A blood-based comprehensive and systems-level analysis of disease stages, immune regulation and symptoms in COVID-19 patients

Anguraj Sadanandam (anguraj.sadanandam@icr.ac.uk)
Institute of Cancer Research (ICR) https://orcid.org/0000-0001-8485-5150
Tobias Bopp
Johannes Gutenberg University Mainz
Santosh Dixit
4Centre for Translational Cancer Research
David JHF Knapp
Université de Montréal
Chitra Priya Emperumal
University of Southern California
Krishnaraj Rajalingam
University Medical Center of the Johannes Gutenberg University Mainz
Alan Melcher
The Institute of Cancer Research
Nagarajan Kannan (Kannan.Nagarajan@mayo.edu)
Mayo Clinic Cancer Center

Keywords: COVID-19, coronavirus, systems biology, immunology, disease stages, symptoms, genes, pathways, networks, immune cells, SARS

Posted Date: May 21st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-30473/v1

Version of Record: A version of this preprint was published at Cell Death Discovery on May 21st, 2020.
See the published version at https://doi.org/10.1038/s41420-020-00376-x.

A blood-based comprehensive and systems-level analysis of disease stages, immune regulation and symptoms in COVID-19 patients

Anguraj Sadanandam[1,2,#], Tobias Bopp[3], Santosh Dixit[4], David JHF Knapp[5,6], Chitra Priya Emperumal[7], Krishnaraj Rajalingam[8,9], Alan Melcher[10], Nagarajan Kannan[2,11,12,#]

[1]Division of Molecular Pathology, The Institute of Cancer Research, London, United Kingdom.
[2]Division of Experimental Pathology, Department of Lab Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, 55905, United States.
[3]Institute for Immunology, University Medical Center, Johannes Gutenberg University Mainz, Mainz, Germany.
[4]Centre for Translational Cancer Research (CTCR; a joint initiative of Indian Institute of Science Education and Research (IISER) Pune and Prashanti Cancer Care Mission), Pune, India.
[5]Institut de recherche en immunologie et en cancérologie, Université de Montréal, Montreal, Quebec, Canada.
[6]Département de pathologie et biologie cellulaire, Université de Montréal, Montreal, Quebec, Canada.
[7]Herman Ostrow School of Dentistry, University of Southern California, Los Angeles, California, USA.
[8]Cell Biology Unit, University Medical Center of the Johannes Gutenberg University Mainz, 55131, Mainz, Germany.
[9]University Cancer Center Mainz, University Medical Center, Mainz, 55131, Germany.
[10]Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, United Kingdom.
[11]Center for Regenerative Medicine, Mayo Clinic, Rochester, Minnesota, 55905, USA.
[12]Mayo Clinic Cancer Center, Mayo Clinic, Rochester, Minnesota, 55905, USA.

Corresponding authors:
Anguraj Sadanandam PhD
Head of Systems and Precision Cancer Medicine
Institute of Cancer Research (ICR)
15 Cotswold Road, Sutton, SM2 5NG
United Kingdom
Email: anguraj.sadanandam@icr.ac.uk

&
Nagarajan Kannan PhD
Head of Laboratory of Stem Cell and Cancer Biology
Mayo Clinic
200 First St SW,
Rochester, MN, 55905, USA
Email: kannan.nagarajan@mayo.edu

Short title: Systems-level analysis for COVID-19 PBMC

Keywords: COVID-19**,** SARS, coronavirus, PBMC, blood gene signatures, cytokine storm, immune suppression, disease aggression, disease stages, personalized medicine, machine-learning, inferential statistics, systems biology

Abstract

COVID-19 patients show significant clinical heterogeneity in presentation and outcomes that makes pandemic control and strategy difficult; optimising management requires a systems biology approach of understanding the disease. Here we sought to understand and infer complex system-wide changes in patients infected with coronaviruses (SARS-CoV and SARS-CoV-2; n=38 and 57 samples) at two different disease stages compared with healthy individuals (n=16) and patients with other infections (n=144). We applied inferential statistics/machine-learning approaches (the COVID-engine platform) to RNA profiles derived from peripheral blood mononuclear cells (PBMCs). Compared to healthy individuals, an integrated blood-based gene signatures distinguished acute-like (mimicking coronavirus-infected patients with prolonged hospitalisation) from recovering-like patients. These signatures also hierarchically represented systems-level parameters associated with PBMC including dysregulated cytokines, genes, pathways, networks of pathways/concepts, immune status, and cell types. Proof-of-principle confirmatory observations included PBMC-associated increases in *ACE2*, cytokine storm-associated *IL6*, enhanced innate immunity (macrophages and neutrophils), and lower adaptive T and B cell immunity in patients with acute-like disease compared to those with recovery-like disease. Patients in the recovery-like stage had significantly enhanced *TNF*, *IFN-*γ, anti-viral, *HLA-DQA1*, and *HLA-F* gene expression and cytolytic activity, and reduced pro-viral gene expression compared to those in the acute-like stage in PBMC. Besides, PBMC-derived surrogate-based approach revealed overlapping genes associated with comorbidities (associated diabetes), and disease-like symptoms (associated with thromboembolism, pneumonia, lung disease and septicaemia). Overall, our study involving PBMC-based RNA profiling may further help understand complex and variable systems-wide responses displayed by coronavirus-infected patients.

Introduction

The spread of COVID-19, a disease caused by Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2), has led to the current global pandemic with already more than 4 million people with confirmed infection and nearly 300,000 deaths within a few months(*1*). According to the World Health Organization (WHO), the mode of infection for COVID-19 is predominantly through respiratory droplets, aerosol transmission due to pathogen-laden viral particles in the air, or close contact with infected people with increased viral loads, especially in the early stages of disease(*2*). The mechanism of human pathogenesis, to a great extent, may simulate that of SARS-CoV (associated with SARS) and Middle East Respiratory Syndrome CoronaVirus (MERS-CoV; MERS) viral infections, including the prolonged persistence of the virus worsening the host immune response(*3-5*). Clinical manifestation of COVID-19 ranges from mild respiratory symptoms to severe disease and death(*6*). However, there are now reports suggesting heterogeneous manifestation of the disease affecting multiple organs, including kidney, liver, and brain(*7*). Although age and compromised health history are considered critical prognostic signatures, certain patients of younger age and good health have shown severe progression of the disease(*8*).

Patients with COVID-19 may be asymptomatic(*9, 10*), but can still transmit infection(*11*). Viral shedding from an infected person may occur, although resolution of symptoms(*12*), and relapse has been reported despite consecutive negative testing(*13*). Currently, there is no standard of care to treat COVID-19 respiratory symptoms(*14, 15*), screen for potential organ failures, disease aggression, and systemic changes in patients.

In this study, we sought to understand and infer changes in coronavirus-infected patients at multiple levels through their peripheral blood mononuclear cells (PBMC) by performing comparisons with healthy volunteers by applying inferential statistics (IS) and machine-learning (ML). The inferences at the systems level in patients provide an efficient way of understanding the heterogeneity and mechanism(s) of disease manifestation as a whole. These inferences can be derived systematically in a hierarchical fashion from the level of gene signatures to the whole organism, to study the pathophysiology of COVID-19 patients. Moreover, the patients' response to insults from the virus along with other associated symptoms can be studied. Nevertheless, we carefully interpreted all these inferences to limit to mononuclear cells in PBMC and its relevance to the disease.

Methods

COVID-engine refers to a pipeline of different IS/ML methods described below. This platform applies gene signatures from PBMC of patients infected with coronavirus to query multiple databases for meta-signatures such as, pathways, mechanistic processes and their associated networks that are connected with different disease manifestations. These signatures and meta-signatures are further modeled systematically to wire the pathophysiology in patients, again, in a hierarchical fashion, from cells to whole organism level (Figure 1A). SARS, COVID-19 and other samples were obtained from published studies(*16, 17*). COVID-engine includes the following methods. PBMC-based differential gene signatures between healthy volunteers and patients were selected by performing supervised Statistical Analysis of Microarrays (SAM)(*18*) using R-based *siggenes* package (*19*). Various gene scores were derived using single sample Gene Set Enrichment Analysis (ssGSEA)(*20*). Hypergeometric gene enrichment analysis were performed using hyperR R-based tool(*21*). Nearest Template Prediction(*22*) was used to derive distance between two signatures – acute-like vs. recovering-like patients. Different gene set databases were from EnrichR(*23*) and MSigDB(*20*). Immune gene sets were from Rooney *et al*(*24*). The references to the used gene sets and databases are provided in the respective figures.

Results

PBMC-based differential gene expression signatures between coronavirus-infected patients and healthy volunteers

To perform an integrative and systematic analysis of heterogeneous patients' responses to the coronavirus infection, we used RNA transcriptome data from PBMC of SARS-CoV patients (n=10) and healthy volunteers (n=4; from a published study(*16*); training data). We identified PBMC gene signatures differentially expressed between patients and healthy volunteers by applying our in-house developed pipeline of IS/ML platform, referred to as "COVID-engine" (see *Methods*). We identified 290 gene signatures that were differentially expressed in patients and healthy volunteers using a supervised SAM approach (Figure 1B and Supplementary Figure 1). Among these 290 genes, 169 (dubbed as CoV-Up-gene signatures) were highly expressed, and 121 highly reduced in patient samples compared to healthy volunteers (dubbed CoV-Down-gene signatures; Figure 1B).

Using COVID-19 patients' PBMC (n=3) and healthy volunteers (n=3; from another published study(*25*)), we further validated our gene signatures. We observed that our CoV-Up-gene signatures from SARS-CoV was higher in COVID-19 patients' PBMC than in the healthy volunteers (see *Methods*). In contrast, CoV-Down-gene signatures was higher in healthy volunteers and lower in COVID-19 patients. This result establishes that our gene signatures from SARS-CoV are applicable for COVID-19 patients (Figure 1C-D).

When our CoV-Up-gene signatures was analysed using PBMC samples (*17, 26*) (n=213) from patients infected with bacteria and influenza, we observe a broadly similar pattern (Figure 1E and Supplementary Figure 2A). The reciprocal analyses with CoV-Down-gene signatures was higher in PBMC from healthy volunteers than from SARS-CoV, COVID-19, and other microbe-infected patients (Figure 1F and Supplementary Figure 2B). Again, we performed enrichment analysis (hypergeometric test) using MSigDB's C7 immune signature and found that 43% (60 out of 138) of the signatures that were derived from PBMC (mostly associated with specific diseases) were significantly (FDR < 0.2) enriched for our CoV-Up-gene signatures (Supplementary Figure 3). Overall, the results suggest that CoV-Up-gene signatures represent primarily diseased PBMC as initially derived.

PBMC gene signatures distinguish disease stages – acute-like vs. recovering-like coronavirus-infected patients

Next, we sought to assess the potential of our CoV-Up-gene signatures to stratify patients into those at different stages of the disease (aggression): acute vs. recovering. For this, we used PBMC transcriptome data from 44 samples from the longitudinal collection over the disease course from a published study (validation data; Lee *et al*.) (*17*). Lee *et al*. defined acute samples (n=25) as those that tested positive (using blood) for SARS-CoV during hospitalization or within 10 days of onset of the disease in patients. These acute samples were also correlated with disease severity including an increased clinical pulmonary infection score (CPIS) (*17, 27*). The remaining samples were labeled as recovering samples (n=19). Hence, acute vs. recovering samples refer to different stages of the disease that can occur in the same patient as specific samples were collected from the same individual during hospitalization.

Interestingly, disease phase appeared to be associated with our identified CoV expression signatures in samples from Lee *et al*(*17*) with recovering patients showing an intermediate signature between acute phase and healthy donors (Figure 1E). We then applied gene signature (39 genes with known gene symbols) for distinguishing acute from recovering

patients identified by Lee *et al(17)* back to our training data (Regunathan *et al*(*16*)). This resulted in 7 of the 10 patient samples showed maximal similarity to acute phase (termed aute-like patients), while 3 scored similar to recovering patients (termed recovering-like patients; measured as signature-based cosine distance; see *Methods*; Figure 1G). These results suggest that the gene signatures predominantly contain patients at the acute-like stage and healthy individual-based gene indicators; however, it may distinguish patients at the recovering-like stage.

Gene expression of *ACE2* and key cytokines in PBMCs in acute-like vs. recovering-like coronavirus patients

We next analyzed the expression patterns of genes associated with coronavirus infection in PBMCs in acute-like and recovering-like using SARS-CoV-infected patients and healthy volunteers (from training data). We compared the expression of *ACE2*, *IL6*, and *TNF* genes in PBMC from infected patients and healthy individuals, which are known to be expressed in circulating monocytes and macrophages after viral infection (*28, 29*)*,*(*30*). Although there was increased expression of *ACE2* and *IL6* in acute-like patients compared to healthy individuals, there was no significant differences between acute-like and recovering-like patients (Figure 2A-B). Interestingly, *TNF* was highly expressed in acute-like patients compared to recovering-like patients and healthy individuals (Figure 2C). Besides, analysis of subunits of lactate dehydrogenase (*LDHA* and *LDHB*; associated with hypoxia) genes showed *LDHB* was highly expressed in healthy individuals compared to coronavirus-infected patients, and inverse trends were observed for *LDHA* (Figure 2D-E). Among these five genes, *TNF* was the only gene that showed differential expression between acute-like and recovering-like patients. These results show increased expression of these key COVID-19/SARS genes in patient PBMC samples.

We further examined multiple other candidate genes that act as chemo-attractants to monocytes and macrophages, specifically those that interfere with innate and adaptive immunity and viral replication (*29*). Among those genes, we observed *CXCL8* (*IL8*) and *CCL13,* which are associated with chemoattraction of neutrophils/macrophages (innate immunity), to be highly expressed in acute-like patients, compared to recovering-like patients and healthy individuals (Figure 3A). On the other hand, *OAS2* and *IL16* associated with T cells (adaptive immunity) and inhibition of viral replication were highly expressed in recovering-like patients and healthy individuals (Figure 3B). These results suggest that PBMC from acute-like patients may be associated with the activity of innate immunity, whereas PBMC from recovering-like patients may be associated with an adaptive immune profile.

Enrichment of TNF-alpha, *IL6* and hypoxia-related pathways in PBMC of coronavirus patients

Based on the expression of key genes, including *TNF*, we next set out to investigate the functional implications of the CoV-Up-signature. To do so, we performed enrichment analysis using the CoV-Up-genes and MSigDB's Hallmarks database (Figure 3C). This revealed multiple highly ranked pathways involved in cytokine storm and acute infection including TNF signalling, IL6 (IL6-JAK-STAT3) and IL2 (IL2-STAT5) signaling, inflammatory response, and KRAS/MTOR and late responses to estrogen pathways (Figure 3C). This is consistent with clinical manifestations including observations of high IL6 levels in COVID-19 patients(*31*).

Interestingly, outside of the inflammatory response, multiple pathways related to hypoxia, angiogenesis and oxygen transport (heme/iron) were also implicated, consistent with the oxygen limitation(*32*) experienced during coronavirus infection (Figure 3C). Of particular interest was the enrichment of complement and coagulation pathways which may explain the high frequency of embolisms observed in COVID-19 patients(*33*) and may represent one of the key pathological mechanisms of the virus. Enrichment of the apoptotic pathway may have relevance to cell death of lymphocytes as suggested elsewhere(*25*). These enriched pathways and processes may be linked together and associated with COVID-19 infection in these patients.

