

Clinical and molecular characteristics of COVID-19 patients with persistent SARS-CoV-2 infection

Chaoyang Sun (✉ suncydoctor@gmail.com)

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology <https://orcid.org/0000-0003-2469-1638>

Junpeng Fan

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Jia Huang

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Ensong Guo

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Yu Fu

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Si Liu

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Rourou Xiao

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Chen Liu

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Funian Lu

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Tianyu Qin

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Chao He

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Zizhuo Wang

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Xu Qin

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Dianxing Hu

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Lixin You

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Xi Li

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Tian Wang

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Peng Wu

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Gang Chen

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Jianfeng Zhou

Department of Hematology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Kezhen Li

Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Research Article

Keywords: COVID-19 virus persistence, single-cell sequencing, clinical and molecular features

Posted Date: October 7th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-86940/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The clinical features, molecular characteristics, and immune responses of COVID-19 patients with persistent SARS-CoV-2 infection are not yet well described. In this study, we investigated the differences in clinical parameters, laboratory indexes, plasma cytokines, and peripheral blood mononuclear cell responses, which were assessed using single-cell RNA-sequencing in patients with non-critical COVID-19 with long durations (LDs) and short durations (SDs) of viral shedding. Our results revealed that clinical parameters and laboratory indexes, such as c-reactive protein (CRP) and D-dimer, were comparable between SDs and LDs. Most inflammatory cytokines/chemokines, such as IL-2, IL2R, TNF α / β , IL1 β , and CCL5 were present at low levels in LDs. Our single-cell RNA-sequencing revealed a reconfiguration of the peripheral immune cell phenotype in LDs, including decreases in natural killer (NK) cells and CD14⁺ monocytes and an increase in regulatory T cells (Tregs). Furthermore, most cell subsets in LDs consistently exhibited reduced expression of ribosomal protein (RP) genes, indicating dysfunctions in cytokine/chemokine synthesis, folding, modification, and assembly. Accordingly, the negative correlation between the RP levels and viral shedding duration was validated in an independent cohort of bulk-RNA-sequencing data from 103 non-critical patients, which may help guide clinical management and resource allocation. Moreover, peripheral T and NK cells and memory B cells in LDs likely failed to activate, which contributed to the persistence of viral shedding.

Introduction

There are more than 27,417,497 confirmed cases of the coronavirus disease (COVID-19) global pandemic and it has caused over 894,241 deaths as of September 9, 2020 (data from WHO at <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>). This has had a profound impact on the global culture and economy¹. Disease manifestation is highly heterogeneous, ranging from asymptomatic infection to severe disease leading to death². Moreover, the duration of viral shedding has been reported to vary dramatically, ranging from 6-63 days, with a median duration of 20 days from disease onset³⁻⁵. Longer durations (LDs) of viral shedding cause additional challenges in controlling the pandemic owing to increased risk of spread, the consumption of additional hospital resources, and greater economic costs. Moreover, we only have a limited understanding of the clinical and molecular characteristics of COVID-19 with long-term SARS-CoV-2 shedding. Thus, here, we delineated and compared the clinical features and molecular characteristics of patients exhibiting long-term and short-term viral shedding by integrating serological surveys and single-cell RNA-sequencing (scRNA-seq).

Results

Demographic characteristics

The duration of virus shedding was defined as the interval from illness onset until successive negative detection of SARS-CoV-2 RNA, consistent with other studies of COVID-19^{6,7}. As of April 30, 2020, a total of 12 non-critical COVID-19 in-patients exhibited LDs of viral shedding durations (viral shedding duration \geq 45

days). Given that the median SARS-CoV-2 viral shedding duration is approximately 3 weeks⁸, we also examined 38 age- and gender-matched non-critical COVID-19 in-patients whose viral shedding durations were less than 21 days (short durations, SDs), for comparison (Supplemental Table 1). All the patients were identified as laboratory-confirmed SARS-CoV-2 infected patients at Tongji Hospital, Wuhan, China. The median viral shedding duration was 57 days (range: 45–100 days) and 16 days (range: 3–21 days) in LDs and SDs, respectively (log-rank $P < 0.0001$) (Supplemental Table 1). The basic demographic information of these patients and the clinical parameters are detailed in Supplemental Table 1. Notably, there were no significant differences in comorbidities, complete blood counts (white blood counts, lymphocyte counts, neutrophil counts, platelet counts, and hemoglobin), blood biochemistry (alanine/aspartate aminotransferase and lactate dehydrogenase), and coagulation function (prothrombin, activated partial thromboplastin time, and D-dimer) between LDs and SDs. Moreover, inflammatory markers, such as procalcitonin (PCT), erythrocyte sedimentation rate (ESR), and C-reactive protein (CRP), which have been well reported as high-risk factors of the development of severe COVID-19⁹⁻¹¹, were comparable in LDs and SDs. Therefore, there is a need for further investigations to identify new biomarkers for viral shedding duration and the underlying mechanism of persistent viral shedding.

Cytokines in SDs, LDs, and healthy donors (HDs)

Cytokines are central to the pathophysiology of COVID-19 and a “cytokine storm” has been described as a feature of COVID-19 severity, which is associated with adverse outcomes^{12,13}. To further elucidate the immune response associated with the viral shedding duration, we checked the serum cytokine/chemokine levels in SDs, LDs, and 22 HDs. Intriguingly, from among the 48 cytokines/chemokines detected, 21 inflammatory cytokines/chemokines had the lowest levels in LDs when compared to SDs or HDs (Fig.1c). Of these, platelet-derived growth factor (PDGF-BB) ($P = 0.000065$), C-C motif ligand 5 (CCL5) ($P = 0.00011$), and macrophage migration inhibitory factor (MIF) ($P = 0.00015$) showed the most significant changes (Fig.1c and Extended Data Fig.1). Additionally, IL-1 β , IL-2, IL-2R, IL-9, IL-18, TNF- α , and TNF- β , the upregulation of which contributed to lung injury, multiorgan failure, and ultimately death¹⁴⁻¹⁶, were present at lower levels in LDs (Fig.1c). Collectively, LDs of viral shedding were associated with a weaker inflammatory response characterized by low circulating concentrations of cytokines and chemokines.

Differences in cell compositions detected using single-cell transcriptomes of human peripheral blood mononuclear cells (PBMCs)

To characterize the immunological features of LDs and SDs compared to HDs, we performed 10X Genomics scRNA-seq to study the transcriptomic profiles of PBMCs from HDs ($n = 3$) and from patients exhibiting LDs ($n = 5$) and SDs ($n = 4$) of viral shedding (Supplementary Table 2). The demographics,

clinical features, and laboratory findings of these patients are listed in Supplementary Table 2. After the unified single-cell analysis pipeline (see Methods), in total, 124,606 cells across all subjects, with an average of 10,384 cells per sample, were integrated into an unbatched and comparable dataset (Supplementary Table 2). In accordance with previous reports¹⁷, we did not detect SARS-CoV-2 RNA expression in the PBMCs of these patients (Extended Data Fig.2).

