bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Systematic temporal analysis of peripheral blood transcriptomes using *TrendCatcher* identifies early and persistent neutrophil activation as a hallmark of severe COVID-19

Xinge Wang[1,2,3], Mark Sanborn[1, 2, 3], Yang Dai[1] and Jalees Rehman[1, 2, 3*]

[1]Department of Biomedical Engineering, University of Illinois Colleges of Engineering and Medicine, Chicago, IL, USA

[2]Department of Pharmacology and Regenerative Medicine, University of Illinois College of Medicine, Chicago, IL, USA

[3]Division of Cardiology, Department of Medicine, University of Illinois College of Medicine, Chicago, IL, USA

*Please address correspondence to Jalees Rehman (jalees@uic.edu)

The University of Illinois at Chicago, College of Medicine (M/C 868)

835 S. Wolcott Ave. Rm. E403, Chicago, IL 60612, USA

Office phone: +1 312-996-5552; Fax: +1 312-996-1225

Conflict-of-interest statement

The authors have declared that no conflict of interest exists.

Keywords:

COVID-19, Immunology, Inflammation, RNA-sequencing, Single cell RNA-sequencing, Temporal analysis, Gene expression dynamics, Differential expression, Transcriptomics

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Studying temporal gene expression shifts during disease progression provides important insights into the biological mechanisms that distinguish adaptive and maladaptive responses. Existing tools for the analysis of time course transcriptomic data are not designed to optimally identify distinct temporal patterns when analyzing dynamic differentially expressed genes (DDEGs). Moreover, there is a lack of methods to assess and visualize the temporal progression of biological pathways mapped from time course transcriptomic datasets. In this study, we developed an open-source R package *TrendCatcher* (https://github.com/jaleesr/TrendCatcher), which applies the smoothing spline ANOVA model and break point searching strategy to identify and visualize distinct dynamic transcriptional gene signatures and biological processes from longitudinal datasets. We used *TrendCatcher* to perform a systematic temporal analysis of COVID-19 peripheral blood transcriptomes, including bulk RNA-seq and scRNA-seq time course data. *TrendCatcher* uncovered the early and persistent activation of neutrophils and coagulation pathways as well as impaired type I interferon (IFN-I) signaling in circulating cells as a hallmark of patients who progressed to severe COVID-19, whereas no such patterns were identified in individuals receiving SARS-CoV-2 vaccinations or patients with mild COVID-19. These results underscore the importance of systematic temporal analysis to identify early biomarkers and possible pathogenic therapeutic targets.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Time-course transcriptomic profiling has been widely used to study and model dynamic biological processes in cells (1). By profiling mRNA levels during consecutive time points, researchers can infer dynamic responses to various external cues that cannot be observed by looking at only initial and terminal states. Recent improvements in high-throughput RNA sequencing (RNA-seq) technologies, including single-cell RNA-sequencing (scRNA-seq) provide viable approaches to study dynamic gene expression changes (2). Especially scRNA-seq allows for the in-depth analysis of temporal changes in distinct cell populations, thus providing insights into the heterogeneity and dynamics of responses to environmental cues or pathogenic stimuli. However, the analysis and visualization of longitudinal bulk RNA-seq data or scRNA-seq can be computationally challenging.

Currently, there are two predominant strategies for the analysis of sequential transcriptomic datasets. One strategy treats the sampling time points as categorical variables and is based on generalized linear models (GLMs). GLM-based packages include DESeq2 (3), edgeR (4) and limma (5). A complementary strategy is to treat time as a continuous variable and fit the time expression data into a spline-like model. These methods include DESeq2Spline (DESeq2 adopted with spline model for temporal RNA-seq datasets) fitting, ImpulseDE2 (6) and Next maSigPro (7). The former strategies focus on the magnitude of change instead of the time order of gene expression, and may also suffer from a relative loss of statistical testing power, especially if many time points are assessed (6). The latter strategy increases the power of detecting dynamic genes, but is based on strong model assumptions which are not optimally suited for multiphasic responses, such as gene trajectory patterns that reflect initial acute stimuli followed by counter-regulatory compensatory responses. Furthermore, there is a lack of tools that leverage existing knowledge of functional pathway databases to infer and visualize pathway trajectories instead of individual gene trajectories only.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We found this is challenging in a complex disease, such as coronavirus disease 19 (COVID-19) caused by the SARS-CoV-2 virus infection (8). COVID-19 is characterized by distinct disease progression patterns that suggest diverse host immune responses (9). Patients with severe disease exhibit profound inflammatory responses and immunopathology (10). COVID-19 immunophenotyping studies involve a large number of time points from corresponding RNA-seq and scRNA seq datasets (11, 12). Existing approaches for temporal transcriptomic analysis either do not take the temporal sequence into account when identifying dynamic differentially expressed genes (DDEGs) using pairwise comparison between time points, nor systematically analyze pathways that are dysregulated at defined time points.