Potential role for a network of related pathways representing cytokine storm and innate immune changes in PBMC of coronavirus-infected patients

Given that different pathways were enriched in infected patients, next, we interrogated how these pathways are linked together to convey a network of processes or changes at the cellular level. Hence, we used the REACTOME pathway database(*34*) to connect different but related pathways that were enriched in infected patients using network analysis. Two evident and distinct networks were: a) interleukins and cytokine signaling (potentially representing cytokine storm), and b) neutrophils and innate immunity showed significant enrichment using CoV-Up-gene signatures (Figure 4A). Nevertheless, we observed an increased enrichment of a unique network linking granulopoiesis, megakaryocyte differentiation, and platelet activation (Figure 4A). This may be linked to coagulation system that can activate the innate immune system (e.g. monocytes/macrophages) to produce TNF (*35*). Nevertheless, this requires further understanding. These clearly show innate immune system activation with potential cytokine storm in coronavirus patients.

Network analysis using Kyoto encyclopedia of genes and genomes (KEGG)-based revealed an extensive network of pathways of different infections, including *Helicobacter pylori* and leishmania (Figure 4B). Interestingly, there was an enrichment of systemic lupus erythematosus (SLE), which is a chronic disease associated with inflammation in connective tissue and affects multiple organs, including the blood-forming system. While SLE patients have been reported to be more vulnerable to COVID-19, clinical validation, and understanding at the systemic level is needed to know whether COVID-19 (*36*) symptoms may be associated with SLE. A more relevant and well-known individual pathway associated with this disease is the renin-angiotensin system associated with *ACE2* function (Figure 4B). There is also an enrichment of RIG-I-like receptor signaling pathway representing potential anti-viral event through pathogen-associated molecular patterns (*37*). Multiple chemokine/cytokine and metabolism pathways were enriched as a part of the CoV-Up-gene signatures in the KEGG database (Figure 4B).

Significant changes in subcellular regulatory networks associated with PBMC of coronavirus-infected patients

While these genes to network processes provide information related to coronavirus infection, we were interested in investigating the next level in the hierarchy and the potential subcellular interaction networks that may inform viral interaction within host cells. Interestingly, cell-cell adhesion processes, secretory granules, vesicles, and exosomes spanning plasma membrane and lipid complexes and cytoplasm were enriched in CoV-Up-gene signatures (Figure 4C), suggesting that this may indicate the viral infection of monocytes that may be associated with increased *ACE2* expression (Figure 2A). Furthermore, these were associated with enrichment of fatty acid synthase complex that is known to be involved in the plasma membrane and vesicle formation (*38*) (Figure 4C). The data also suggests potential interaction of the virus with host immune cells through cell-cell adhesion processes, which requires further investigation.

Nonetheless, the host-specific subcellular changes in PBMC are also evident from this analysis. An increased replication and proliferation of potential host cells, mainly involving the innate immune system, may be evident based on the enrichment of genes associated with DNA polymerase processivity factor and proliferating cell nuclear antigen (PCNA) complex. Also, the production of immunoglobin (IgG) complexes along with NFkB complex is higher in these patient gene signatures, again, representing increased immune responses. At the same time, the host's potential responses to death signals associated with BCL2 complex are also enriched in this analysis. Again, this potentially represents lymphocyte apoptosis in connection

with an enriched apoptotic pathway in Figure 3C. Neutrophil specific S100A8/A9 complexes are also enriched (Figure 4C). Overall, these results suggest PBMC-based subcellular level changes associated with the viral integration in immune cells and associated pathophysiology.

Recovery from coronavirus infection is associated with increased cytolytic activity and IFN-γ but not increased B-cell levels

In order to gain insight into the cellular dynamics of the SARS-CoV-2 immune response, we calculated immune signature scores from Rooney et al(*24*) (see *Methods*). In this case, we separated the coronavirus patient samples into those with acute-like or recovering-like disease and compared these with healthy control samples. As expected, we observed that the innate immune system involving macrophages and neutrophils was highly active in the acute-like patients, suggesting that they were the first to encounter the coronavirus. with these decreasing in recovering patients (Figure 5A).

Perhaps most interestingly, there was a significant increase in NK cells, cytolytic activity, and plasmacytoid dendritic cells (pDCs) in recovering-like patients compared to acute-like patients (Figure 5A). It is noteworthy that the absolute levels of *CD8*$^+$ T cells and co-stimulating helper T cells are not different between recovering-like and acute-like patients (Figure 5A). This result suggests that the *CD8*$^+$ T cells are potentially activated (cytolytic) in the recovering patients. These results from SARS patients were confirmed using PBMC from COVID-19 samples (Figure 5B). There were no differences in B cells between both types of patients and healthy individuals. Intriguingly, there was an increase in interferon (IFN)-γ type-II genes only in the recovering-like patients. These genes were low in both acute-like patients and healthy volunteers (Figure 5A). These results suggest that T-cell responses may be pivotal in successful response to SARS-CoV-2 infection, consistent with the recent study by Grifoni et. al. which found SARS-CoV-2 reactive T cells in 70% of convalescent COVID-19 patients.

Next, we examined the differential expression of major histocompatibility complex (MHC) class-I and class-II HLA that may reflect antigen presentation to and/or activation of *CD4*$^+$/*CD8*$^+$ T cells, and whose levels are increased by IFN-γ (Figure 5C). Among the MHC class-II HLA genes, there were four of them that were lower in acute-like patients compared to recovering-like and/or healthy individuals (borderline significance with nominal $p<0.1$ due to low sample size). While most of the MHC class-II HLA genes showed no difference in expression between the acute-like and recovering-like patients, *HLA-DQA1* showed an increasing trend in the recovering-like patients towards the healthy individuals and low level

in acute-like patients. Similarly, *HLA-F* is the only MHC class I HLA gene that showed a trend akin to *HLA-DQA1* (Figure 5C). On the other hand, there was no change in the antigen processing machinery (data not shown). However, all these speculations from limited data warrant further validation, and their functional immunological significance is currently unclear.

To confirm these analyses of immunologic composition, we performed two additional independent analyses. Hypergeometric test enrichment analysis of MSigDB's C7 immune signatures showed similar conclusions that innate immunity and myeloid (neutrophils and macrophages) cells are upregulated in acute-like patients, whereas adaptive immunity is downregulated in these patients (Figure 5D). Similarly, analysis using the BioGPS database – gene sets(*39*) demonstrated an increased enrichment of $CD33^+$ myeloid and $CD14^+$ monocytes associated with upregulated genes in our CoV expression signature while $CD8^+$ and $CD4^+$ T cells showed up along with enrichment of $CD56^+$ NK cells and $CD19^+$ B in the downregulated fraction (Figure 6A-B). Compared to healthy individuals, these cells are under-represented in CoV-Up-genes or lower in coronavirus-infected patients (Figure 6B).

Coronavirus-infected patients' PBMC reveals genes that overlap with known disease symptoms

Based on Supplementary Figure 3, we reasoned that the disease symptoms from coronavirus infection may be similar to other diseases. To study this, we applied enrichment analysis to study the overlap of genes between CoV-Up-gene signature and other disease symptoms. We found that CoV-Up-gene signatures was enriched for various diseases, including septicaemia, pneumonia, lung disease, arthritis, cystic fibrosis, thalassemia, pre-eclampsia, bacterial infections, asthma, acute coronary syndrome and others (Figure 7A). The overlap of CoV-Up-gene signatures and those genes from selected diseases - septicaemia, pneumonia, lung diseases, arthritis, and cystic fibrosis are shown in Figure 7B. These results suggest that the disease symptoms due to coronavirus may be complex and highly variable and may affect patients with pre-existing disease conditions as recently reported (*7*). Further refinement of COVID-engine may help to identify hidden/silent pre-existing symptoms and develop an effective personalized COVID-19 treatment strategy using PBMC.

Discussion

The clinical course of COVID-19 patients remains enigmatic, and no treatment options exist with proven efficacy(*40*). The variety of clinical presentations of this disease has alarmed healthcare providers across the globe. The rampant spread of COVID-19 during asymptomatic stages is attributed to the high SARS-CoV-2 shedding in the upper respiratory tract(*41*). We reasoned that the blood, in real-time, provides changes occurring in immune and other cells and potentially infected tissues in PBMC, thereby acting as a potential remote biosensor of highly complex system-wide changes. There is no systematic study performed to our knowledge that attempts to use PBMC samples to understand the system-wide changes along with the disease symptoms in COVID-19 patients. This type of study will have the strength to distinguish systemic changes during acute and recovering stages of the patient's infection. This, in the future with further refinement and validation, may support the development of personalized prognostic biomarkers and may provide the opportunity to save patients who are most likely to die of the disease.

In this study, we performed a comprehensive analysis using publicly available blood cell RNA profiles from SARS and COVID-19 patients and cross-validated with patients with other infections or healthy individuals. Using IS/ML approaches in our COVID-engine, we were able to develop a blood cell RNA profile-based gene signatures that are differentially expressed at different stages of infection (acute vs. recovering). In addition, our COVID-engine provided end-to-end hierarchical and comprehensive analysis describing infection-associated changes in genes, pathways, networks, subcellular, and cellular, covering almost the whole system. This 'integrated' analysis can help understand which other disease-related symptoms could manifest in COVID-19 patients.

Our results represent that the innate immune system associated with increased neutrophils, macrophages, and monocytes with potential cytokine storm (including the expression of *IL6*, *IL8* (*CXCL8*) and *CCL13*) is high in CoV-Up-gene signatures and specifically in acute-like patients. Macrophages and monocytes are known to serve as factories for viral replication in other disease conditions (*29*). These changes in immune cells may also be connected to increased neutrophil counts in these patients (*42*). The cytokine storm-related to innate immune changes may be linked to changes in angiogenesis and coagulation, suggesting a potential relationship between inflammation, thromboembolism(*43*), and coagulation(*44*). While there is no change in overall $CD8^+$ T cell population between patients in their acute-like vs. recovering-like stages, the change in cytolytic activity, pDCs, and NK cells suggests that the adaptive immunity is a late event represented in patients recuperating from this disease

as represented in multiple reports(*45*). Congruently, this is associated with lower innate immunity in recovering-like patients than acute-like patients and associated with increased expression of anti-viral genes *OAS2* and *IL16*.

Similarly, the high *HLA-F* gene, which is known to be associated with the interaction between CD4 T cells and NK cells to inactivate human immunodeficiency virus (HIV)(*46*), in recovering-like patients suggests the anti-viral effect in these individuals. Remarkably, this supposedly anti-viral CD4 T and NK cells along with B cells, are low in COVID-19 patients and are associated with low HLA-DR monocytes in these patients with severe respiratory failure (*31, 47, 48*). This report corroborates with our results that *HLA-DRA* gene and all the above three cell types are low in acute-like patients. Specifically, B-cell-based adaptive immunity seems to vary among patients and mostly low in COVID-19 patients with severe respiratory failure (*31, 49*). Our data suggest that this may impact the development of effective vaccines for this infection. In addition, CoV-Up-gene signatures was high in methicillin-resistant *Staphylococcus aureus* (Supplementary Figure 2), which, similar to COVID-19, colonizes upper respiratory track and causes pneumonia. It is interesting to note that there is no vaccine for *S. aureus* infection (*50*). While there are more changes in T cells than B cells, it may be interesting to consider T cell therapy for COVID-19 patients (*51*). In our study, there are disease links and potential comorbidities (Figure 7) that has evidence from recent reports (*7*).

There are a number of questions that arose from this study that could be of relevance in tackling the current pandemic. How does coronavirus infection downregulate the adaptive immune system? Is the dysregulation of the immune system described causally linked with clinical outcomes? Does the dysregulated immune system alert the body to respond, and how? The dysregulated immune system could alert the host response to produce an active adaptive immune response, which typically takes 10-14 days. Our findings may suggest that individuals with activated, appropriate immune responses, especially with increased INF-γ type-II responses and cytolytic activity, which may also serve as biomarkers with further validation, maybe on their way to recovery from symptoms. Moreover, those patients' incapable of such progression may have multiorgan failures that maybe represented in our data.

Although our study may be timely for the current pandemic, there are limitations. We have performed analysis using publicly available small number of coronavirus-based training (n=10) and test (n=47) samples from less annotated datasets and limited clinical data, which may be appreciated provided the current global lock-down scenario. Also, the acute-like and recovering-like patients may overlap with the symptomatic and asymptomatic patients,

respectively, described in the original publication from where the training dataset was derived(*16*). Further validation was curtailed due to lack of associated clinical data, which is difficult to obtain in the current scenario.

In conclusion, PBMC has information related to infection status, immune states, disease aggression, severity, and disease symptoms that are likely going to be manifested due to coronavirus infection and COVID-19 disease (Figure 8).

Contributions

AS conceived the idea, developed the concept, collected the data, developed the IS/ML pipeline, interpreted the results, performed the experiments, supervised the project and wrote the manuscript. NK co-developed the concept, interpreted the results, co-supervised the project and wrote the manuscript. SD and CPE interpreted certain analysis. TB and KR assisted to further develop the immune analysis and interpreting the data. AM and DK critically interpreted the results and assisted in writing the manuscript.

Conflict of Interest

None of the authors have any conflict of interests with respect to the current study. AS serves as a private consultant to develop biomarkers for a different immune-disorder related disease for a company.

Acknowledgements

We thank Dr. Krisha Desai and Ms. Aasia Hussain for critically reading the manuscript.

Funding

There is no funding to declare.

Figure legends

Figure 1. COVID-engine and PBMC-based gene signatures show association with SARS and COVID-19. Schematic showing the identification of PBMC RNA gene signatures genes associated with disease staging, and hierarchical modeling of genes, pathways, networks, sub-cellular contents, cells, and disease symptoms using COVID-engine (A). This figure was prepared using Servier Medical Art (https://smart.servier.com) under a Creative Commons Attribution 3.0 Unported License (http://creativecommons.org/licenses/by/3.0/). Heatmap showing 290 gene signatures indicator genes in 10 SARS patients and 4 healthy individuals. Both CoV-Up-gene signatures (169 genes) and CoV-Down-gene signatures (121 genes) are shown (B). CoV-Up-gene signatures scores (C) and CoV-Down-gene signatures scores (D) and their association with SARS-CoV-2-infected (COVID-19; n=3) patients and healthy individuals (n=3; datasets from Raghunathan et al(*16*)). For C) and D) statistical significance was not considered due to low sample size. CoV-Up-gene signatures scores (E) and CoV-Down-gene signatures scores (F) and their association with acute and recovering SARS-CoV-infected patients (n=25; 44 samples), bacteria-infected patients (n=16) and healthy individuals (n=9; datasets from Lee et al(*17*)). Kruskal-Wallis statistical with $p < 0.001$ for E) and F). Acute vs. recovering gene signatures from Lee et al(*17*) predicted acute-like and recovering-like SARS-CoV-infected patients (n=10) from Raghunathan et al(*16*) using NTP method and cosine distance measure (G).

Figure 2. Significantly represented key genes at different stages of coronavirus infection. Expression levels of genes: *ACE2* (A) cytokines *IL6* (B), *TNF* (C), and hypoxia related *LDHA* (D) and *LDHB* (E) in acute-like and recovering-like SARS patients and healthy individuals. Kruskal-Wallis statistical with nominal p values reported. ns represents not significant. Multiple testing was not done due to low sample size.

Figure 3. Highly represented signalling pathways/processes in coronavirus-infected patients. Expression levels of genes related to immune cell chemoattractants *CXCL8* and *CCL13* (A), genes involved in T cells and suppression of viral replication *OAS2* and *IL16* (B) in acute-like and recovering-like SARS patients and healthy individuals. Kruskal-Wallis statistical with nominal p values reported. Multiple testing was not done due to low sample size. Enrichment statistical analysis using CoV-Up-gene signatures and pathways/processes based on genesets from MSigDB's Hallmarks database (C).

Figure 4. Highly represented related networks of molecular, cellular and development pathways show cytokine (storm) network and innate immunity in coronavirus-infected patients. REACTOME(*34*) database-based connection of different but related pathways that

were enriched in infected patients using network plots (A). Similarly, KEGG(*52*) pathways network showing enrichment of different diseases and infection related pathways (B). Enrichment statistical analysis using CoV-Up-gene signatures and pathways/processes based on genesets from COMPARTMENTS database(*53*) (C).