Using unsupervised clustering of uniform manifold approximation and projection (UMAP), we identified 20 cell populations based on the expression of canonical cell-type gene markers (Fig.2a-b, Extended Data Fig.3). To reveal the differences in cell composition across LDs and SDs and to compare them with that of HDs, we investigated the relative proportions of immune cells among the three groups (Fig.2c-d). While there were limited differences between SDs and HDs, significant differences were observed between LDs and HDs. Compared to that in the HDs, the proportion of natural killer (NK) cells and CD14⁺ monocytes were significantly decreased in LDs (Fig.2d). The massive decreases in NK cells and CD14⁺ monocytes in LDs were in accordance with the observed decrease in inflammatory cytokines in LDs. Notably, inflammatory monocytes, induced by T cells, have been reported to incite the cytokine storm in COVID-19¹⁸. In contrast, regulatory T cells (Tregs) were particularly elevated in LDs (Fig.2d). Given the importance of Tregs in secreting immunosuppressive cytokines and inhibiting the activation of both innate and adaptive immune cells^{11,19,20}, elevated levels of Tregs could be one cause of the suppressed immune response observed in LDs. Taken together, these results demonstrated that the decreased NK cells and CD14⁺ monocytes, as well as increased Tregs in LDs may prevent the immune system from overreacting but contribute to the persistence of the virus.

Transcriptional signatures associated with LDs

Next, we performed hierarchical clustering based on relative gene expression changes with respect to the HDs to evaluate the molecular difference of each cell type in LDs and SDs. Unexpectedly, all cell types among the PBMCs clustered together according to the disease groups rather than by cell-types, with the exception of hematopoietic stem cells (HSCs), plasma B cells, and megakaryocytes (Extended Data Fig.4). This indicates that the molecular features of PBMCs in LDs and SDs are markedly different, regardless of the cell type. Therefore, we sought to identify variations in the relevant biological functions in individual cell types through differentially expressed gene (DEG) and Gene Ontology (GO) analyses. Most importantly, we found that protein targeting to the membrane, endoplasmic reticulum (ER) related pathways, translation related pathways, and ribosome small subunit assembly pathway was consistently downregulated in all cell types in LDs, with the exception of gamma delta ($\gamma\delta$) T cells, mucosal associated invariant T (MAIT) cells, and megakaryocytes (Fig.3a). In agreement with the GO results, we found that many genes encoding ribosomal proteins and the protein synthesis related proteins were specifically downregulated in LDs (Fig.3b-c). Interestingly, RPL41, RPS29, RPL36A, RPS27, RPS21, RPS10, RPL38, RPL39, and RPS28 localize to the ER and participate in protein synthesis, folding, and assembly, as detailed in the information provided on <https://www.proteinatlas.org/>. TMA7²¹, TAF10²², and PTOV1²³

were also specifically downregulated in LDs. These genes have previously been reported to be associated with ribosomes, and their overexpression promotes global protein synthesis. Given that antibodies²⁴ and cytokines²⁵ are synthesized, folded, modified, and assembled by the rough ER and attached ribosomes, these findings suggest that immune cells of LDs tend to have reduced cytokine synthesis, folding, and assembly functions, which is consistent with the lower levels of inflammatory cytokines observed in LDs (Fig.1c). CEBPB and MAP3K8, which are involved in the production of pro-inflammatory cytokines, and ZFP36²⁶ and PDE4D²⁷, which are involved in IL-2 production, were selectively reduced in LDs (Fig.3b). Additionally, genes involved in T cell activation (PCBP1, ARPC2), migration (FMNL1), cytotoxic function (GNLY, SRM), transcription factors (LYN), and downstream signal transduction (COTL1) were all reduced in LDs (Fig.3b-c). Given that cytokines are produced by several immune cells, including adaptive T cells²⁸, the reduced cytokine levels in LDs are at least partially explained by these findings. Meanwhile, IFITM2, an interferon (IFN)-stimulated gene (ISG), vital for viral clearance²⁹, was downregulated in LDs and may contribute to the longer viral persistence in LDs (Fig.3b-c).

Extended viral shedding time leads to higher transmission probability and requires stricter infection control, the use of additional hospital resources, and result in more economic costs. Therefore, there is increased interest in early patient stratification to identify those who are more likely to recover rapidly in order to make hospital resources available for those at higher risk³⁰. We assessed whether the RP levels could be used for stratification of the viral shedding duration by integrating bulk-RNA-seq data from the 103 independent non-critical COVID-19 patients whose viral shedding durations were available. Remarkably, we found that lower expression of RPs was associated with longer viral shedding durations, including the following RPs identified in our scRNA-seq data: RPL38, RPL41, and RPS10 (Fig.3d-e). In summary, lower RP levels were associated with persistent viral shedding, and specific RPs could be applied as indicators of longer viral shedding.

Molecular features of T and NK cells in LDs and SDs

We next performed sub-clustering analysis on T and NK lymphocytes considering their crucial anti-viral effects^{31,32}. UMAP embedding of T and NK cells from all the samples identified substantial differences in the cellular phenotypes of CD4⁺ T, CD8⁺ T, NKT, and NK cells (Fig.4a-b). Although the proportion of T cells was comparable between LDs and SDs (Fig.2d), the correlation matrix revealed that the molecular features differed between the two groups (Fig.4c), such as memory CD8⁺ T cells and NK cells.

For example, in memory CD8⁺ T cells, DEGs involved in T-cell activation (CD74, SELENOK, FYN, CCL5, and RNF125), positive regulation of cytokine production (IRF1, SELENTOK, and HMGB2), proinflammatory mediators of secretion, and interferon- γ (IFN- γ) pathways (IRF1, HLA-DPB1, HLA-DPA1, HLA-DRB1, CCL5, and CCL4) were specifically downregulated in LDs, while they were upregulated in SDs compared to HDs (Fig.4d-e). In addition, in NK cells, DEGs associated with the positive regulation of T-cell activation pathways (ZFP36L2, SELENOK, SLA2, ZBTB1, MAP3K8, IRF1, CEBPB, RUNX3, and PIK3R1) were

profoundly enriched in SDs but not in LDs (Fig.4f-g), suggesting consistent immune cell dysfunction in LDs that might be associated with the mild secretion of proinflammatory cytokines and contribute to viral persistence.

