-C motif chemokine ligand 8 (CXCL8). The authors concluded that an imbalanced host response to SARS-CoV-2 comprising reduced innate antiviral response failing to clear the pathogen combined with harmful inflammatory response constitute a hallmark of COVID-19 disease process.

Wilk and colleagues sought to provide a cell atlas of the peripheral immune response to severe COVID-19[2]. Single cell RNA sequencing was performed on peripheral blood mononuclear cells (PBMC) from six healthy donors and seven patients suffering from COVID-19 (including four patients with acute respiratory distress syndrome). A major finding was reconfiguration of peripheral immune cell phenotype including a heterogeneous interferon-stimulated gene signature, HLA class II downregulation and a developing neutrophil population, closely related to plasmablasts. Cellular trajectory analysis revealed a differentiation bridge from plasmablast to neutrophil, whereby the cells progressively lost expression of genes encoding canonical plasmablast markers (CD27, CD38, TNFRSF17) and acquired expression of genes encoding neutrophil granule proteins (defensin, elastase and myeloperoxidase) in the blood of COVID-19 patients with acute respiratory failure requiring mechanical ventilation. The authors did not observe any over-expression of pro-inflammatory cytokines by the peripheral monocytes and lymphocytes. Overall, the authors reported marked changes in the immune cell composition and phenotype in infection with specific immunological features of severe COVID-19 patients with acute respiratory distress syndrome (ARDS).

Viruses deploy molecular mimicry to harness or disrupt cellular functions including modulation of immune responses. It was already shown that over 140 host cellular proteins are structurally mimicked

by coronaviruses (CoV) with complement and coagulation pathways as targets of this strategy across all CoV strains[6]. Follow-up work by Ramlall and colleagues included a multimodal approach that combined clinical informatics, transcriptomics, genomics and molecular structure-function analysis to identify factors associated with poor clinical outcome of COVID-19[3]. First, analysis of the health records revealed that prior clinical history of coagulation or complement-activation disorders are risk factors for SARS-CoV-2-associated morbidity and mortality. Second, transcriptome analysis of nasopharyngeal swabs from a large cohort (650 control + case) revealed enrichment of complement and coagulation pathway (Fig. 1). Additional genetic evidence made a compelling argument that complement and coagulation functions play a major role in the outcome of SARS-CoV-2 infection.

Immune gene expression profiling was performed in three patients of COVID-19, one with severe and two with moderate illness[1]. Although the number of patients was small, blood samples were collected at multiple time points for each subject, thus enabling dissection of gene expression dynamics during the course of the illness. The authors made some key observations. First, in the severe case, there was increased cytokine expression, but timing of up-regulation was variable. Most pro-inflammatory cytokines followed the lowest point in respiratory dysfunction suggesting that these were the consequence rather than cause of respiratory dysfunction. However, interleukin-1 (IL-1) signalling was observed early on, suggesting that IL-1 and related pro-inflammatory pathways may be prognostic markers and/or therapeutic targets for COVID-19. Second, pro-inflammatory response was intertwined with T cell activation leading to the dichotomy between viral clearance and exacerbation of disease. For example, the case with severe illness was associated with rapid viral clearance, but increased respiratory dysfunction and requirement for supplemental oxygen. On the other hand, the cases with moderate illness were consistently positive for SARS-CoV-2 suggesting failure of viral clearance by the host immune system.

Lee and colleagues[4] performed high-resolution transcriptome profiling (single-cell RNA seq) of PBMC