Figure 5. Immune cells, their activities and HLA types in stage-specific coronavirus-infected patients. Enrichment of immune cell scores using ssGSEA analysis and specific immune signatures from Rooney *et al*. representing specific immune cells and their activities in acute-like (n=7), recovering-like (n=3) SARS-CoV patients and control healthy individuals (=4) (A), and COVID-19 patients and control healthy individuals (B). HLA profiles in acute-like, recovering-like SARS patients and control healthy individuals (C). Kruskal-Wallis statistical nominal p values were significant for all reported plots. Similar enrichment analysis of CoV-Up-gene signatures signatures using MSigDB's C7 immune gensets (D).

Figure 6. PBMC-based gene signatures show association with subset of immune cells in coronavirus-infected patients. Enrichment of subsets of immune cell genes in multiple BioGPS – gene portal system(*39*) using CoV-Up-gene signatures (A) and CoV-Down-gene signatures (B).

Figure 7. PBMC-based gene signatures identify known and novel disease symptoms in coronavirus-infected patients. Enrichment analysis of disease-based genesets from DisGeNET(*54*) (A). Top 5 diseases and associated PBMC genes (B).

Figure 8. Schematic summarizing the gene signatures and their relevance at systems-level to disease stages, and symptoms for personalized COVID-19 medicine.

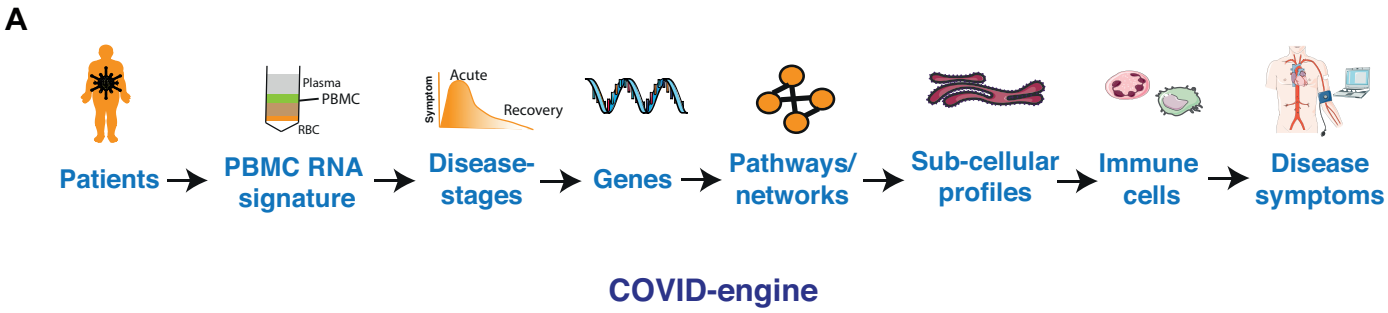
References:

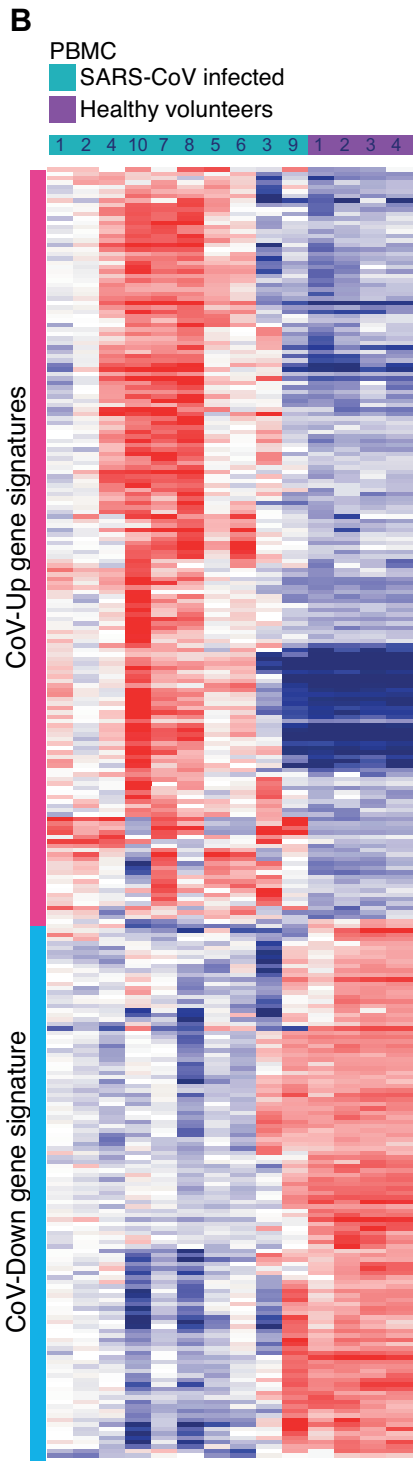
1. WHO. (World Health Organization, 2020).
2. J. E. L. Wong, Y. S. Leo, C. C. Tan, COVID-19 in Singapore-Current Experience: Critical Global Issues That Require Attention and Action. *Jama*, (2020).
3. E. Kindler, V. Thiel, F. Weber, Interaction of SARS and MERS Coronaviruses with the Antiviral Interferon Response. *Adv Virus Res* **96**, 219-243 (2016).
4. S. Shokri, S. Mahmoudvand, R. Taherkhani, F. Farshadpour, Modulation of the immune response by Middle East respiratory syndrome coronavirus. *J Cell Physiol* **234**, 2143-2151 (2019).
5. M. Kikkert, Innate Immune Evasion by Human Respiratory RNA Viruses. *J Innate Immun* **12**, 4-20 (2020).
6. D. L. Swerdlow, L. Finelli, Preparation for Possible Sustained Transmission of 2019 Novel Coronavirus: Lessons From Previous Epidemics. *Jama* **323**, 1129-1130 (2020).
7. M. Wadman, J. Couzin-Frankel, J. Kaiser, C. Matacic, How does coronavirus kill? Clinicians trace a ferocious rampage through the body, from brain to toes. *Science* **368**, 356-360 (2020).

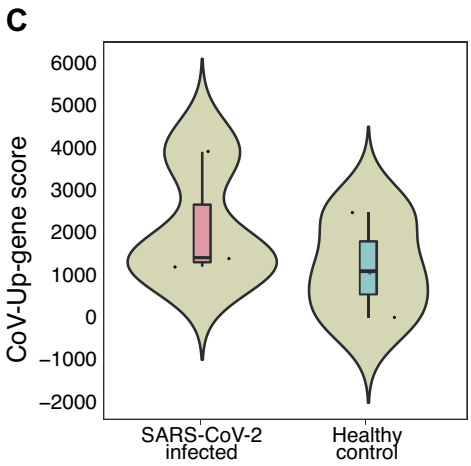
8. A. L. Phelan, R. Katz, L. O. Gostin, The Novel Coronavirus Originating in Wuhan, China: Challenges for Global Health Governance. *Jama*, (2020).
9. H. Nishiura *et al.*, Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int J Infect Dis*, (2020).
10. O. T. Ng *et al.*, SARS-CoV-2 Infection among Travelers Returning from Wuhan, China. *N Engl J Med* **382**, 1476-1478 (2020).
11. Y. Bai *et al.*, Presumed Asymptomatic Carrier Transmission of COVID-19. *Jama*, (2020).
12. L. Wang, Y. H. Gao, L. L. Lou, G. J. Zhang, The clinical dynamics of 18 cases of COVID-19 outside of Wuhan, China. *Eur Respir J*, (2020).
13. D. Chen *et al.*, Recurrence of positive SARS-CoV-2 RNA in COVID-19: A case report. *Int J Infect Dis* **93**, 297-299 (2020).
14. B. Shanmugaraj, A. Malla, W. Phoolcharoen, Emergence of Novel Coronavirus 2019-nCoV: Need for Rapid Vaccine and Biologics Development. *Pathogens* **9**, (2020).
15. J. Pang *et al.*, Potential Rapid Diagnostics, Vaccine and Therapeutics for 2019 Novel Coronavirus (2019-nCoV): A Systematic Review. *J Clin Med* **9**, (2020).
16. R. Reghunathan *et al.*, Expression profile of immune response genes in patients with Severe Acute Respiratory Syndrome. *BMC Immunol* **6**, 2 (2005).
17. Y. S. Lee *et al.*, Molecular signature of clinical severity in recovering patients with severe acute respiratory syndrome coronavirus (SARS-CoV). *BMC Genomics* **6**, 132 (2005).
18. V. G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences* **98**, 5116-5121 (2001).
19. H. Schwender, siggenes: Multiple Testing using SAM and Efron's Empirical Bayes Approaches. *R package version 1.62.0.*, (2020).
20. A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545-15550 (2005).
21. A. Federico, S. Monti, hypeR: an R package for geneset enrichment workflows. *Bioinformatics* **36**, 1307-1308 (2020).
22. Y. Hoshida, Nearest template prediction: a single-sample-based flexible class prediction with confidence assessment. *PLoS One* **5**, e15543 (2010).
23. E. Y. Chen *et al.*, Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
24. M. S. Rooney, S. A. Shukla, C. J. Wu, G. Getz, N. Hacohen, Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48-61 (2015).
25. Y. Xiong *et al.*, Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg Microbes Infect* **9**, 761-770 (2020).
26. O. Ramilo *et al.*, Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood* **109**, 2066-2077 (2007).
27. J. Pugin *et al.*, Diagnosis of ventilator-associated pneumonia by bacteriologic analysis of bronchoscopic and nonbronchoscopic "blind" bronchoalveolar lavage fluid. *Am Rev Respir Dis* **143**, 1121-1129 (1991).
28. M. Rutkowska-Zapala *et al.*, Human monocyte subsets exhibit divergent angiotensin I-converting activity. *Clin Exp Immunol* **181**, 126-132 (2015).
29. E. Nikitina, I. Larionova, E. Choinzonov, J. Kzhyshkowska, Monocytes and Macrophages as Viral Targets and Reservoirs. *Int J Mol Sci* **19**, (2018).
30. D. E. Gordon *et al.*, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, (2020).
31. E. J. Giamarellos-Bourboulis *et al.*, Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. *Cell Host Microbe*, (2020).

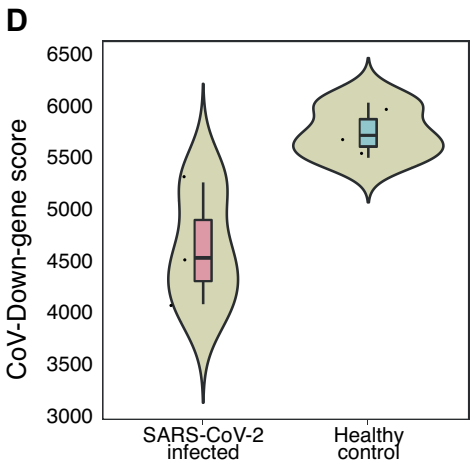
32. W. Ottestad, M. Seim, J. O. Mæhlen, COVID-19 with silent hypoxemia. *Tidsskr Nor Laegeforen* **140**, (2020).
33. J. Poissy *et al.*, Pulmonary Embolism in COVID-19 Patients: Awareness of an Increased Prevalence. *Circulation*, (2020).
34. B. Jassal *et al.*, The reactome pathway knowledgebase. *Nucleic Acids Res* **48**, D498-D503 (2020).
35. C. T. Esmon, J. Xu, F. Lupu, Innate immunity and coagulation. *J Thromb Haemost* **9 Suppl 1**, 182-188 (2011).
36. A. Mathian *et al.*, Clinical course of coronavirus disease 2019 (COVID-19) in a series of 17 patients with systemic lupus erythematosus under long-term treatment with hydroxychloroquine. *Ann Rheum Dis*, (2020).
37. A. M. Kell, M. Gale, Jr., RIG-I in RNA virus recognition. *Virology* **479-480**, 110-121 (2015).
38. M. Lorizate, H. G. Krausslich, Role of lipids in virus replication. *Cold Spring Harb Perspect Biol* **3**, a004820 (2011).
39. C. Wu, X. Jin, G. Tsueng, C. Afrasiabi, A. I. Su, BioGPS: building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Res* **44**, D313-316 (2016).
40. H. Ledford, How does COVID-19 kill? Uncertainty is hampering doctors' ability to choose treatments. *Nature* **580**, 311-312 (2020).
41. M. Gandhi, D. S. Yokoe, D. V. Havlir, Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid-19. *N Engl J Med*, (2020).
42. Y. Liu *et al.*, Neutrophil-to-lymphocyte ratio as an independent risk factor for mortality in hospitalized patients with COVID-19. *J Infect*, (2020).
43. T. Wang *et al.*, Attention should be paid to venous thromboembolism prophylaxis in the management of COVID-19. *Lancet Haematol* **7**, e362-e363 (2020).
44. R. J. Jose, A. Manuel, COVID-19 cytokine storm: the interplay between inflammation and coagulation. *Lancet Respir Med*, (2020).
45. M. Z. Tay, C. M. Poh, L. Renia, P. A. MacAry, L. F. P. Ng, The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol*, (2020).
46. W. F. Garcia-Beltran *et al.*, Open conformers of HLA-F are high-affinity ligands of the activating NK-cell receptor KIR3DS1. *Nat Immunol* **17**, 1067-1074 (2016).
47. D. Zhang *et al.*, COVID-19 infection induces readily detectable morphological and inflammation-related phenotypic changes in peripheral blood monocytes, the severity of which correlate with patient outcome. *medRxiv*, 2020.2003.2024.20042655 (2020).
48. Y. Zhou *et al.*, Aberrant pathogenic GM-CSF⁺ T cells and inflammatory CD14⁺CD16⁺ monocytes in severe pulmonary syndrome patients of a new coronavirus. *bioRxiv*, 2020.2002.2012.945576 (2020).
49. X. Cao, COVID-19: immunopathology and its implications for therapy. *Nat Rev Immunol* **20**, 269-270 (2020).
50. D. Parker, A live vaccine to Staphylococcus aureus infection. *Virulence* **9**, 700-702 (2018).
51. V. Bachanova *et al.*, Chimeric Antigen Receptor T Cell Therapy During the COVID-19 Pandemic. *Biol Blood Marrow Transplant*, (2020).
52. M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353-D361 (2017).
53. J. X. Binder *et al.*, COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database (Oxford)* **2014**, bau012 (2014).
54. J. Pinero *et al.*, DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833-D839 (2017).