Next, we reconstructed T cell antigen receptor (TCR) sequences from the TCR sequencing data. Briefly, more than 70% of the cells in all the subsets had matched TCR information, with the exception of the $\gamma\delta$ T and NKT subsets (Fig.4h). Compared to HDs, clonal expansion was obvious in patients with COVID-19, especially in those with SDs of viral shedding (Fig.4i). Meanwhile, the proportion of large clonal expansions (clonal size ≥ 30), primarily in cytotoxic cells, was higher in SDs (Fig.4i), indicating that SDs have more efficient clonal expansion of effector T cells than LDs to promote viral clearance.

To explore the preferential V and J combinations in SDs and LDs, we first analyzed and listed the V and J combinations most frequently observed in the TCRs in all samples (Fig.4j). Among these combinations, relatively frequent pairings of the TCR in HDs were TRBV29::TRVJ2-7 and TRAV29/DV5::TRAJ20, while TRAV29/DV5::TRAJ49 and TRBV9::TRBJ1-3 were frequent in LD patients, and TRAV12-3::TRAJ54 and TRAV1-2::TRAJ49 were frequent in SD patients (Fig.4j). The selective usage of V(D)J genes suggests that different immunodominant epitopes may drive the molecular composition of T-cell responses and may be associated with LD and SD infection.

Features of B cell subsets in LDs

Some T cells and cytokines prime B cells for maturation, which go on to become plasma cells and produce pathogen neutralizing antibodies³³. We subclustered B cells into three subsets according to the expression and distribution of canonical B-cell markers (Extended data Fig.5a-b). Compared with HDs, plasma B cells were not significantly increased in SDs, which may be due to sampling during the convalescent period³⁴ (Extended Data Fig.5a, Fig.2d). In LDs, despite viral persistence, the proportion of plasma cells was also extremely low, which may indicate that LDs fail to produce sufficient neutralizing antibodies (Extended Data Fig.5a, Fig.2d). Previous studies^{5,35} have suggested that antibodies produced by plasma cells in response to SARS-CoV-2 during initial exposure disappeared within a few weeks, but memory B cells persisted for much longer. Therefore, we compared the expression profiles of memory B cells in the three groups. Interestingly, the pathways involved in T-cell differentiation (CD83, ZFP36L2, and GPR183) and cell growth and activation (CD83, ZFP36L2, GPR183, and PELI1) were selectively enriched in SDs but not LDs, indicating that B and T cells in LDs failed to synergize in order to clear the virus (Extended Data Fig.4d-e). Moreover, HMGB2 and PDE4B, which positively regulate the production of cytokines such as IL-2, and the pathways involved in leukocyte chemotaxis (LYN, DUSP1, and RAC1) were exclusively enriched in SDs (Extended data Fig. 4c-d).

Discussion

COVID-19 patients with LDs of viral shedding require more stringent infection control measures, the utilization of additional hospital resources, and increased economic expenditures, along with posing a higher risk of transmission. In this study, we found that clinical indexes, including contemporaneous CRP, D-dimers, IL-6, IL-8, and ESR (Fig.1b, Extended Data Fig.1) failed to distinguish the patients with LDs of viral shedding. Furthermore, although persistently positive for SARS-CoV-2 RNA, patients with LDs of viral shedding appeared to exhibit a hypo-inflammatory response, as evidenced by lower cytokine levels (Fig.1c). Such patients may not experience a cytokine storm and tend to exhibit mild symptoms³⁶. Moreover, the pathways and genes associated with the ER and ribosome dysregulation may contribute to the persistent viral shedding, declined secretion of inflammatory cytokines, and mild symptoms in LDs. Moreover, the RP levels were further validated to be negatively correlated with the duration of viral shedding in an independent cohort of non-critical COVID-19 patients. From a practical perspective, it should be noted that RP levels were better able to identify LDs than the commonly used clinical parameters such as D-dimer, CRP, and ESR. Therefore, it may be reasonable to introduce RP profiling into routine clinical tests to aid in the early risk-based stratification of patients, which could provide a major benefit when healthcare systems are overwhelmed.

Recently, Matthias et al. reported that nonstructural protein 1 (Nsp1) from SARS-CoV-2 efficiently interferes with the cellular translation machinery (40S ribosome subunit), inhibits all cellular antiviral defense mechanisms, including the interferon response and other proinflammatory cytokines, and then facilitates efficient viral replication and immune evasion³⁷. Indeed, we observed massively decreased RP expression in LDs in our study; however, whether the reduced levels of RPs in LDs were the cause or the consequence of viral persistence requires further investigation.

We used single-cell transcriptomics to characterize and visualize the peripheral immune responses in patients with long-term COVID-19 compared to the responses in those with short-term COVID-19 and in healthy donors. We observed marked changes in the immune cell composition, phenotype, and immunological features in LDs. Thus, this work provides new insights into the pathophysiology of COVID-19 and a resource for understanding peripheral immune heterogeneity in patients with LDs of viral shedding.

Lastly, there are a few limitations of our study. For example, our sample size is small, and the timing of the clinical presentation of the patients varied, which may influence their transcriptional landscapes. Additionally, only peripheral blood was evaluated as it was challenging to obtain immune cells from the bronchoalveolar lavage fluid owing to biosafety concerns during the COVID-19 outbreak when this study was performed. Therefore, future studies with longitudinal samples from lesion sites, such as the lung, and with more patients can provide a more systematic overview and more comprehensive conclusions.

Declarations

Author contribution

C.S. and K.L. conceived the experiments. B.Y. and J.H. collected the samples and performed single-cell library construction. J.F. and C.S. performed single-cell data analysis. B.Y. wrote the manuscript and C.S. revised it. E.G., Y.F., S.L., R.X., C.L., F.L., Z.W., C.H., X.Q., D.H., L.Y., T.Q. and T.W. collected clinical information, interpretation of data and performed cytokines measurements. X.L., J.Z. and P.W. performed bulk RNA-seq analysis. G.C., K.L. and C.S. provided expertise and feedback. C.S. and K.L. conceived and coordinated the project.

Competing Interests statement

The authors declare no conflicts of interest.

References

- 1 Wu, F., Zhao, S., Yu, B., Chen, Y. M. & Zhang, Y. Z. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 1-8 (2020).
- 2 Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H. & Zhong, N. S. Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine* **382** (2020).
- 3 Young, B. E. *et al.* Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. *Jama* **323**, 1488-1494 (2020).
- 4 Widders, A., Broom, A. & Broom, J. SARS-CoV-2: the viral shedding vs infectivity dilemma. *Infection, Disease & Health* (2020).
- 5 Long, Q.-X. *et al.* Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nature medicine* **26**, 1200-1204 (2020).
- 6 Xu, K. *et al.* Factors associated with prolonged viral RNA shedding in patients with COVID-19. *Clinical Infectious Diseases* (2020).
- 7 Qi, L. *et al.* Factors associated with duration of viral shedding in adults with COVID-19 outside of Wuhan, China: A retrospective cohort study. *International Journal of Infectious Diseases* (2020).
- 8 He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine* **26**, 672-675, doi:10.1038/s41591-020-0869-5 (2020).
- 9 Ling, W. C-reactive protein levels in the early stage of COVID-19. *Medecine et maladies infectieuses* (2020).