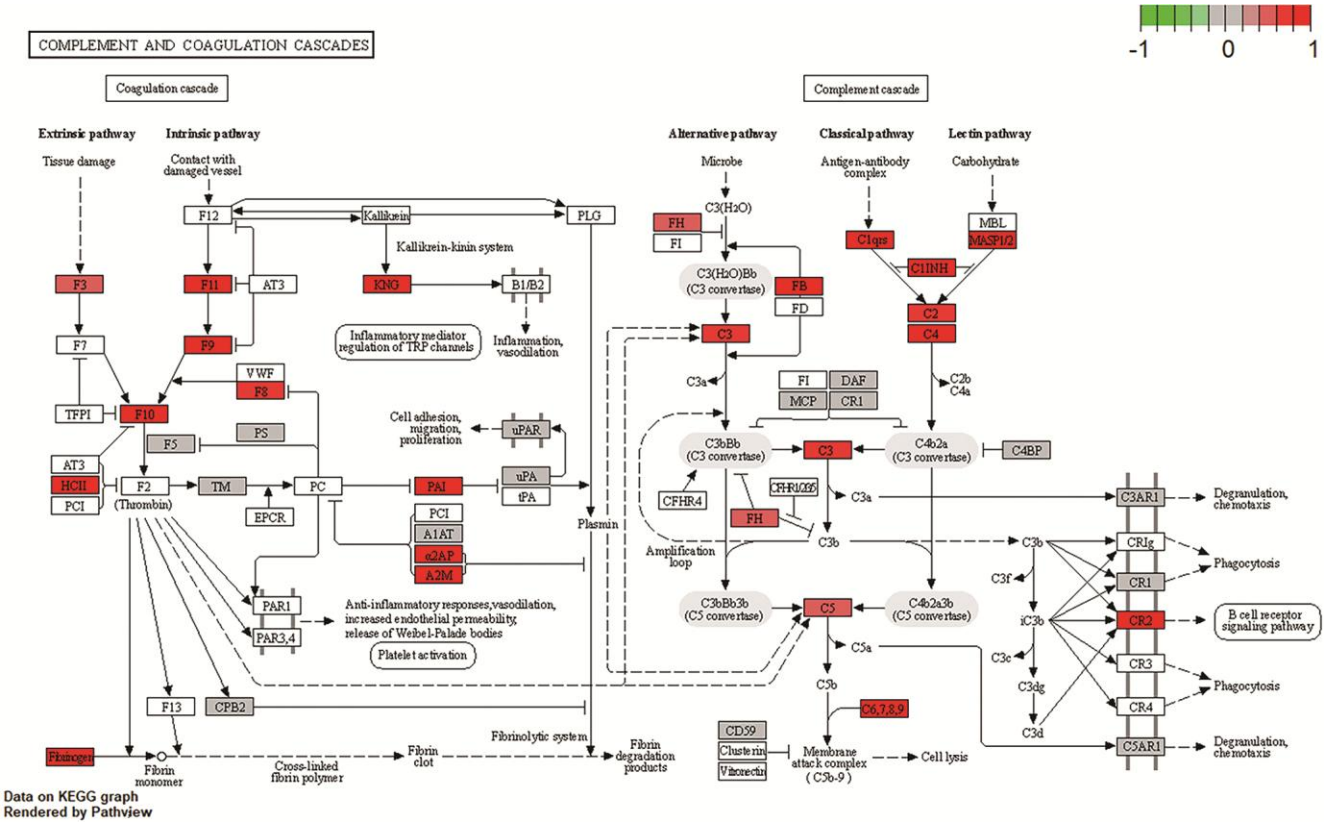

Fig. 1 — Up-regulation of complement and coagulation cascades. Each node corresponds to a gene belonging to the KEGG pathway "Complement and Coagulation Cascade". Up-regulated genes are shown as red nodes in the pathway map. Data were extracted from[3] and drawn using R and Bioconductor[13]

from healthy subjects, influenza, COVID-19 of varying severity. Comparison was made between the host response to the two respiratory viruses (influenza and SARS-CoV-2) with an effort to better define the immune response in COVID-19. Prominently, a TNF-α/IL-1β-driven hyperinflammatory response was observed in patients with severe COVID-19, but not in patients with mild illness. They proposed that type I interferon response plays a key role in exacerbating inflammation in severe COVID-19. The finding of this study stands in contrast to that of other study[2], who did not observe the up-regulation of proinflammatory cytokines.

Liao and colleagues performed single-cell transcriptomic profiling of broncho-alveolar lavage fluids from SARS-CoV-2 patients of varying disease severity[7]. They observed an increased proportion of pro-inflammatory monocyte derived macrophage in patients with severe illness, and clonal expansion of $CD8^+$ T cells in patients with moderate illness.