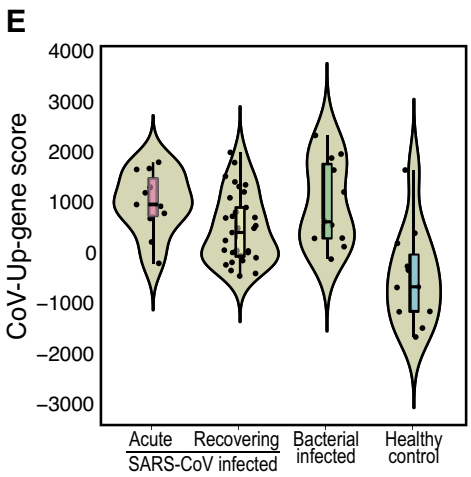
Figure 1

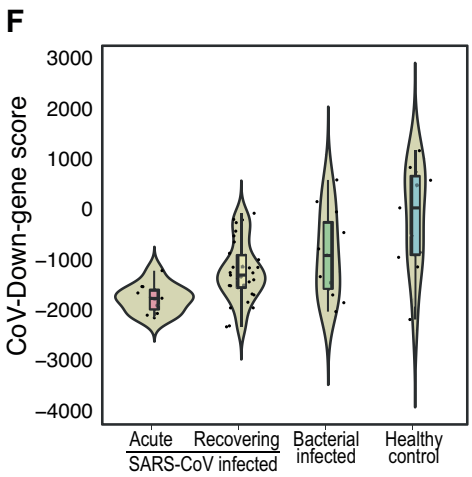

A
Plasma
PBMC
RBC
Acute
Recovery
Symptom
Patients
PBMC RNA signature
Disease-stages
Genes
Pathways/ networks
Sub-cellular profiles
Immune cells
Disease symptoms
COVID-engine


B
PBMC
SARS-CoV infected
Healthy volunteers
1 2 4 10 7 8 5 6 3 9 1 2 3 4
CoV-Up gene signatures
CoV-Down gene signature


C
CoV-Up-gene score
6000
5000
4000
3000
2000
1000
0
−1000
−2000
SARS-CoV-2 infected
Healthy control


D
CoV-Down-gene score
6500
6000
5500
5000
4500
4000
3500
3000
SARS-CoV-2 infected
Healthy control


E
CoV-Up-gene score
4000
3000
2000
1000
0
−1000
−2000
−3000
Acute
Recovering
SARS-CoV infected
Bacterial infected
Healthy control


F
CoV-Down-gene score
3000
2000
1000
0
−1000
−2000
−3000
−4000
Acute
Recovering
SARS-CoV infected
Bacterial infected
Healthy control

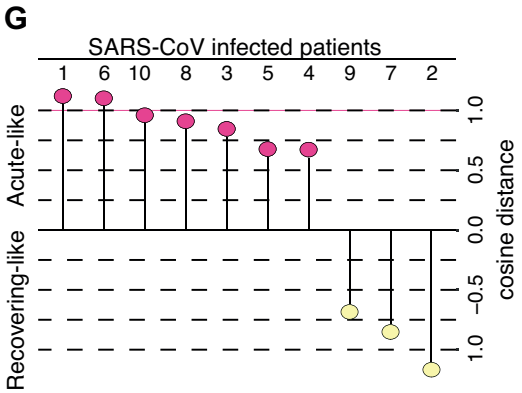

G
SARS-CoV infected patients
1 6 10 8 3 5 4 9 7 2
Acute-like
Recovering-like
1.0
0.5
0.0
−0.5
1.0
cosine distance

Figure 2

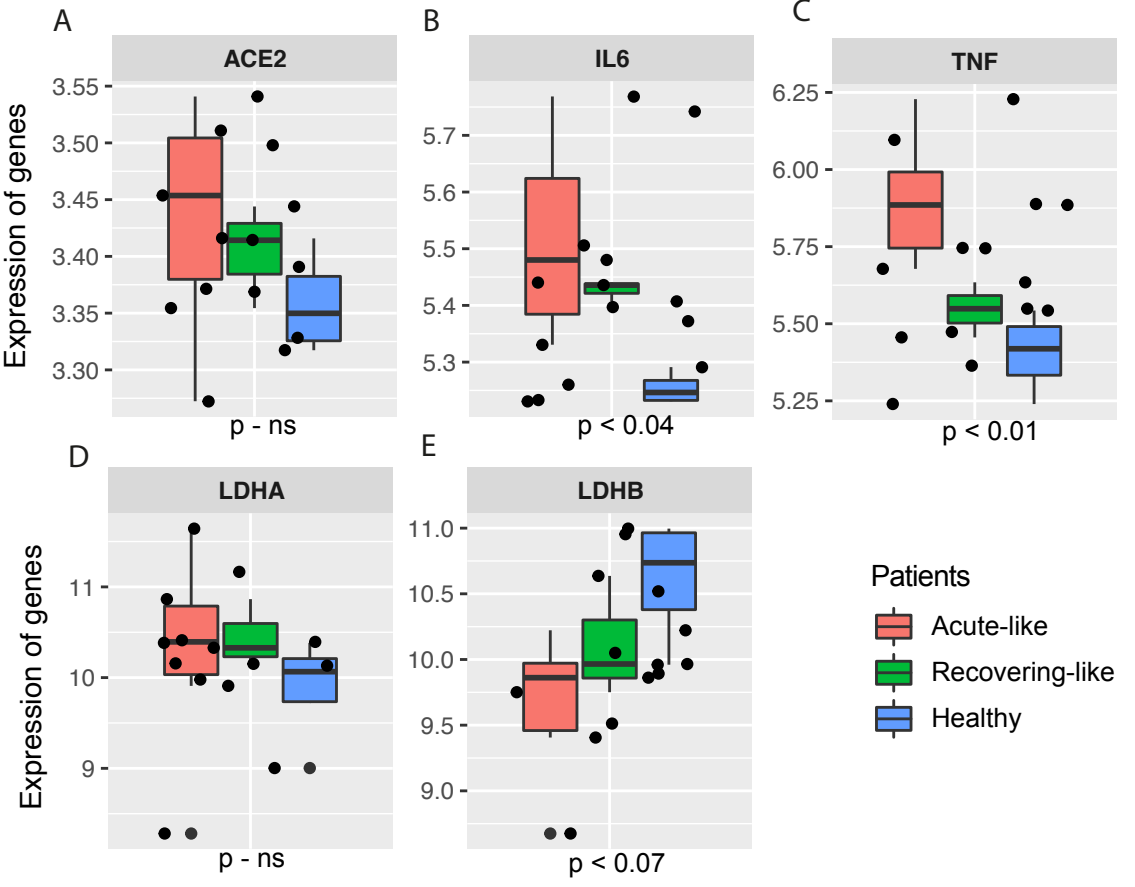

A
ACE2
3.55
3.50
3.45
3.40
3.35
3.30
Expression of genes
p - ns
B
IL6
5.7
5.6
5.5
5.4
5.3
p < 0.04
C
TNF
6.25
6.00
5.75
5.50
5.25
p < 0.01
D
LDHA
11
10
9
Expression of genes
p - ns
E
LDHB
11.0
10.5
10.0
9.5
9.0
p < 0.07
Patients
Acute-like
Recovering-like
Healthy

Figure 3

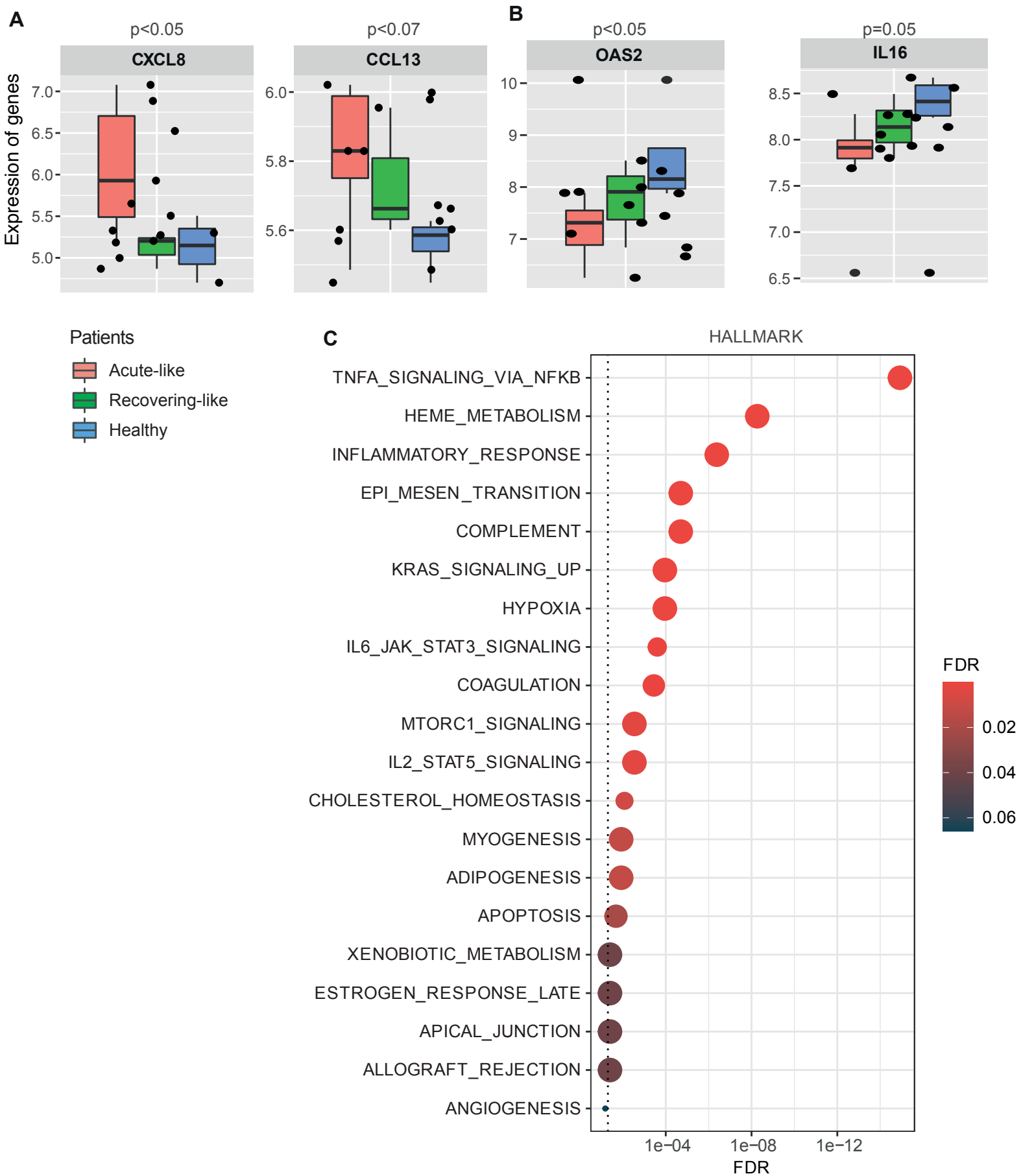

A
p<0.05
CXCL8
p<0.07
CCL13
Expression of genes
7.0
6.5
6.0
5.5
5.0
6.0
5.8
5.6
B
p<0.05
OAS2
p=0.05
IL16
10
9
8
7
8.5
8.0
7.5
7.0
6.5
Patients
Acute-like
Recovering-like
Healthy
C
HALLMARK
TNFA_SIGNALING_VIA_NFKB
HEME_METABOLISM
INFLAMMATORY_RESPONSE
EPI_MESEN_TRANSITION
COMPLEMENT
KRAS_SIGNALING_UP
HYPOXIA
IL6_JAK_STAT3_SIGNALING
COAGULATION
MTORC1_SIGNALING
IL2_STAT5_SIGNALING
CHOLESTEROL_HOMEOSTASIS
MYOGENESIS
ADIPOGENESIS
APOPTOSIS
XENOBIOTIC_METABOLISM
ESTROGEN_RESPONSE_LATE
APICAL_JUNCTION
ALLOGRAFT_REJECTION
ANGIOGENESIS
1e−04
1e−08
1e−12
FDR
FDR
0.02
0.04
0.06

Figure 4

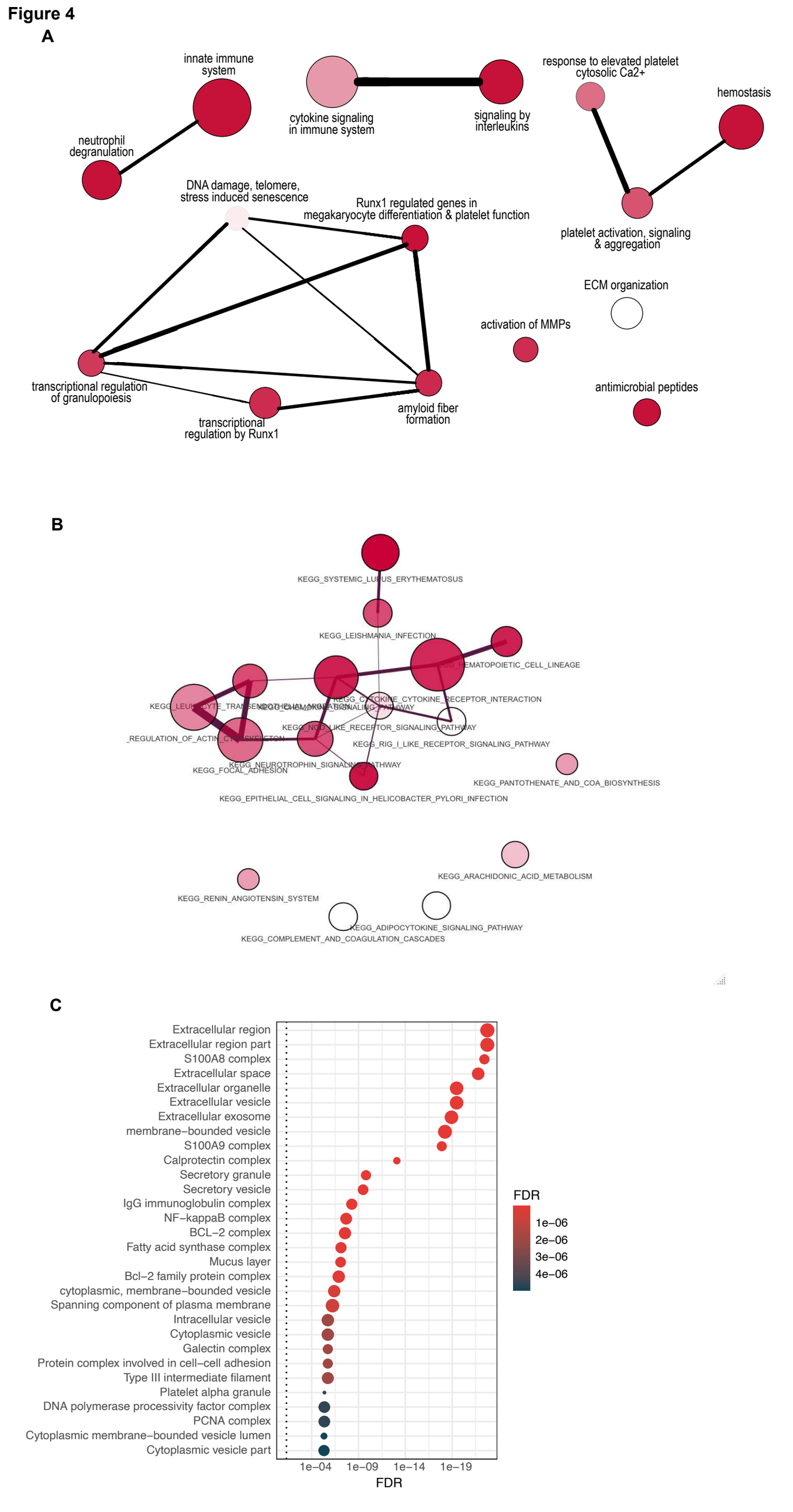

A
innate immune system
neutrophil degranulation
cytokine signaling in immune system
signaling by interleukins
response to elevated platelet cytosolic Ca2+
hemostasis
platelet activation, signaling & aggregation
DNA damage, telomere, stress induced senescence
Runx1 regulated genes in megakaryocyte differentiation & platelet function
transcriptional regulation of granulopoiesis
transcriptional regulation by Runx1
amyloid fiber formation
ECM organization
activation of MMPs
antimicrobial peptides
B
KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS
KEGG_LEISHMANIA_INFECTION
KEGG_HEMATOPOIETIC_CELL_LINEAGE
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION
KEGG_NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY
KEGG_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY
KEGG_NEUROTROPHIN_SIGNALING_PATHWAY
KEGG_FOCAL_ADHESION
KEGG_PANTOTHENATE_AND_COA_BIOSYNTHESIS
KEGG_EPITHELIAL_CELL_SIGNALING_IN_HELICOBACTER_PYLORI_INFECTION
KEGG_ARACHIDONIC_ACID_METABOLISM
KEGG_RENIN_ANGIOTENSIN_SYSTEM
KEGG_ADIPOCYTOKINE_SIGNALING_PATHWAY
KEGG_COMPLEMENT_AND_COAGULATION_CASCADES
C
Extracellular region
Extracellular region part
S100A8 complex
Extracellular space
Extracellular organelle
Extracellular vesicle
Extracellular exosome
membrane–bounded vesicle
S100A9 complex
Calprotectin complex
Secretory granule
Secretory vesicle
IgG immunoglobulin complex
NF–kappaB complex
BCL–2 complex
Fatty acid synthase complex
Mucus layer
Bcl–2 family protein complex
cytoplasmic, membrane–bounded vesicle
Spanning component of plasma membrane
Intracellular vesicle
Cytoplasmic vesicle
Galectin complex
Protein complex involved in cell–cell adhesion
Type III intermediate filament
Platelet alpha granule
DNA polymerase processivity factor complex
PCNA complex
Cytoplasmic membrane–bounded vesicle lumen
Cytoplasmic vesicle part
1e–04
1e–09
1e–14
1e–19
FDR
FDR
1e–06
2e–06
3e–06
4e–06