- 10 Gao, Y. *et al.* Diagnostic utility of clinical laboratory data determinations for patients with the severe COVID-19. *Journal of medical virology* (2020).
- 11 Qin, C. *et al.* Dysregulation of immune response in patients with COVID-19 in Wuhan, China. *Clinical Infectious Diseases* (2020).
- 12 Mehta, P. *et al.* COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet (London, England)* **395**, 1033 (2020).
- 13 Ye, Q., Wang, B. & Mao, J. The pathogenesis and treatment of the Cytokine Storm in COVID-19. *Journal of infection* **80**, 607-613 (2020).
- 14 Shimizu, M. in *Cytokine Storm Syndrome* 31-41 (Springer, 2019).
- 15 Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet* **395**, 497-506 (2020).
- 16 Chen, G. *et al.* Clinical and immunologic features in severe and moderate forms of Coronavirus Disease. *J Clin Invest.* **137244** (2019).
- 17 Zhu, L. *et al.* Single-cell sequencing of peripheral blood mononuclear cells reveals distinct immune response landscapes of COVID-19 and influenza patients. *Immunity* (2020).
- 18 Zhou, Y. *et al.* Pathogenic T-cells and inflammatory monocytes incite inflammatory storms in severe COVID-19 patients. *National Science Review* (2020).
- 19 Chen, G. *et al.* Clinical and immunological features of severe and moderate coronavirus disease 2019. *The Journal of clinical investigation* **130** (2020).
- 20 Stephen-Victor, E. *et al.* Potential of regulatory T cell-based therapies in the management of severe COVID-19. *European Respiratory Journal* (2020).
- 21 Fleischer, T. C., Weaver, C. M., McAfee, K. J., Jennings, J. L. & Link, A. J. Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. *Genes & development* **20**, 1294-1307 (2006).
- 22 Kamenova, I. *et al.* Co-translational assembly of mammalian nuclear multisubunit complexes. *Nature communications* **10**, 1-15 (2019).
- 23 Marques, N. *et al.* Regulation of protein translation and c-Jun expression by prostate tumor overexpressed 1. *Oncogene* **33**, 1124-1134 (2014).
- 24 Van Anken, E. *et al.* Efficient IgM assembly and secretion require the plasma cell induced endoplasmic reticulum protein pERp1. *Proceedings of the National Academy of Sciences* **106**, 17019-17024 (2009).

- 25 So, J.-S. Roles of endoplasmic reticulum stress in immune responses. *Molecules and cells* **41**, 705 (2018).
- 26 Moore, M. J. *et al.* ZFP36 RNA-binding proteins restrain T cell activation and anti-viral immunity. *Elife* **7**, e33057 (2018).
- 27 Peter, D., Jin, S. C., Conti, M., Hatzelmann, A. & Zitt, C. Differential expression and function of phosphodiesterase 4 (PDE4) subtypes in human primary CD4+ T cells: predominant role of PDE4D. *The Journal of Immunology* **178**, 4820-4831 (2007).
- 28 Cronkite, D. A. & Strutt, T. M. The regulation of inflammation by innate and adaptive lymphocytes. *Journal of immunology research* **2018** (2018).
- 29 Schoggins, J. W. & Rice, C. M. Interferon-stimulated genes and their antiviral effector functions. *Current opinion in virology* **1**, 519-525 (2011).
- 30 Steinberg, E., Balakrishna, A., Habboushe, J., Shawl, A. & Lee, J. Calculated decisions: COVID-19 calculators during extreme resource-limited situations. *Emergency medicine practice* **22**, CD1-CD5 (2020).
- 31 Zheng, M. *et al.* Functional exhaustion of antiviral lymphocytes in COVID-19 patients. *Cellular & molecular immunology* **17**, 533-535 (2020).
- 32 Maucourant, C. *et al.* Natural killer cell immunotypes related to COVID-19 disease severity. *Science Immunology* **5** (2020).
- 33 LeBien, T. W. & Tedder, T. F. B lymphocytes: how they develop and function. *Blood* **112**, 1570-1580, doi:10.1182/blood-2008-02-078071 (2008).
- 34 Zhang, J.-Y. *et al.* Single-cell landscape of immunological responses in patients with COVID-19. *Nature Immunology*, 1-12 (2020).
- 35 Kellam, P. & Barclay, W. The dynamics of humoral immune responses following SARS-CoV-2 infection and the potential for reinfection. *Journal of General Virology*, jgv001439 (2020).
- 36 Coperchini, F., Chiovato, L., Croce, L., Magri, F. & Rotondi, M. The cytokine storm in COVID-19: an overview of the involvement of the chemokine/chemokine-receptor system. *Cytokine & Growth Factor Reviews* (2020).
- 37 Thoms, M. *et al.* Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* **369**, 1249-1255, doi:10.1126/science.abc8665 (2020).
- 38 Lee, J. S. *et al.* Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Science immunology* **5** (2020).

Methods

Patient information and data source

This study was reviewed and approved by the Institutional Review Board of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology (TJ-IRB20200405). All the enrolled patients signed an informed consent form, and all the blood samples were collected using the rest of the standard diagnostic tests, with no burden to the patients. A waiver of informed consent was obtained to query the patient electronic health records.

As April 30, 2020, whole blood samples from 103 COVID-19 patients diagnosed for SARS-CoV-2 viral infection status by PCR were enrolled for bulk RNA-seq. Serum from 50 patients and 22 healthy donors as controls were for the multiplex cytokine panel measurements. And PBMCs isolated from 12 patients and 3 healthy donors as controls were performed 10X scRNA-seq. Routine laboratory measurements and blood counts were obtained as part of standard medical care.