Xiong and colleagues performed transcriptome profiling of broncho-alveolar lavage fluids and PBMC samples from COVID-19 patients and observed strong association between COVID-19 pathogenesis and excessive cytokine release such as CCL2/MCP-1, CXCL10/IP-10, CCL3/MIP-1A, and CCL4/MIP1B[8]. Additionally, the authors suggested that lymphopenia (observed in COVID-19 patients) is possibly due to SARS-CoV-2-induced activation of apoptosis and p53 signalling in lymphocytes.

Sequencing platforms

The studies cited here employed different platforms such as single cell RNA seq, bulk RNA seq, or more targeted approaches such as Nanostring nCounter. Often the choice of platform is linked with the question being asked by the investigators. The newer technology of single cell RNA sequencing provides data at a very high cellular granularity. However, this is an evolving technology with heterogeneous workflow and may lead to contradictory results[2,4]. Bulk RNA sequencing is a safer option, with well-established analysis pipeline and reproducible results. A well-established platform such as bulk RNA sequencing produces reproducible result, but is limited to coarse tissue level data. The Nanostring nCounter technology is especially suitable for evaluation of disease-specific signature gene set [1].

A series of studies mentioned inthe preceding sections have generated extensive transcriptomic data from human subjects with COVID-19. The findings from these studies add up to a set of interesting observations and are summarized in (Table 1). Most commonly observed transcriptional alterations include that of cytokines, interferon and HLA II. Gene expression profiling can be de-convoluted to reason about the cellular source of the RNA. This has led to frequent reporting of differential immune cell abundance in the blood or tissue of COVID-19 patients with varying illness severity. Further, innate immune processes such as complement and coagulation have been shown to be significantly up-regulated in patients with COVID-19 illness. Three important motifs emerge from the cumulative evidence collected by the transcriptomic studies.

Cytokine storm

Multiple transcriptome analyses from blood immune cells and lung tissue suggest the presence of higher level of cytokines in COVID-19 patients[1,3,5,8]. This is also consistent with the immunophenotyping data[4] and the prevailing paradigm of an uncontrolled pro-inflammatory cytokine storm being the principal driver of the disease. However, cytokine storm was not observed in all COVID-19 cohorts. Wilk and colleagues observed no evidence of a cytokine storm in the patients of COVID-19[2]. Also, different cytokines were identified as the dominant driver of pathophysiology by different studies. IL-1, IL-6 and TNF-α have all been implicated in the disease progression and outcome of COVID-19, but there is no consensus. Of note, two single cell RNA seq studies have conspicuously differed on the finding of pro-inflammatory cytokines such as IL-1β and TNF-α in COVID-19[2,4]. Further, there is significant difference in the dynamics of multiple cytokine gene expression during COVID-19 disease, where IL-1 signalling preceded (while other cytokines succeeded) the respiratory function nadir in a case of severe COVID-19 illness[1]. While COVID-19 illness cannot be completely explained in terms of a "cytokine storm", unravelling the interaction between specific cytokines and biologically relevant immuno-inflammatory processes are likely to impact precise therapeutics (*e.g.*, anti-IL-6 Tocilizumab) for COVID-19.

Immune cells in COVID-19

Severe COVID-19 illness is strongly associated with lymphopenia and neutrophilia[9]. Apoptosis and p53 signalling have been implicated in lymphopenia[8]. However, there is targeted expansion of $CD8^+$ T cells