Figure 5

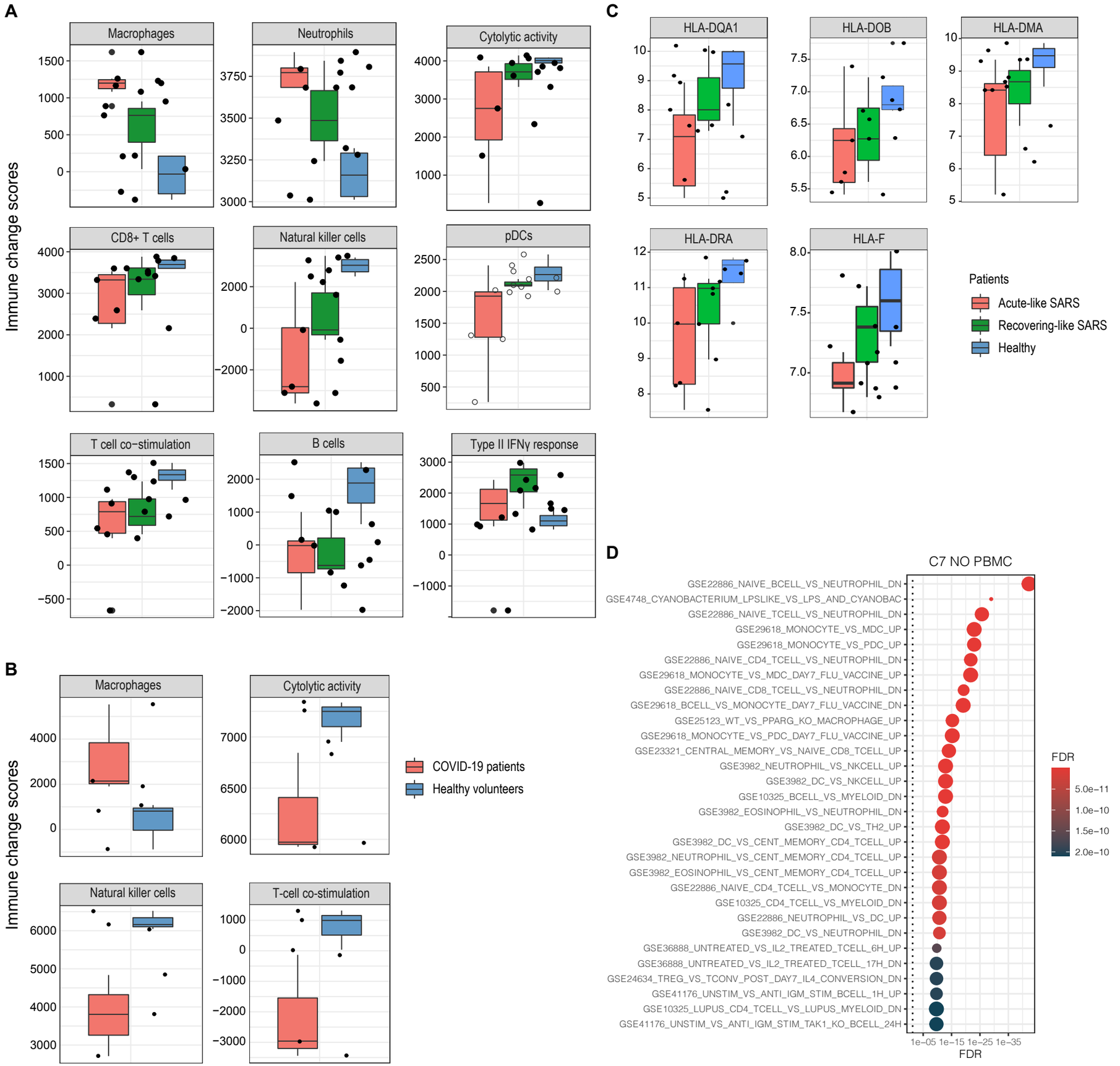

A
Macrophages
Neutrophils
Cytolytic activity
CD8+ T cells
Natural killer cells
pDCs
T cell co−stimulation
B cells
Type II IFNγ response
Immune change scores
C
HLA-DQA1
HLA-DOB
HLA-DMA
HLA-DRA
HLA-F
Patients
Acute-like SARS
Recovering-like SARS
Healthy
B
Macrophages
Cytolytic activity
Natural killer cells
T-cell co-stimulation
Immune change scores
COVID-19 patients
Healthy volunteers
D
C7 NO PBMC
GSE22886_NAIVE_BCELL_VS_NEUTROPHIL_DN
GSE4748_CYANOBACTERIUM_LPSLIKE_VS_LPS_AND_CYANOBAC
GSE22886_NAIVE_TCELL_VS_NEUTROPHIL_DN
GSE29618_MONOCYTE_VS_MDC_UP
GSE29618_MONOCYTE_VS_PDC_UP
GSE22886_NAIVE_CD4_TCELL_VS_NEUTROPHIL_DN
GSE29618_MONOCYTE_VS_MDC_DAY7_FLU_VACCINE_UP
GSE22886_NAIVE_CD8_TCELL_VS_NEUTROPHIL_DN
GSE29618_BCELL_VS_MONOCYTE_DAY7_FLU_VACCINE_DN
GSE25123_WT_VS_PPARG_KO_MACROPHAGE_UP
GSE29618_MONOCYTE_VS_PDC_DAY7_FLU_VACCINE_UP
GSE23321_CENTRAL_MEMORY_VS_NAIVE_CD8_TCELL_UP
GSE3982_NEUTROPHIL_VS_NKCELL_UP
GSE3982_DC_VS_NKCELL_UP
GSE10325_BCELL_VS_MYELOID_DN
GSE3982_EOSINOPHIL_VS_NEUTROPHIL_DN
GSE3982_DC_VS_TH2_UP
GSE3982_DC_VS_CENT_MEMORY_CD4_TCELL_UP
GSE3982_NEUTROPHIL_VS_CENT_MEMORY_CD4_TCELL_UP
GSE3982_EOSINOPHIL_VS_CENT_MEMORY_CD4_TCELL_UP
GSE22886_NAIVE_CD4_TCELL_VS_MONOCYTE_DN
GSE10325_CD4_TCELL_VS_MYELOID_DN
GSE22886_NEUTROPHIL_VS_DC_UP
GSE3982_DC_VS_NEUTROPHIL_DN
GSE36888_UNTREATED_VS_IL2_TREATED_TCELL_6H_UP
GSE36888_UNTREATED_VS_IL2_TREATED_TCELL_17H_DN
GSE24634_TREG_VS_TCONV_POST_DAY7_IL4_CONVERSION_DN
GSE41176_UNSTIM_VS_ANTI_IGM_STIM_BCELL_1H_UP
GSE10325_LUPUS_CD4_TCELL_VS_LUPUS_MYELOID_DN
GSE41176_UNSTIM_VS_ANTI_IGM_STIM_TAK1_KO_BCELL_24H
1e−05 1e−15 1e−25 1e−35
FDR
FDR
5.0e−11
1.0e−10
1.5e−10
2.0e−10

Figure 6

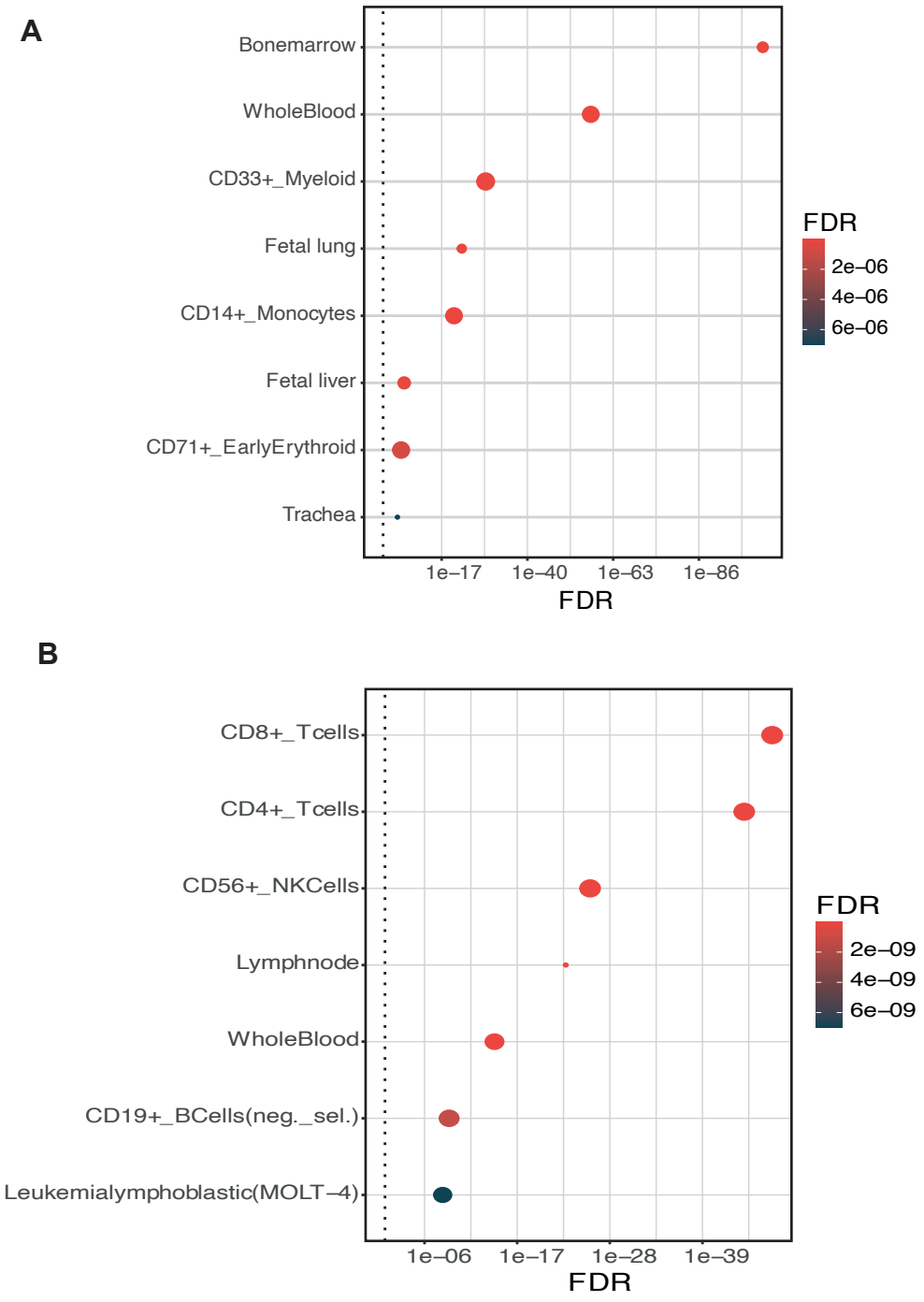

A
Bonemarrow
WholeBlood
CD33+_Myeloid
Fetal lung
CD14+_Monocytes
Fetal liver
CD71+_EarlyErythroid
Trachea
1e−17
1e−40
1e−63
1e−86
FDR
FDR
2e−06
4e−06
6e−06
B
CD8+_Tcells
CD4+_Tcells
CD56+_NKCells
Lymphnode
WholeBlood
CD19+_BCells(neg._sel.)
Leukemialymphoblastic(MOLT−4)
1e−06
1e−17
1e−28
1e−39
FDR
FDR
2e−09
4e−09
6e−09

Figure 7

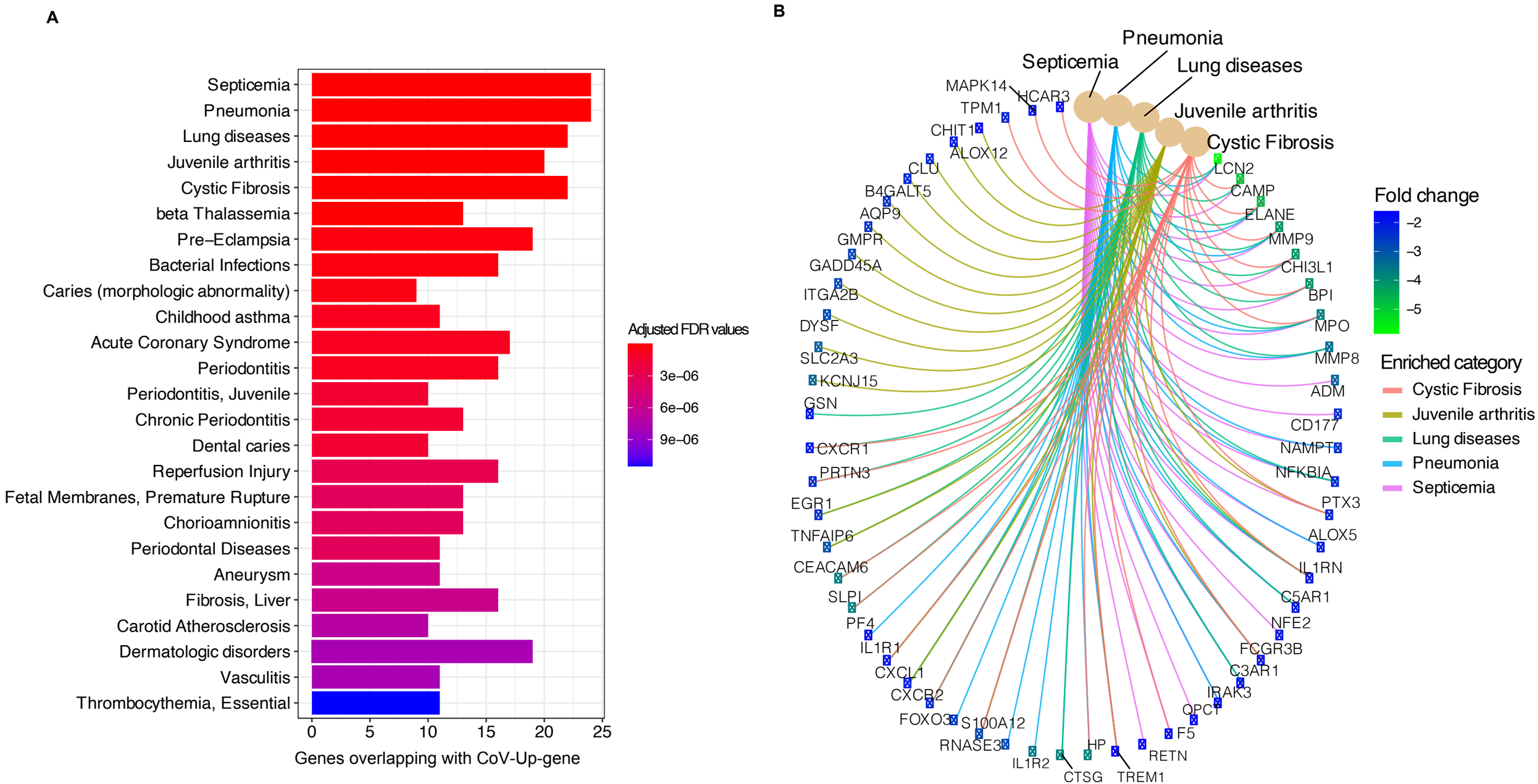
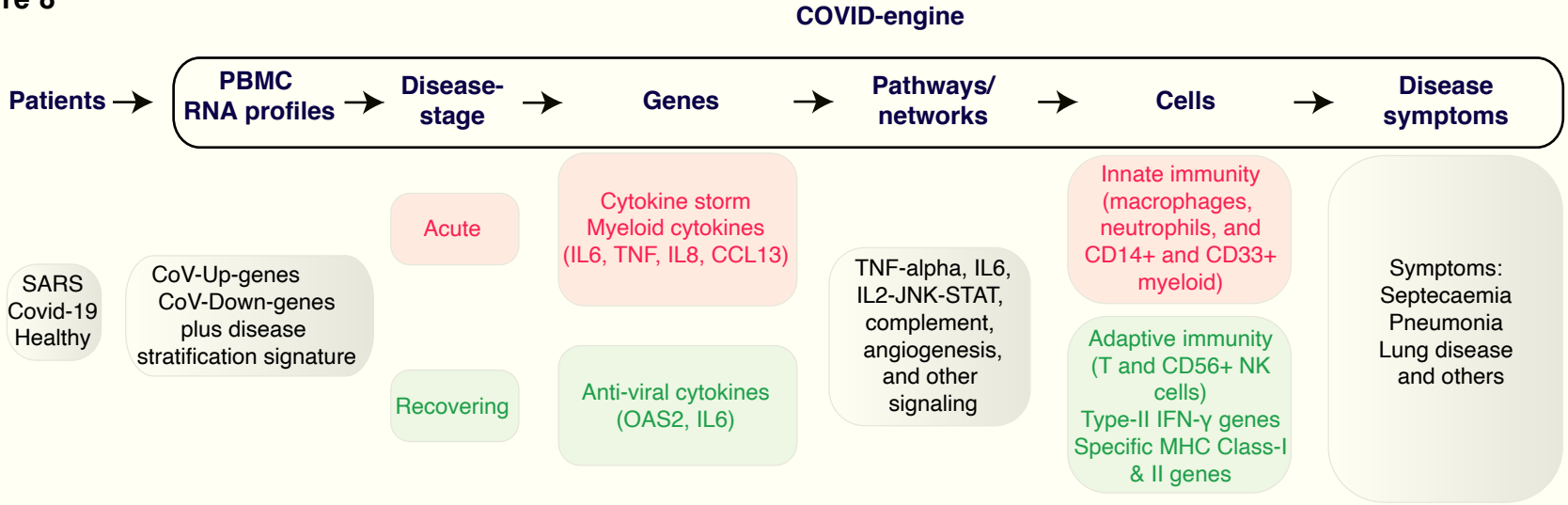