Cytokine measurements

The levels of serum cytokines were determined by Bio-Plex Pro Human Cytokines 48-Plex Screening assay (Bio-Rad Life Sciences, Hercules, CA, USA) using a Luminex FLEXMAP 3D system (Luminex, Austin, TX, USA) according to the manufacturer's protocols. The 48-Plex Screening panel: Basic FGF, CTACK, eotaxin, G-CSF, GM-CSF, GRO- α , HGF, ICAM-1, IFN- α 2, IFN- γ , IL-1 α , IL-1ra, IL-2, IL-2R α , IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12, IL-13, IL-15, IL-16, IL-17A, IL-18, IP-10, LIF, MCP-1, MCP-3, M-CSF, MIF, MIG, MIP-1 α , MIP-1 β , β -NGF, PDGF-BB, RANTES, SCF, SCGF- β , SDF-1 α , TNF- α , TNF- β , TRAIL, VCAM-1, VEGF-A. Data were analyzed using Bio-Plex Manager 6.2 software (Bio-Rad Life Sciences, Hercules, CA, USA). Undetected values were inputted with a random value between 0 and the limit of detection 1 to avoid an artificial reduction in the standard deviation.

scRNA library construction sequencing by 10X genomics

For both patients with COVID-19 and healthy controls, blood was collected into heparin tubes (Becton, Dickinson and Co.) and PBMCs were isolated by density gradient centrifugation using Ficoll-Paque Plus medium (GE Healthcare) and washed with Ca/Mg-free PBS. Blood was processed within 4 h of collection for all samples, and within 1 h for most. PBMC cells were examined by microscope after 0.4% Trypan blue coloring. When the viability of cells was higher than 80%, the experiment of library construction was performed using the ChromiumTM Controller and ChromiumTM Single Cell 5' Reagent Version 2 Kit (10x Genomics, Pleasanton, CA). Briefly, single cells, reagents and Gel Beads containing barcoded oligonucleotides were encapsulated into nanoliter-sized GEMs (Gel Bead in emulsion) using the GemCode Technology. Lysis and barcoded reverse transcription of polyadenylated mRNA from single cells were

performed inside each GEM. Post RT-GEMs were cleaned up and cDNA were amplified. cDNA was fragmented and fragments end were repaired, as well A-tailing was added to the 5' end. The adaptors were ligated to fragments which were double sided SPRI selected. Another double sided SPRI selecting was carried out after sample index PCR. The final library was quality and quantitated in two methods: check the distribution of the fragments size using the Agilent 2100 bioanalyzer, and quantify the library using real-time quantitative PCR (QPCR) (TaqMan Probe). The final products were sequenced using the Illumina HiSeq 4000 or Xten platform (BGIShenzhen, China).

TCR V(D)J sequencing

Full-length TCR V(D)J segments were enriched from amplified cDNA from 5' libraries via PCR amplification using a Chromium Single-Cell V(D)J Enrichment kit according to the manufacturer's protocol (10X Genomics).

Detection of SARS-CoV-2 transcripts

Mock sample with SARS-CoV-2 transcripts was generated by add 200 SARS-CoV-2 paired reads to the health control. SARS-CoV-2 transcripts were identified from sequencing data using Viral-Track and Cell Ranger (version 3.0.1, 10x Genomics) with a modified reference contain SARS-CoV-2 genome (NC_045512.2).

Single cell RNA-seq data processing

The sequencing data of patients were processed using Cell Ranger against the GRCh38 human reference genome. Quality of cells were then assessed based on the UMI counts per cell, genes expressed per cell and the proportion of mitochondrial gene counts using Seurat (version 3.1.5). Cells that had UMIs between 500 and 30,000, more than 200 genes expressed and fewer than 15% of UMIs from mitochondrial genes were considered high quality and retained for further analysis. We next identified and removed the doublets following previous described method (Pijuan-Sala et al., 2019). After removing the doublets, a total of 124,606 cells were retained for downstream analysis.

To remove batch effect, the function "NormalizeData" and "FindVariableFeatures" in Seurat was performed respectively for each sample. After that, the normalized data were integrated using "FindIntegrationAnchors" and "IntegrateData" function with parameter set to "nfeatures=300, dims=1:30". Then, the integrated dataset was scaled and PCA conducted with features exclude ribosomal protein and mitochondria protein and the cells were clustered using "FindNeighbors" and "FindClusters" function with

parameter set to “k.param = 10, resolution = 1”. Finally, the cells were visualized by UMAP using the top 30 principal components.

Cell types annotation

The markers of each clusters identified by the “FindAllMarker” function, as well as some canonical markers, were visualized using violin plot and feature plot, the expression of them were used to classify and annotate the clusters. HSC cells and clusters expression more than 2 canonical cell-type markers were excluded and 122,865 cells were retained for further analysis.

Identification of differential expressed genes (DEGs) analysis and GO enrichment

DEGs were performed using “FindMarkers” function with MAST algorithm in Seurat based on a Bonferroni-adjusted $p < 0.05$ and a \log_2 fold change > 0.25 . For GO enrichment, DEGs identified were conducted using function enrichGO in ClusterProfiler with parameter set to “OrgDb = org.Hs.eg.db, ont = ‘BP’, pAdjustMethod = ‘BH’ ”.

Hierarchical clustering of gene expression changes among disease groups at cell type resolution

Hierarchical clustering of gene expression changes was conducted following previous described method³⁸. Briefly, the UMI count of each gene were normalized by the total UMI count in each cell type and multiplied by 100,000. The gene expression of each disease groups was divided by the values in the healthy donors and the highly variable genes in terms of the top 3000 standard deviation followed by \log_2 -transformation. Hierarchical clustering was conducted based on the Pearson correlation coefficient (PCC) of the highly variable genes.

Hierarchical clustering of ribosomal genes in whole blood RNA-seq

RNA-seq data of 103 patients divided into three groups separated by 21 days and 45 days were aligned to the reference genome using hisat2 and the gene expression level (FPKM) was calculated by RSEM. The expression of ribosomal genes was \log_2 transformed followed by z-transformation, which was used for hierarchical clustering.

Analysis of T cells, NK cells and B cells

T and NK cells were extracted from PBMCs and the cells were principal component analysis and visualized as described above. B cells in PBMCs were also extracted and processed using the procedure used for T cells.

Hierarchical clustering of DEGs

DEGs of COVID-19 groups compared to HD were selected and the expression of them in individuals are calculated by the function “Average Expression” in Seurat followed by z-transformation. Then the scaled expression was used for hierarchical clustering.

TCR V(D)J analysis

The sequencing data were performed using Cell Ranger V(D)J pipeline with GRCh38 as reference. The TCR matrix containing barcode information and clonotype frequency was obtained, the cells with at least one productive TRA and one productive TRB were retained for further analysis. Each unique TRA(s)-TRB(s) pair was defined as a clonotype.

Boxplot

All of the boxplots in this paper were performed using “ggboxplot()” function in ggpubr R package. Each point represents for one sample. The horizontal line with each box represents the median, and the top and bottom of each box indicate the 25th and 75th percentile.