Table 1 — Salient findings of some key transcriptomic studies on COVID-19

	Ong 2020	Wilk 2020	Lee 2020	Xiong 2020	Liao 2020	Blanco-Melo 2020	Ramllal 2020
Country	Singapore	USA	South Korea	China	China	USA	USA
Number of subjects	Control (10), Moderate (2), Severe (1)	Control (6), NonVent (4), ARDS (4)	Control (4) Severe influenza (5), Asymptomatic nCOV case (1), Mild nCOV cases (4), Severe nCOV cases (6)	Control (3), Case PBMC (2), Case BALF (3)	Control (3), Moderate (3), Severe (6)	Non infected Lung (2), Deceased Lung (2)	Control (519), Case (216)
RNA source	Blood	PBMC	PBMC	BALF, PBMC	BALF	Lung	Nasopharyngeal Swab
Platform Technology	NanoString nCounter	Single Cell Sequencing	Single Cell Sequencing	Bulk RNA Sequencing	Single Cell Sequencing	Bulk RNA Sequencing	Bulk RNA Sequencing
Cytokine	Up-regulation of cytokines including IL-1	No evidence of up-regulation of pro-inflammatory cytokines	Up-regulation of TNFa, IL1b	Up-regulation of CCL2 (MCP-1), CXCL10 (IP-10), CCL3 (MIP1A), CCL4 (MIP1B)	Increased number of pro-inflammtory monocyte-derived macrophage	Up-regulation of IL1RA, IL-6	Up-regulation of IL-6 response
Interferon	N.R.	Heterogeneous IFN response	Type I IFN up-regulated in severe patients	N.R.	N.R.	No up-regulation of Type I, Type III IFN	Up-regulation of Type-I IFN response
HLA II	Down	Down	N.R.	N.R.	N.R.	N.R.	N.R.
Lymphocyte	T cell activation	N.R.	N.R.	Lymphopenia is caused by apoptosis of Lymphocytes and p53 signaling	Highly clonally expanded CD8+ T cells	N.R.	N.R.
Neutrophil	N.R.	Developing Neutrophils, segregate with plasmablasts	N.R.	N.R.	N.R.	Up-regulation of neutrophil chemoattractants CXCL2, CXCL8	N.R.
Complement /Coagulation	N.R.	N.R.	N.R.	Up-regulation of complement cascade	N.R.	N.R.	Up-regulation

BALF-Broncho-alveolar lavage fluid; N.R.-Not reported; nCOV-COVID-19; PBMC-Peripheral Blood Mononuclear Cells

in specific tissues such as lungs[7] and T cell activation in blood[1]. Apart from increase in the neutrophil count, there is up-regulation of circulating chemo attractants CXCL2, CXCL8 (for neutrophils), CCL2, CCL8 (for monocytes), *etc*. in COVID-19 patients[5]. Moreover, not all neutrophils observed in the severe COVID-19 illness are mature neutrophils. There is evidence of increased number of "developing" neutrophils in patients of severe COVID-19 with respiratory distress, either derived from plasma blasts, as a consequence of emergency granulopoiesis[2], or by another mechanism[10]. Of note, their role in severe and critical COVID-19 illness remains to be unmasked.

Complements and Coagulation pathway

Multi-modal transcriptomic analysis revealed that SARS-CoV-2 engages robust activation of complement and coagulation cascades[3,8]. Abnormal coagulative states predispose individuals to adverse outcomes associated with SARS-CoV-2 outcomes[11]. In the laboratory, hypercoagulable state is frequently encountered in the patients of COVID-19. A high level of D-dimer (a fibrin degradation product) is particularly associated with severity of illness and poor outcome. Up-regulation of complement pathway is strongly associated with coagulopathy and respiratory dysfunction[12]. The cellular and molecular pathways that communicate with

complement and coagulation activation in COVID-19 is an active area of research. It is expected that new knowledge in this branch of innate immunity shall pave the way for precision therapy of those who are critically ill with SARS-CoV-2 infection.

Conclusion

In this article, we have discussed some recent transcriptomic studies performed in the patients of COVID-19. The studies were conducted with cutting-edge technology and provided insight into the disease biology of human infection by SARS-CoV-2. First, different cytokines were reported to be up-regulated by different authors. Also, the expression of these cytokines had a dynamic component vis-à-vis clinical parameters, especially respiratory function in severe COVID-19 illness. Therefore, instead of a broad cytokine storm, specific and transient cytokine expression is more likely a driver of the down-stream molecular and cellular modules. Second, neutrophilia and lymphopenia are clearly evident from the transcriptomic studies. While gene expression consistent with apoptosis and p53 signalling underlie lymphopenia, neutrophilia is associated with increased abundance of immature or developing neutrophils in the blood of the severe COVID-19 patients. This novel finding warrants deeper investigation and shall likely lead to novel mechanistic insight and therapeutic advances. Third, complement and coagulation cascades are progressively up-regulated in critically ill patients with COVID-19. A hypercoagulable state is a predictor of poor outcome of COVID-19. Along this line, interactions of the complement pathway with coagulation cascade and with other immune-inflammatory processes are being pursued intensively and are likely to generate actionable knowledge for precise therapy of COVID-19 patients. In conclusion, while transcriptomic profiling has provided valuable insight into COVID-19 disease processes, the future looks brighter with integration of clinical metadata and gene expression profiling that shall produce meaningful data generated in a realistic time-frame for the treating physician. Further, with increased automation of clinical transcriptomics and related technologies that generate high-dimensional data in a narrow time window, application of artificial intelligence shall revolutionize critical care for infectious diseases such as COVID-19 and sepsis.