A
Septicemia
Pneumonia
Lung diseases
Juvenile arthritis
Cystic Fibrosis
beta Thalassemia
Pre–Eclampsia
Bacterial Infections
Caries (morphologic abnormality)
Childhood asthma
Acute Coronary Syndrome
Periodontitis
Periodontitis, Juvenile
Chronic Periodontitis
Dental caries
Reperfusion Injury
Fetal Membranes, Premature Rupture
Chorioamnionitis
Periodontal Diseases
Aneurysm
Fibrosis, Liver
Carotid Atherosclerosis
Dermatologic disorders
Vasculitis
Thrombocythemia, Essential
0
5
10
15
20
25
Genes overlapping with CoV-Up-gene
Adjusted FDR values
3e–06
6e–06
9e–06
B
Pneumonia
Septicemia
Lung diseases
Juvenile arthritis
Cystic Fibrosis
MAPK14
HCAR3
TPM1
CHIT1
ALOX12
CLU
B4GALT5
AQP9
GMPR
GADD45A
ITGA2B
DYSF
SLC2A3
KCNJ15
GSN
CXCR1
PRTN3
EGR1
TNFAIP6
CEACAM6
SLPI
PF4
IL1R1
CXCL1
CXCR2
FOXO3
S100A12
RNASE3
IL1R2
CTSG
HP
TREM1
RETN
F5
QPCT
IRAK3
C3AR1
FCGR3B
NFE2
C5AR1
IL1RN
ALOX5
PTX3
NFKBIA
NAMPT
CD177
ADM
MMP8
MPO
BPI
CHI3L1
MMP9
ELANE
CAMP
LCN2
Fold change
–2
–3
–4
–5
Enriched category
Cystic Fibrosis
Juvenile arthritis
Lung diseases
Pneumonia
Septicemia

Figure 8


COVID-engine
Patients
PBMC RNA profiles
Disease-stage
Genes
Pathways/ networks
Cells
Disease symptoms
SARS Covid-19 Healthy
CoV-Up-genes CoV-Down-genes plus disease stratification signature
Acute
Recovering
Cytokine storm Myeloid cytokines (IL6, TNF, IL8, CCL13)
Anti-viral cytokines (OAS2, IL6)
TNF-alpha, IL6, IL2-JNK-STAT, complement, angiogenesis, and other signaling
Innate immunity (macrophages, neutrophils, and CD14+ and CD33+ myeloid)
Adaptive immunity (T and CD56+ NK cells) Type-II IFN-γ genes Specific MHC Class-I & II genes
Symptoms: Septecaemia Pneumonia Lung disease and others

A blood-based comprehensive and systems-level analysis of disease aggression, immune suppression and symptoms in COVID-19 patients

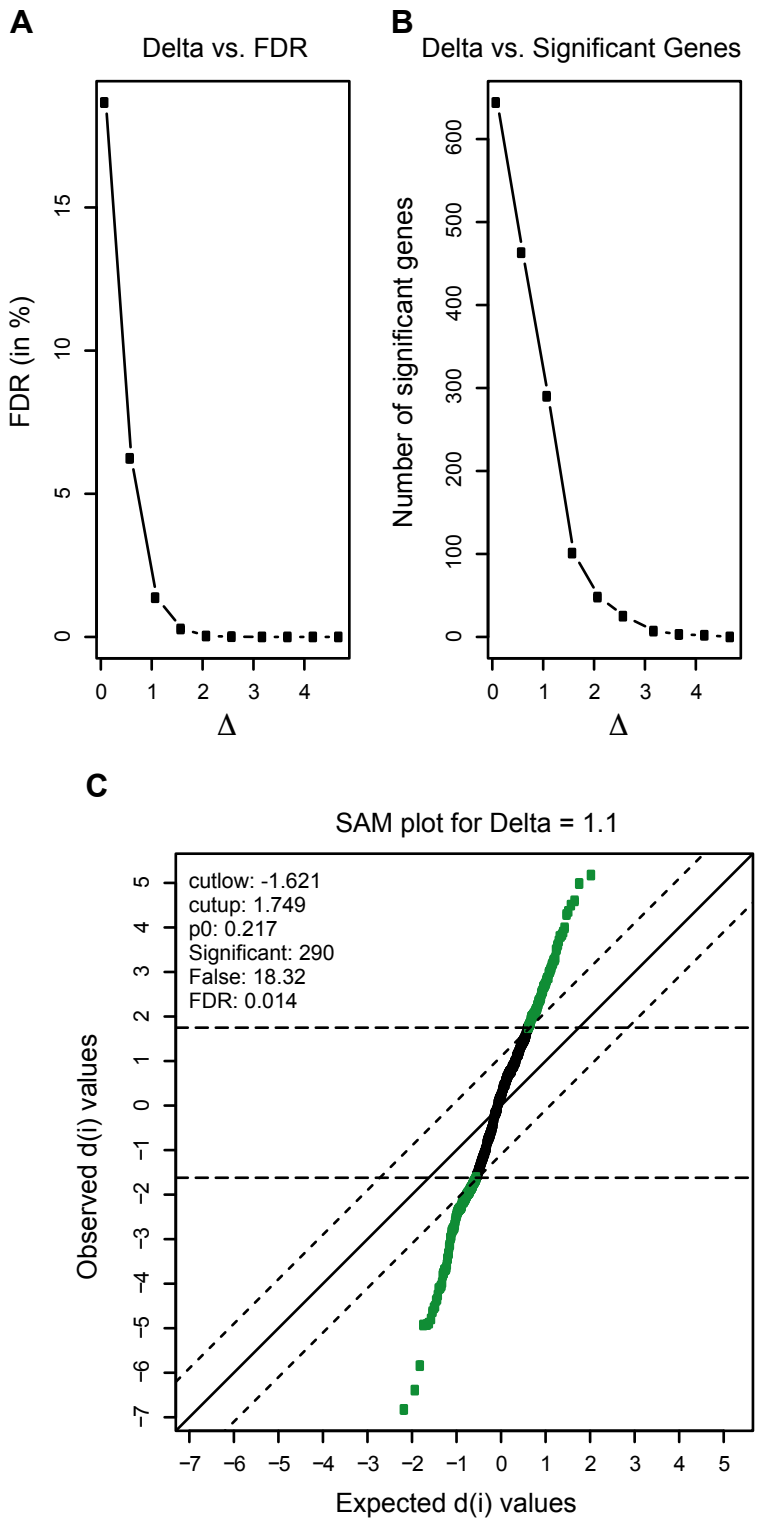
Supplementary Information

Supplementary Figure 1. SAM analysis output for CoV gene signature identification. A) Delta vs. FDR plot. B) Delta vs. Significant genes. C) SAM plot showing significant differentially expressed genes.

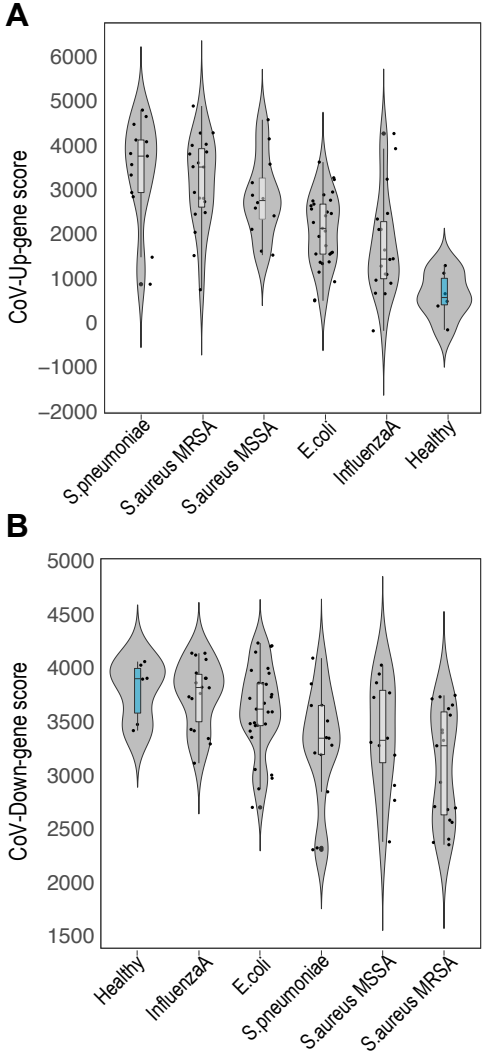
Supplementary Figure 2. PBMC show association with bacterial and other viral infections. CoV-Up-gene scores (A) and CoV-Down-gene scores (B) and their association with patients affected by different bacterial and viral infections and healthy individuals. Transcriptome data for 144 samples for this analysis was from Ramilo *et al*[22]. Kruskal-Wallis statistical with nominal $p < 0.0001$.

Supplementary Figure 3. Enrichment of PBMC from other diseases. A) Enrichment analysis of CoV-Up-gene signatures using MSigDB's C7 immune gensets.

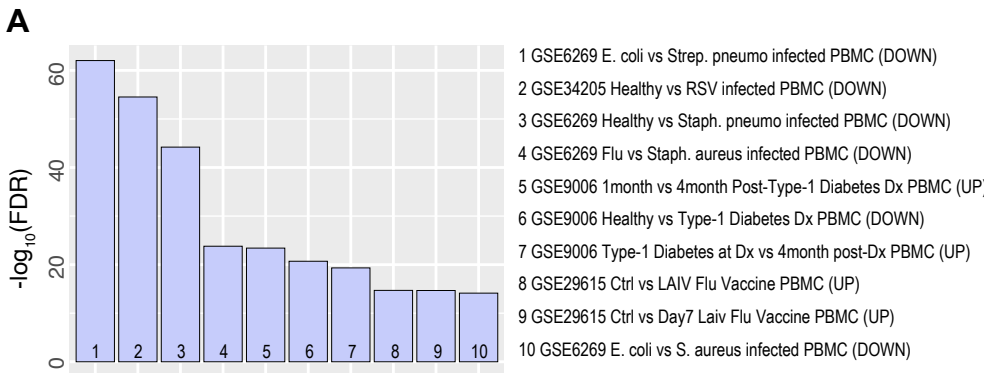
Supplementary Figure-1



Supplementary Figure-2



Supplementary Figure-3



Figures

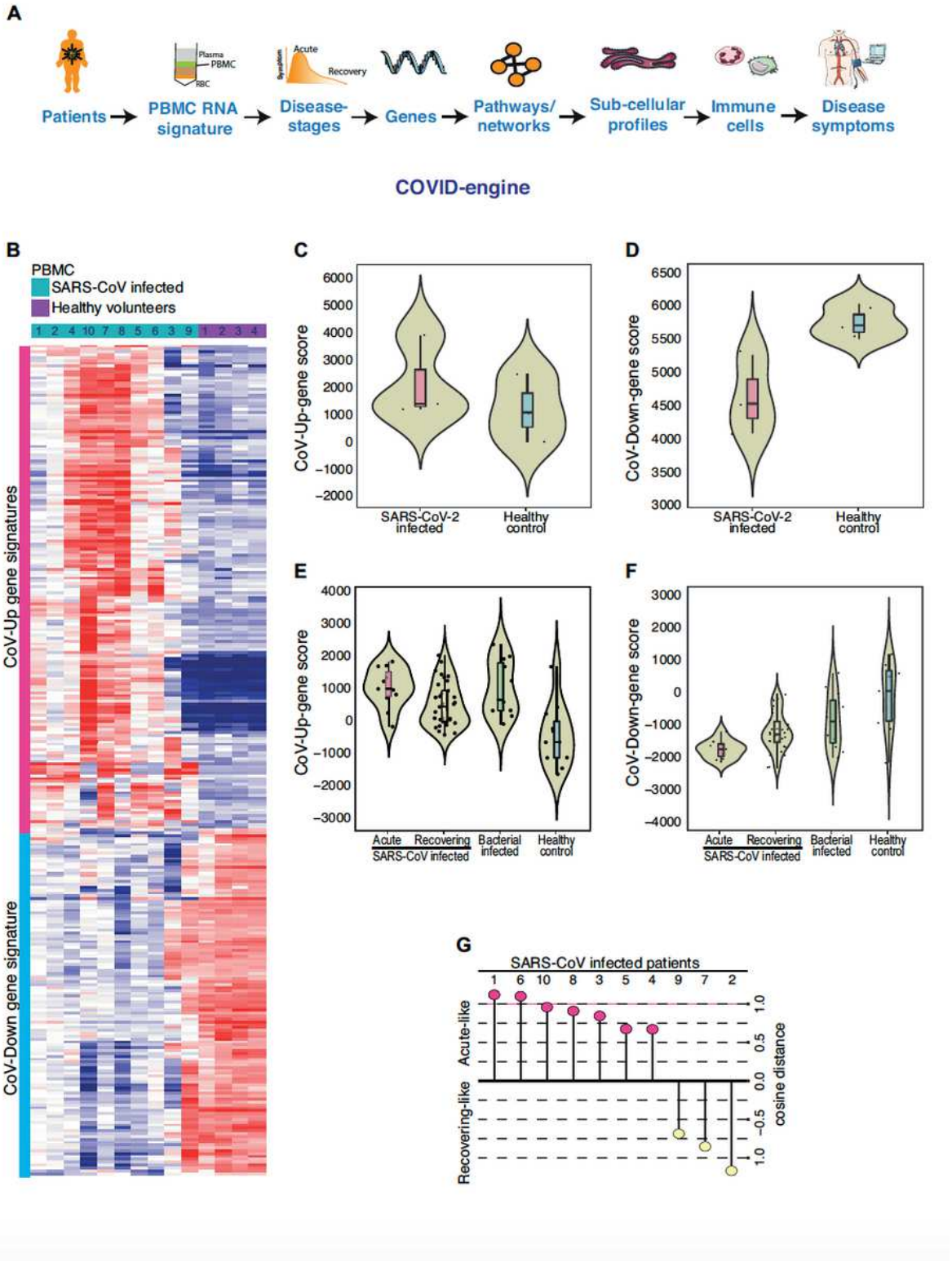


Figure 1

COVID-engine and PBMC-based gene signatures show association with SARS and COVID-19. Schematic showing the identification of PBMC RNA gene signatures genes associated with disease staging, and hierarchical modeling of genes, pathways, networks, sub-cellular contents, cells, and disease symptoms