Statistical analysis

Wilcoxon rank-sum test were performed using R (version 3.6.1) in this study. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figures

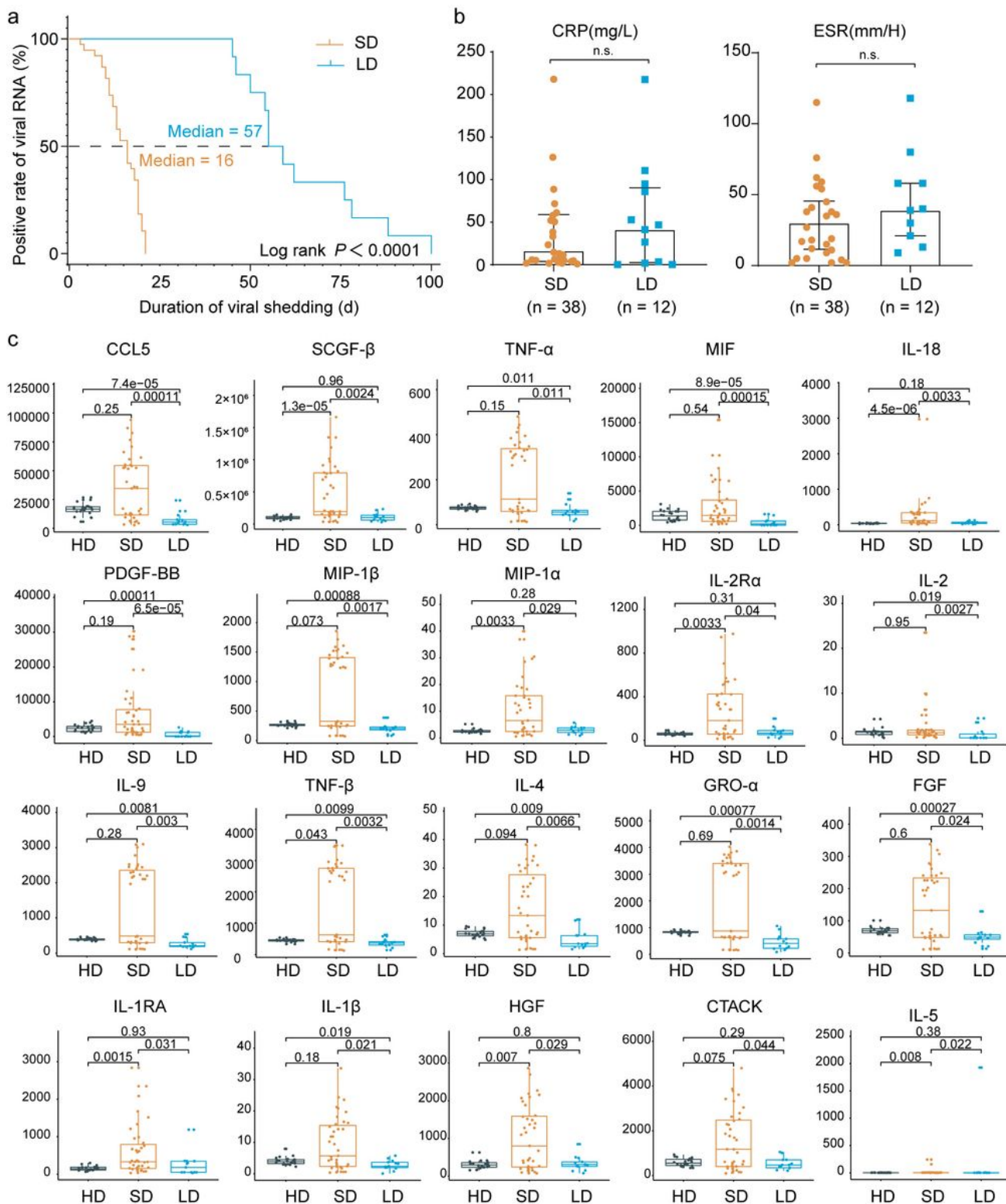


Figure 1

Virological, clinical and cytokines/chemokines characteristics in LDs and SDs. a, The Kaplan-Meier method was used to estimate the positive rate of viral RNA, and the two-sided log-rank test was applied to evaluate the significance difference of the duration of viral shedding in the LDs (n=12) and SDs (n=38). b, The C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR) values obtained from clinical records of LDs and SDs. The box plots show the median (middle line) and the first and third quartiles

(boxes), whereas the whiskers show 1.5× the IQR above and below the box. Wilcoxon rank-sum test was performed, and the significant p values cutoff was set at 0.05. c, Samples from LDs and SDs were collected during hospitalization, and assays were performed to measure the concentrations of 48 cytokines/chemokines. The box plots show the median (middle line) and the first and third quartiles (boxes), whereas the whiskers show 1.5× the IQR above and below the box. Wilcoxon rank-sum test are performed, and the significant p values cutoff was set at 0.05. *p<0.05, **p<0.01, ***p<0.001.