Conflict of interest

All authors declare no conflict of interest.

References

1 Ong EZ, Chan YFZ, Leong WY, Lee NMY, Kalimuddin S, Mohideen SMH, Chan KS, Tan AT, Bertoletti A, Ooi EE & Low JGH, A Dynamic Immune Response Shapes COVID-19 Progression. *Cell Host Microbe*, 27 (2020) 879.

2 Wilk AJ, Rustagi A, Zhao NQ, Roque J, Martínez-Colón GJ, McKechnie JL, Ivison GT, Ranganath T, Vergara R, Hollis T, Simpson LJ, Grant P, Subramanian A, Rogers AJ & Blish CA, A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nat Med*, 26 (2020) 1070.

3 Ramlall V, Thangaraj PM, Meydan C, Foox J, Butler D, Kim J, May B, De Freitas JK, Glicksberg BS, Mason CE, Tatonetti NP & Shapira SD, Immune complement and coagulation dysfunction in adverse outcomes of SARS-CoV-2 infection. *Nat Med*, 26 (2020) 1609.

4 Lee JS, Park S, Jeong HW, Ahn JY, Choi SJ, Lee H, Choi B, Nam SK, Sa M, Kwon JS, Jeong SJ, Lee HK, Park SH, Park SH, Choi JY, Kim SH, Jung I & Shin EC, Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Sci Immunol*, 5 (2020) eabd1554.

5 Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Møller R, Jordan TX, Oishi K, Panis M, Sachs D, Wang TT, Schwartz RE, Lim JK, Albrecht RA & tenOever BR, Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell*, 181 (2020) 1036.

6 Lasso G, Honig B, Shapira SD. A Sweep of earth's virome reveals host-guided viral protein structural mimicry and points to determinants of human disease. *Cell Syst*, 12 (2020) 82.

7 Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, Cheng L, Li J, Wang X, Wang F, Liu L, Amit I, Zhang S & Zhang Z, Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med*, 26 (2020) 842.

8 Xiong Y, Liu Y, Cao L, Wang D, Guo M, Jiang A, Guo D, Hu W, Yang J, Tang Z, Wu H, Lin Y, Zhang M, Zhang Q, Shi M, Liu Y, Zhou Y, Lan K & Chen Y, Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg Microbes Infect*, 9 (2020) 761.

9 Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J & Cao B,Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395 (2020) 497.

10 Alquicira-Hernandez J, Powell JE & Phan TG, No evidence that plasma blasts transdifferentiate into developing neutrophils in severe COVID-19 disease. *bioRxiv*, (2020) doi: https://doi.org/10.1101/2020.09.27.312538.

11 Rauch A, Labreuche J, Lassalle F, Goutay J, Caplan M, Charbonnier L, Rohn A, Jeanpierre E, Dupont A, Duhamel A, Faure K, Lambert M, Kipnis E, Garrigue D, Lenting PJ, Poissy J & Susen S, Coagulation biomarkers are independent predictors of increased oxygen requirements in COVID-19. *J Thromb Haemost*, 18 (2020) 2942.

12 Fletcher-Sandersjöö A & Bellander BM, Is COVID-19 associated thrombosis caused by overactivation of the complement cascade? A literature review. *Thromb Res*, 194 (2020) 36.

13 R Core Team, A language and environment for statistical computing. (2008) URL: https://www.R-project.org/.