using COVID-engine (A). This figure was prepared using Servier Medical Art (https://smart.servier.com) under a Creative Commons Attribution 3.0 Unported License (http://creativecommons.org/licenses/by/3.0/). Heatmap showing 290 gene signatures indicator genes in 10 SARS patients and 4 healthy individuals. Both CoV-Up-gene signatures (169 genes) and CoV-Down-gene signatures (121 genes) are shown (B). CoV-Up-gene signatures scores (C) and CoV-Down-gene signatures scores (D) and their association with SARS-CoV-2-infected (COVID-19; n=3) patients and healthy individuals (n=3; datasets from Raghunathan et al(16)). For C) and D) statistical significance was not considered due to low sample size. CoV-Up-gene signatures scores (E) and CoV-Down-gene signatures scores (F) and their association with acute and recovering SARS-CoV-infected patients (n=25; 44 samples), bacteria-infected patients (n=16) and healthy individuals (n=9; datasets from Lee et al(17)). Kruskal-Wallis statistical with $p < 0.001$ for E) and F). Acute vs. recovering gene signatures from Lee et al(17) predicted acute-like and recovering-like SARS-CoV-infected patients (n=10) from Raghunathan et al(16) using NTP method and cosine distance measure (G).

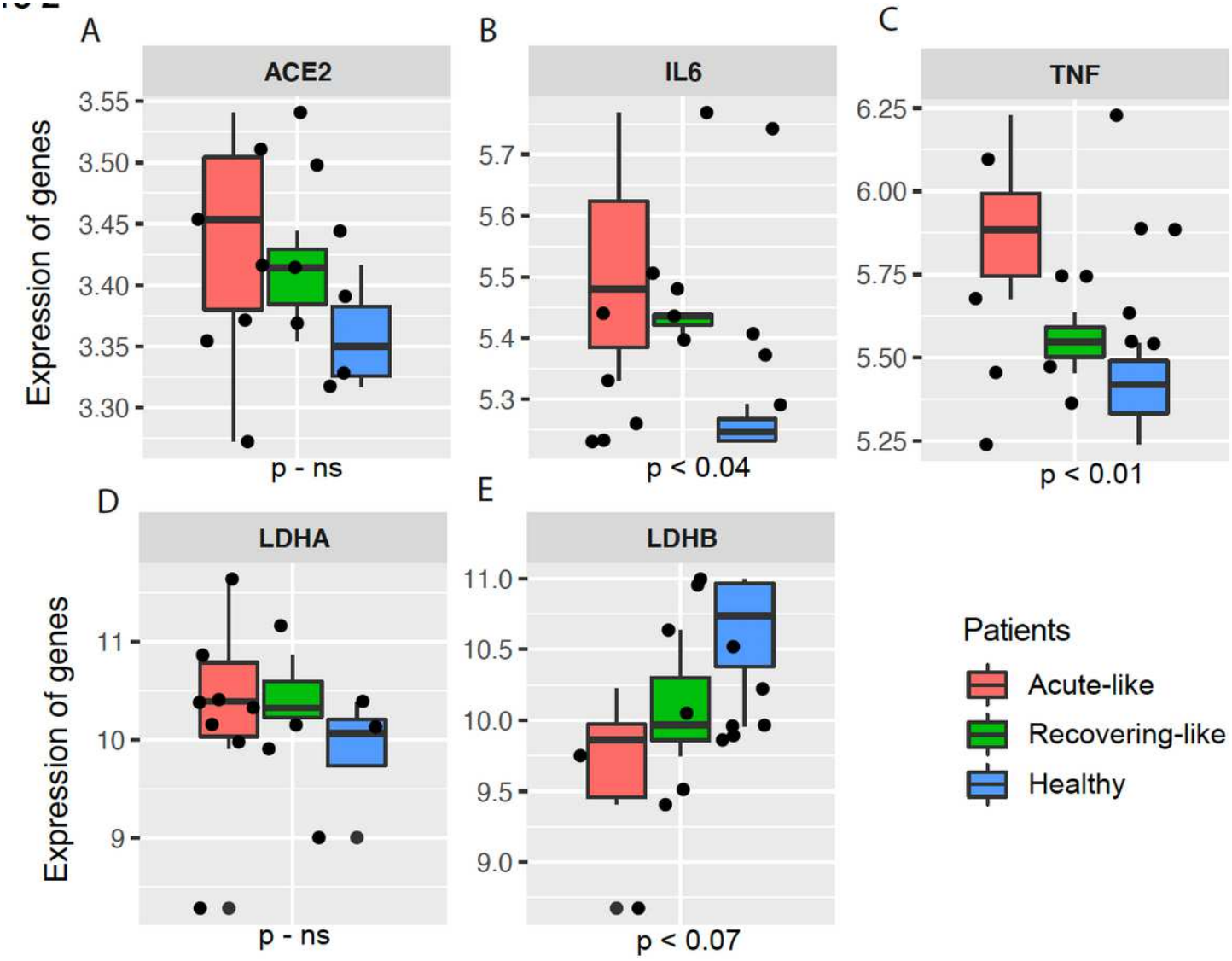


Figure 2

Significantly represented key genes at different stages of coronavirus infection. Expression levels of genes: ACE2 (A) cytokines IL6 (B), TNF (C), and hypoxia related LDHA (D) and LDHB (E) in acute-like and recovering-like SARS patients and healthy individuals. Kruskal-Wallis statistical with nominal p values reported. ns represents not significant. Multiple testing was not done due to low sample size.

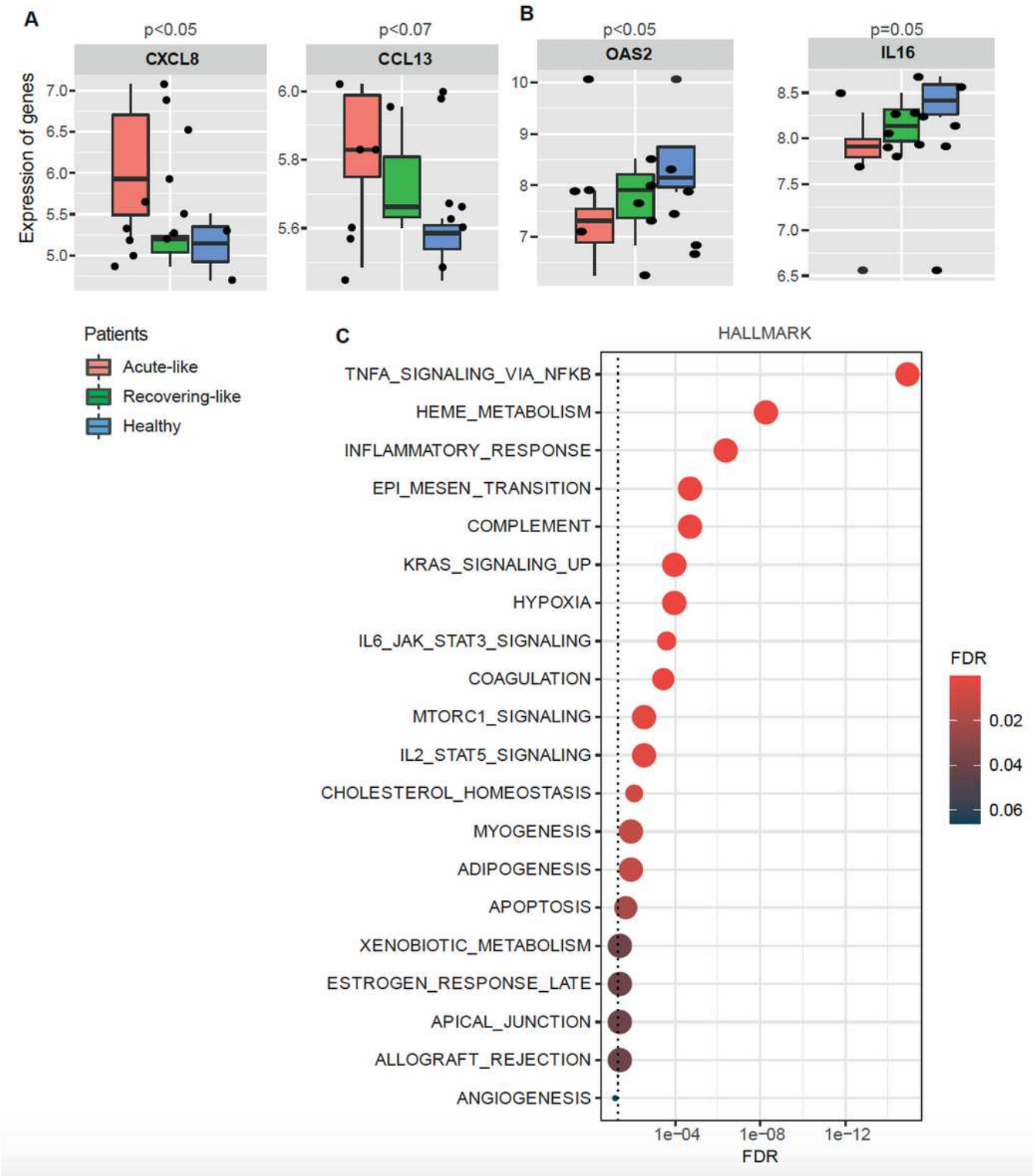


Figure 3

Highly represented signalling pathways/processes in coronavirus-infected patients. Expression levels of genes related to immune cell chemoattractants CXCL8 and CCL13 (A), genes involved in T cells and suppression of viral replication OAS2 and IL16 (B) in acute-like and recovering-like SARS patients and healthy individuals. Kruskal-Wallis statistical with nominal p values reported. Multiple testing was not done due to low sample size. Enrichment statistical analysis using CoV-Up-gene signatures and pathways/processes based on genesets from MSigDB's Hallmarks database (C).

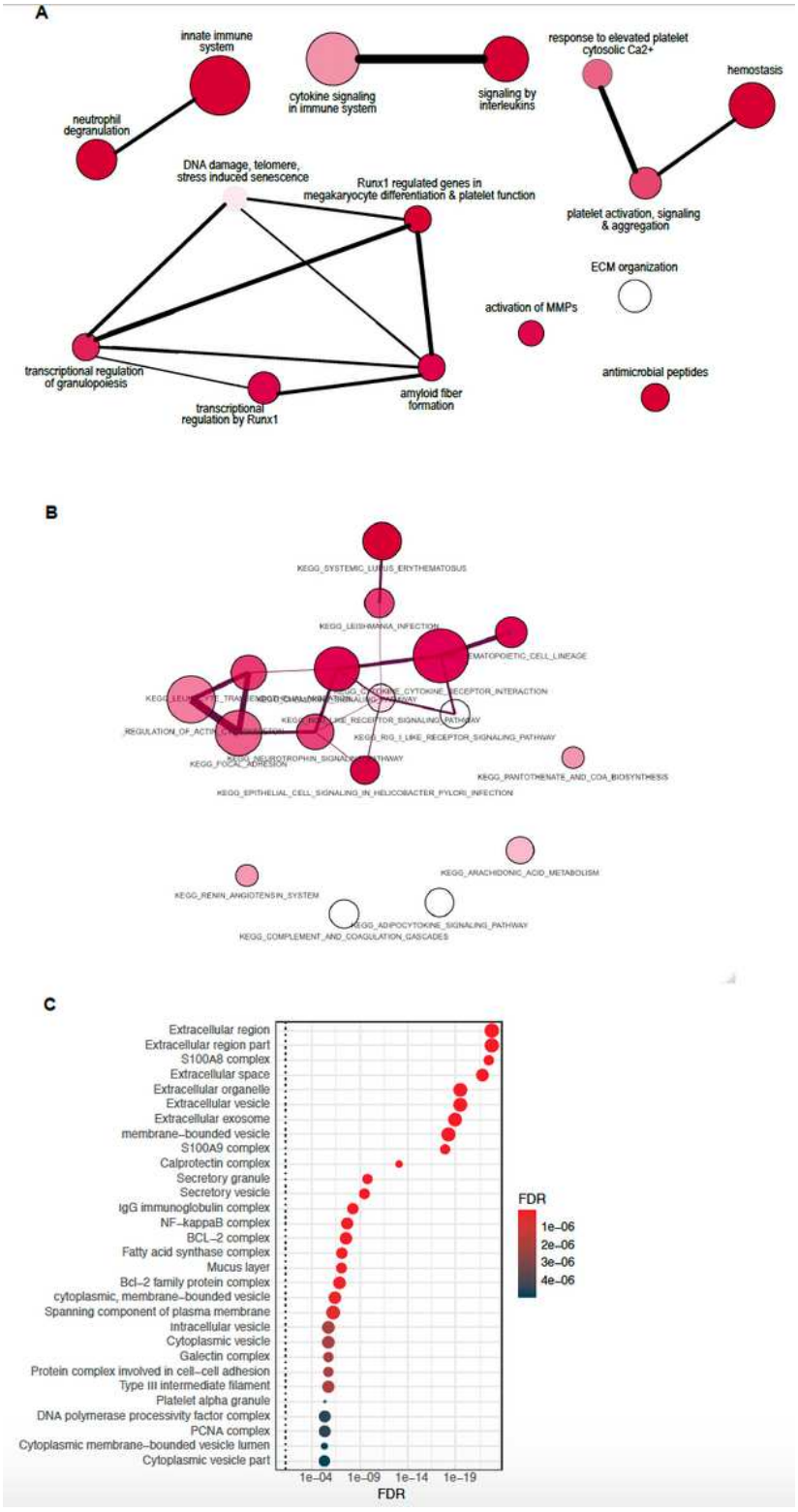


Figure 4

Highly represented related networks of molecular, cellular and development pathways show cytokine (storm) network and innate immunity in coronavirus-infected patients. REACTOME(34) database-based connection of different but related pathways that were enriched in infected patients using network plots (A). Similarly, KEGG(52) pathways network showing enrichment of different diseases and infection related pathways (B). Enrichment statistical analysis using CoV-Up-gene signatures and pathways/processes based on genesets from COMPARTMENTS database(53) (C).