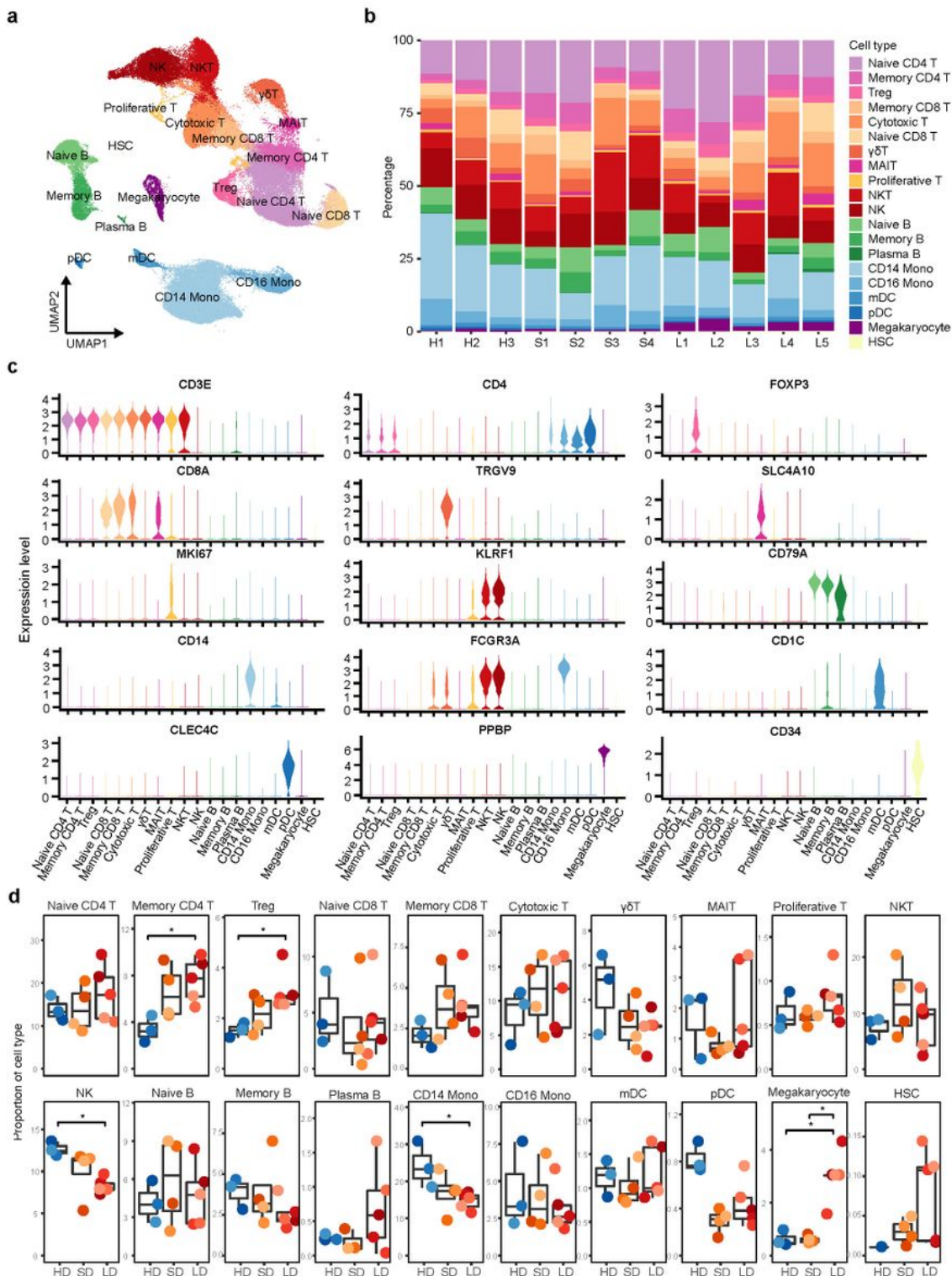


Figure 2

Differences in cell compositions by Single-cell transcriptomes of PBMCs. a, UMAP plot of 122,865 single cells colored by cell types identified from HDs (n=3), SDs (n=4), LDs (n=5). b, Proportion of each cell type at single sample level. c, Expression distribution of selected canonical markers showed by violin plots in the 20 clusters. d, Box plots of proportion of each cell cluster in each group. Samples were shown in different colors. Horizontal lines represent median values, with a maximum of $1.5 \times$ interquartile range. Wilcoxon rank-sum tests were conducted between each group. * $p < 0.05$.

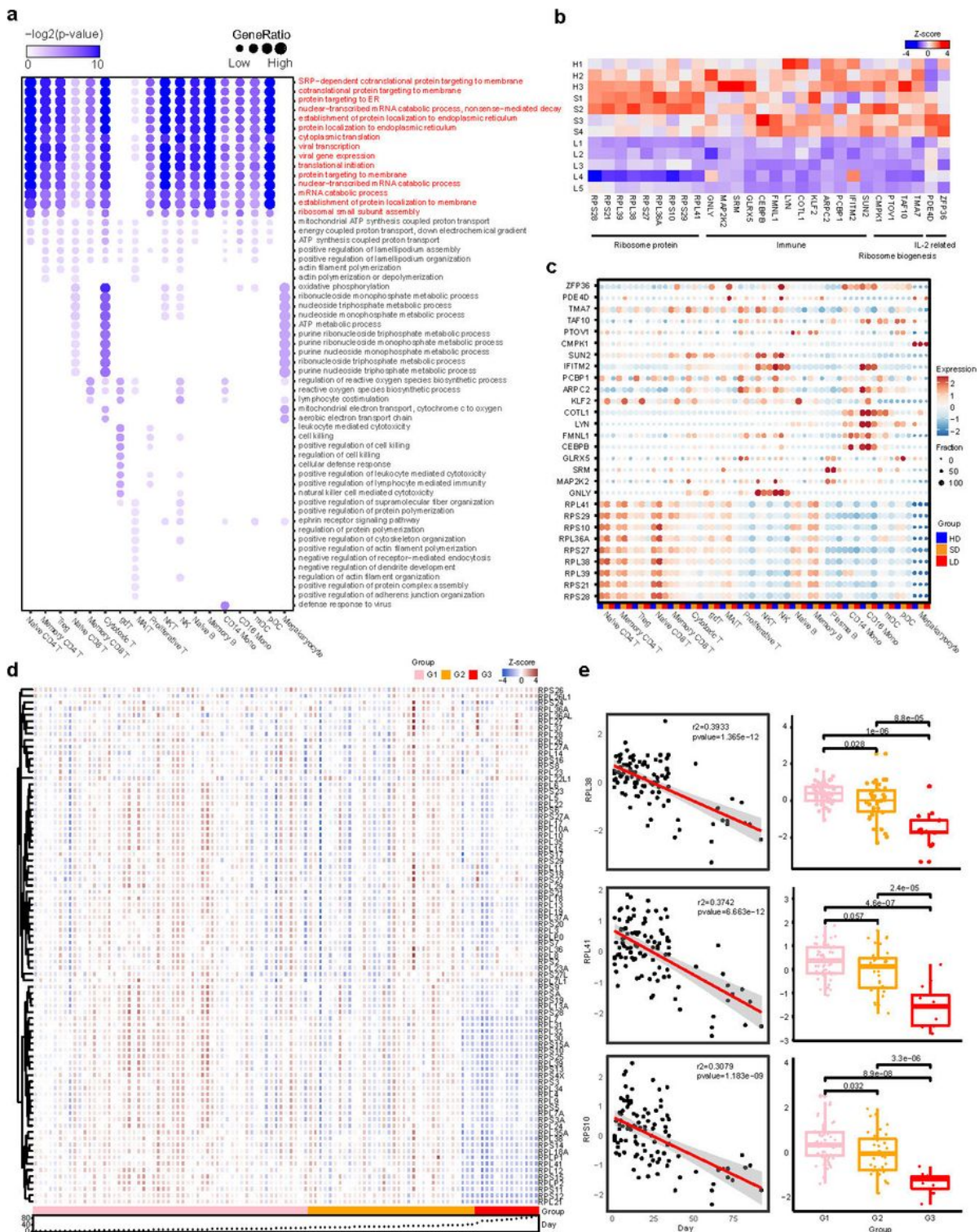


Figure 3

Transcriptional signatures associated with long viral shedding duration. a, Enriched GO pathways of down-regulated genes between LDs and HDs at cell type resolution. Pathways enriched by ribosomal genes are labeled in red. The color intensity indicates the enrichment p-values and the point size indicates the ratio of gene enrich in each pathway. b, Expression level of selected genes across 12 samples. The color intensity indicates the relative expression level. c, Expression of selected genes among groups at cell type resolution. The color intensity indicates the relative expression level and the point size indicates the ratio of cells with each gene expressed. The color bar under the dot plot indicates the group. d, Expression level of ribosomal genes in whole blood bulk-RNA across 103 COVID-19 patients. The color intensity indicates the relative expression level, the color bar under the heatmap indicates the disease group and the scatter indicates the duration time of COVID-19. e, Relationship between expression levels of selected ribosomal genes and duration time of COVID-19 among 103 COVID-19 patients. Boxplot of expression levels of selected genes in three groups among 103 COVID-19 patients. Horizontal lines represent median values, with a maximum of $1.5 \times$ interquartile range. Difference between three groups were performed by wilcoxon rank-sum tests. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