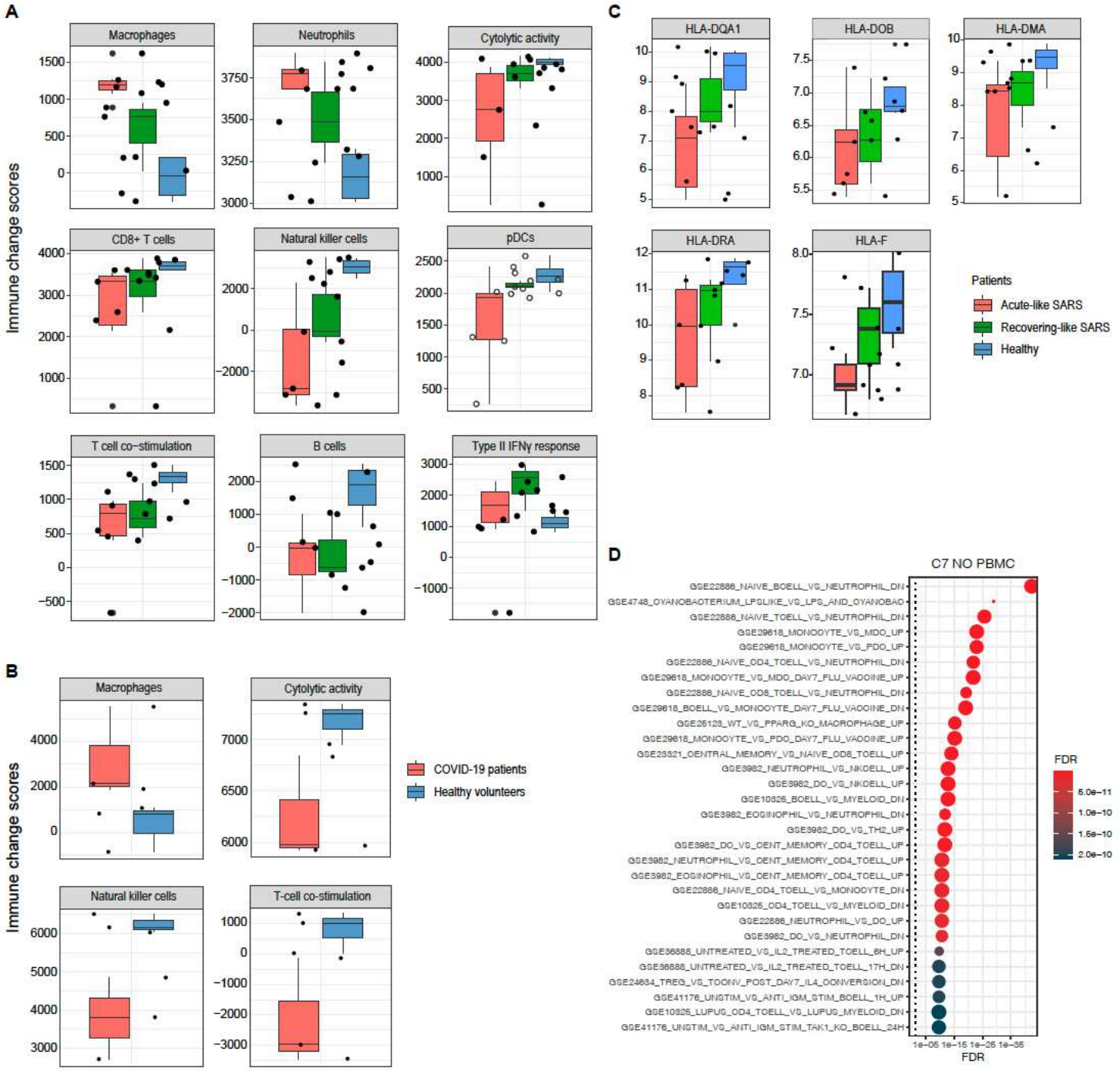
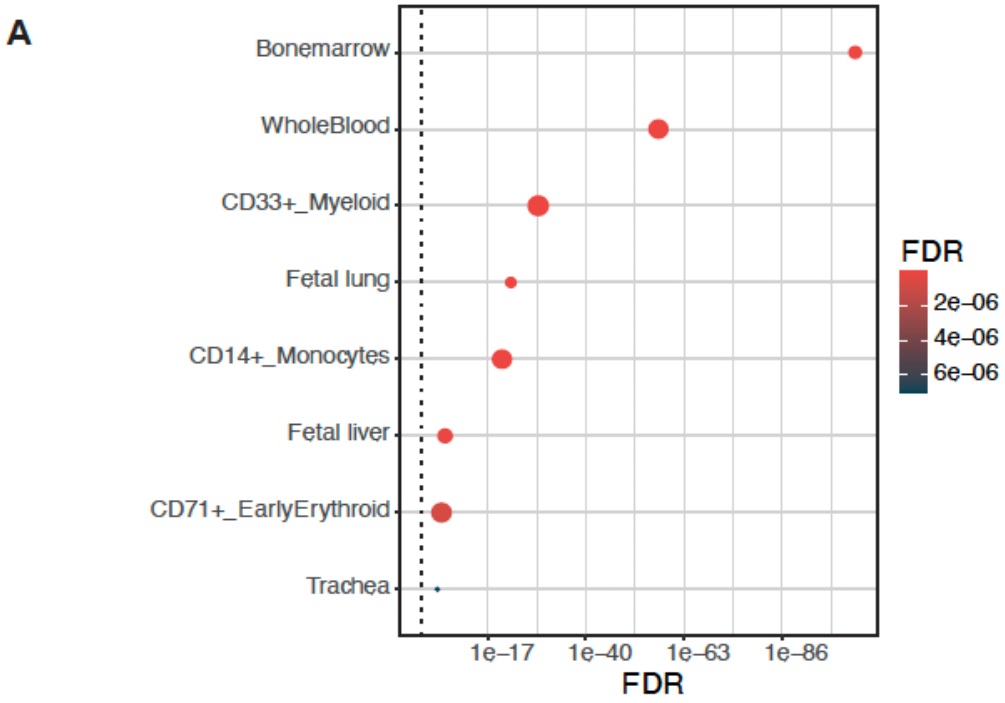


Figure 5

Immune cells, their activities and HLA types in stage-specific coronavirus-infected patients. Enrichment of immune cell scores using ssGSEA analysis and specific immune signatures from Rooney et al. representing specific immune cells and their activities in acute-like (n=7), recovering-like (n=3) SARS-CoV patients and control healthy individuals (=4) (A), and COVID-19 patients and control healthy individuals (B). HLA profiles in acute-like, recovering-like SARS patients and control healthy individuals (C). Kruskal-Wallis statistical nominal p values were significant for all reported plots. Similar enrichment analysis of CoV-Up-gene signatures signatures using MSigDB's C7 immune gensets (D).



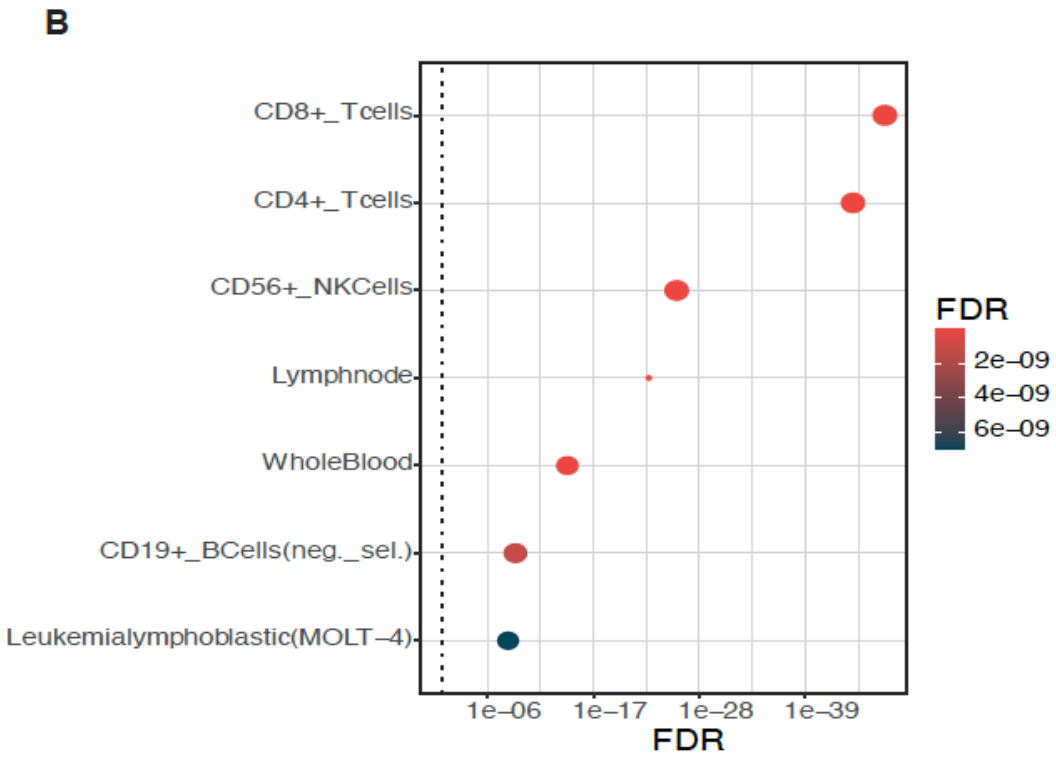


Figure 6

PBMC-based gene signatures show association with subset of immune cells in coronavirus-infected patients. Enrichment of subsets of immune cell genes in multiple BioGPS – gene portal system(39) using CoV-Up-gene signatures (A) and CoV-Down-gene signatures (B).

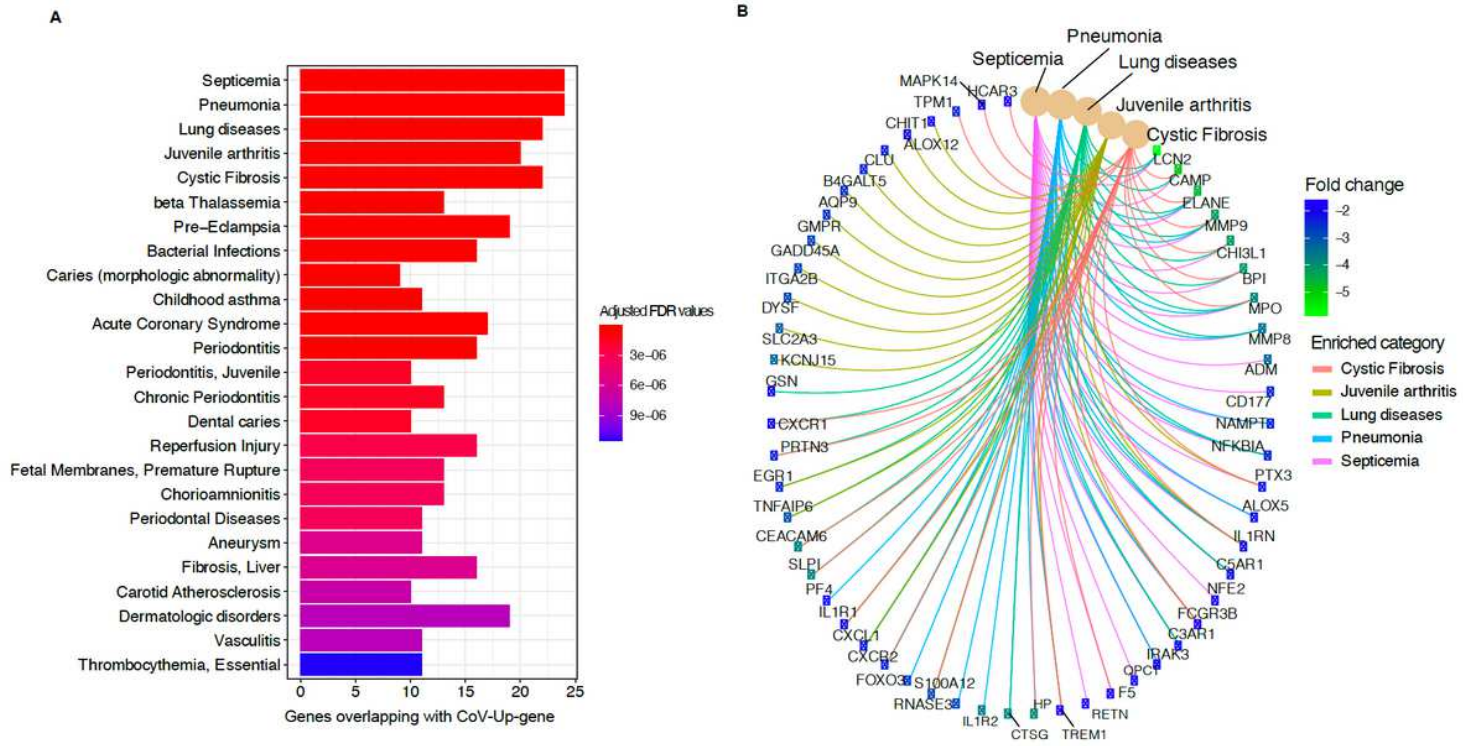


Figure 7

PBMC-based gene signatures identify known and novel disease symptoms in coronavirus-infected patients. Enrichment analysis of disease-based genesets from DisGeNET(54) (A). Top 5 diseases and associated PBMC genes (B).

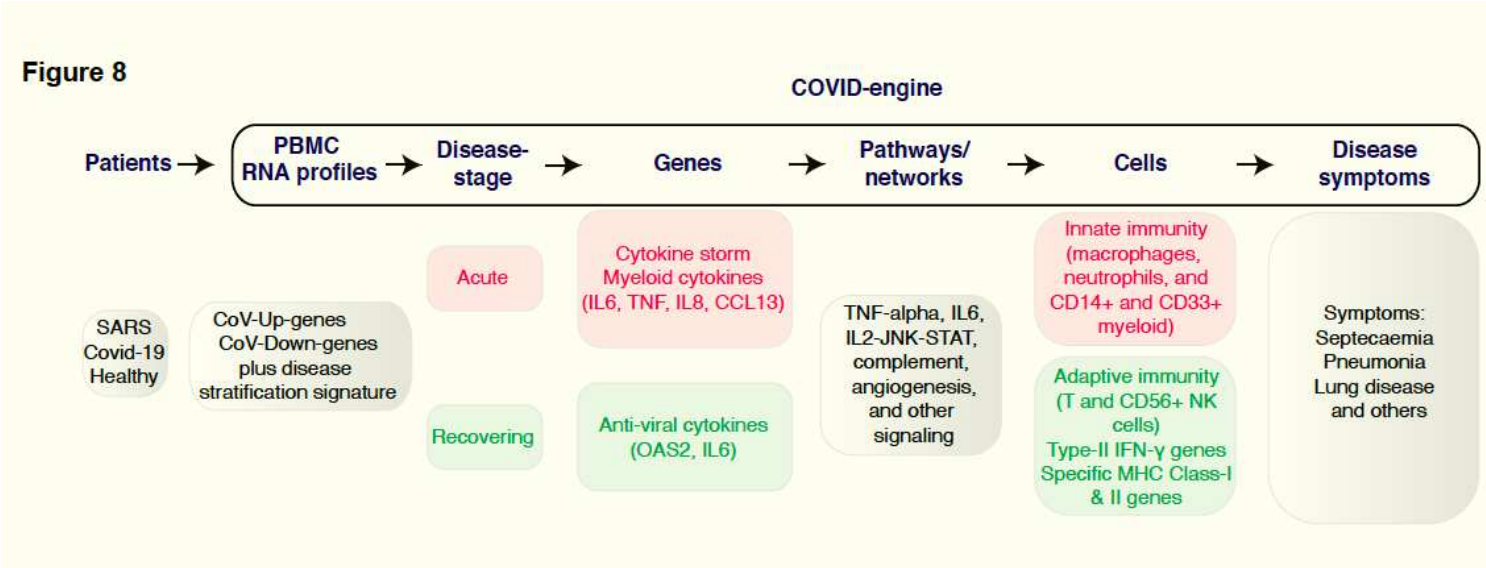


Figure 8

Schematic summarizing the gene signatures and their relevance at systems-level to disease stages, and symptoms for personalized COVID-19 medicine.

Supplementary Files

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

- SupplementalData.pdf