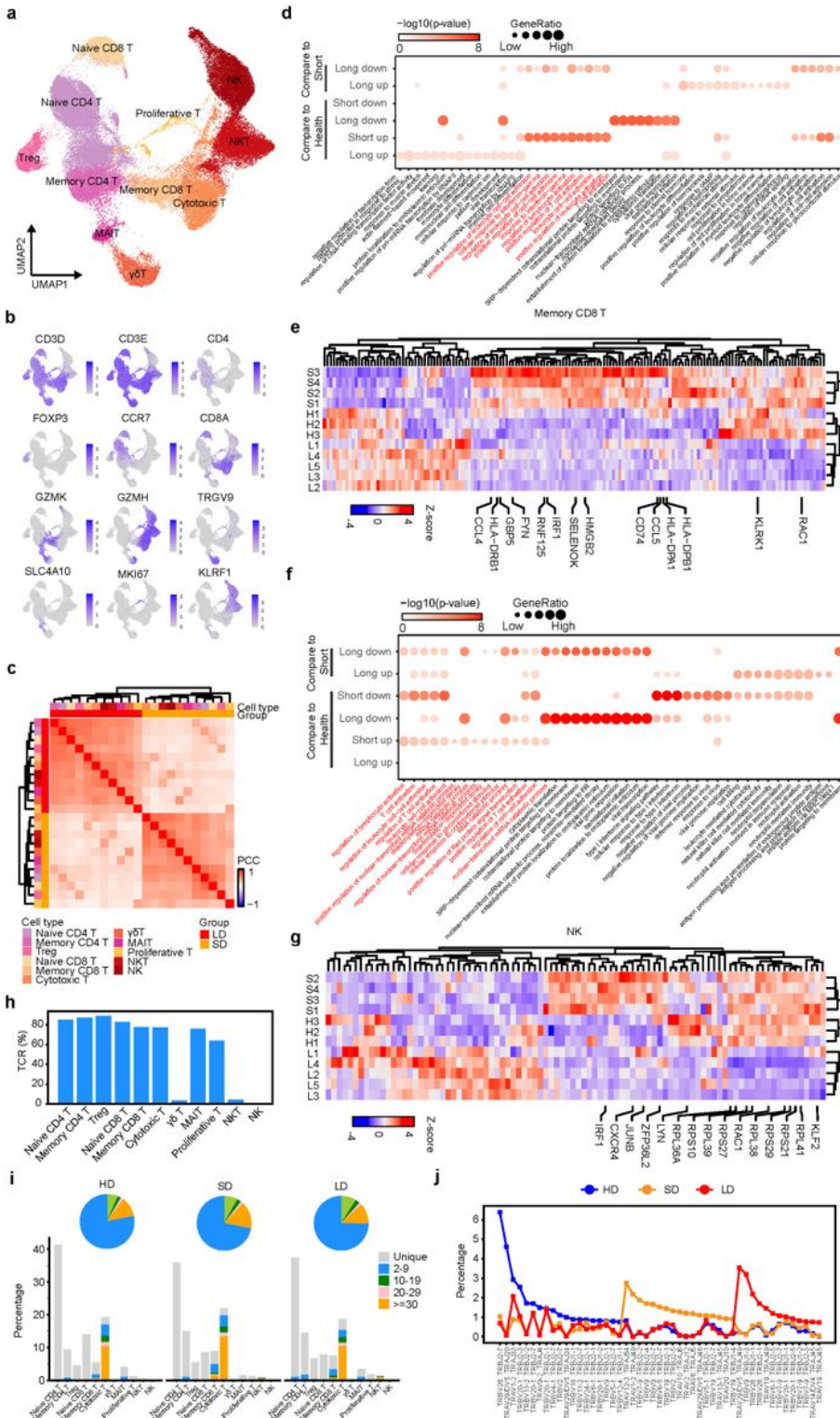


Figure 4

Molecular features of T and NK cells in LDs and SDs. **a**, UMAP projection of T and NK cells. Each dot corresponds to a single cell, colored by cell type. **b**, UMAP plot of canonical markers in 11 cell clusters. Data are colored according to log scaled expression levels. **c**, Hierarchical clustering using the Pearson correlation coefficient (PCC) of a normalized transcriptome between disease groups in T and NK cells. The color intensity indicates the PCC and the color bars above the heatmap indicate the cell type and

disease group. d, Enriched go pathways of COVID-19 groups in memory CD8+ T cells (top 2 columns: DEGs between LDs and SDs, bottom 4 columns: DEGs for SDs and LDs compared to HDs. Pathway enriched by SDs up-regulated genes are labeled in red. The color intensity indicates the enrichment p-values and the point size indicates the ratio of gene enrich in each pathway. e, Hierarchical clustering of expression of DEGs in memory CD8+ T cells at sample level. The color intensity indicates the relative expression of each gene. f, Enriched GO pathways of COVID-19 groups in NK cells (top 2 columns: DEGs between LDs and SDs, bottom 4 columns: DEGs for SDs and LDs compared to HDs. Pathways enriched by SDs up-regulated genes are labeled in red. The color intensity indicates the enrichment p-values and the point size indicates the ratio of gene enrich in each pathway. g, Hierarchical clustering of expression of DEGs in NK cells at sample level. The color intensity indicates the relative expression of each gene. h-i, The clonal status percentage of T cells and that at each cell type resolution in three groups. j, The percentage of selected clonal types in three groups.

Supplementary Files

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

- [FigS1.jpg](#)
- [FigS2.jpg](#)
- [FigS3.jpg](#)
- [FigS4.jpg](#)
- [FigS5.jpg](#)
- [Supplementaltable1.docx](#)
- [SupplementalTable2.docx](#)