In this study, we developed *TrendCatcher*, an open-source R package tailored for longitudinal bulk RNA-seq and scRNA-seq analysis. *TrendCatcher* uses a framework that combines the smooth spline ANOVA model and break point searching strategy, which identifies inflection points when gene expression trends reverse. We show that *TrendCatcher* outperformed commonly used methods for longitudinal RNA-seq analysis when using simulated time course data for benchmarking. We also analyzed bulk RNA-seq and scRNA-seq gene expression profiles of peripheral blood cells in COVID-19 patients at various disease time points. *TrendCatcher* allowed us to identify and visualize dynamic gene expression signature profiles in peripheral blood that were associated with poor disease outcomes during the early phases of disease, and could thus serve as novel mechanistic targets as well as early biomarkers for patient prognostication.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

TrendCatcher accurately identifies DDEGs in simulated datasets

First, we tested the prediction performance of the *TrendCatcher* platform (**Figure 1A**) using a set of simulated time course RNA-seq datasets, because simulated data provides defined standards to assess the accuracy of novel analytical platforms. We considered a comprehensive collection of gene temporal trajectory patterns to simulate a set of realistic data with biological characteristics (1). We embedded 10,000 simulated trajectories with varied temporal patterns, including 90% non-dynamic trajectories, 2.5% monotonous transition trajectories (continuously increasing or continuously decreasing gene expression levels throughout the time course), 2.5% impulse shaped single break point trajectories (only one temporal inflection point, i.e. up-peak-down or down-trough-up), 2.5% two break point trajectories and 2.5% three break point trajectories (multimodal dynamic response, e.g., a combination of 2 or more basic types of trajectories). Compared to DESeq2, DESeq2Spline and ImpulseDE2, *TrendCatcher* had the highest area under the ROC curve (AUC) in a mixed simulated dataset for time-course data with 7 time points (**Figure 1B**). We also tested each model's performance on a varying number of time points, including 3, 5, 7, 9 and 11 time points. As shown in **Figure 1C**, *TrendCatcher* had the highest prediction AUC across all time points, with the AUC values range from 0.88 to 0.90.

We next evaluated the prediction performance for each type of temporal trajectories. As shown in (**Supplemental Figure 1**), although other three methods achieved slightly higher accuracy than *TrendCatcher* in monotonic trajectories, their AUCs dropped markedly when more complicated trajectories were embedded. DESeq2 only achieved AUC values of 0.49 to 0.62 for both bi-phasic trajectory and multimodal trajectory. The DESeq2Spline approach using a spline curve fitting model also dropped to an AUC of approximately 0.7 once multiphasic trajectories were introduced. These results suggests that existing approaches for longitudinal or time course analyses are well-suited for monotonic trajectories (continuously up or

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

continuously down) but that *TrendCatcher* maybe more broadly applicable, because it identifies monotonic, biphasic (up-down, down-up) and multi-phasic shifts in gene expression, which are especially important in complex pathological setting when initial biological responses are followed by counter-regulatory adaptive or maladaptive responses.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

TrendCatcher identifies rapid but transient upregulation of interferon signaling in peripheral blood following SARS-CoV-2 infection in a non-human primate model

To define the key dynamic gene signatures associated with SARS-CoV-2 infection in peripheral blood, we first analyzed the global transcriptomics profiles from a non-human primate dataset (13), in which samples were collected on Days 0 (uninfected controls), 1, 2, 4, 7, 10 and 14 following the live SARS-CoV-2 inoculation. This experiment in non-human primates had the advantage of clearly defining the timing of inoculation and following the time course of dynamic genes. *TrendCatcher* identified 962 DDEGs out of 12,754 total genes, accounting for 7.6% of total expression, suggesting that over 90% of the expressed genes in the peripheral blood remain close to the baseline expression levels even in the setting of the SARS-CoV-2 infection. We observed two major types of dynamic trajectories: (a) 167 genes followed a biphasic "0D-2D up, 2D-14D down" pattern, with their expression level peaking at Day 2 and gradually returning close to baseline levels at Day 14 (**Figure 2A**). These dynamic genes were primarily associated with host defense biological pathways, such as defense response to viruses, regulation of viral life cycle and type I interferon signaling pathways (**Figure 2B**). (b) 263 genes follow a "0D-14D down" pattern (**Figure 2A**). This set of genes followed a monotonous trajectory, with their expression gradually decreasing until Day 14. Interestingly, we found these genes were primarily associated with mitochondrial ATP synthesis and oxidative phosphorylation, suggesting that disruption of mitochondrial respiration and ATP generation may be an important feature of the circulating cells transcriptomic shift in a SARS-CoV-2 infection (**Figure 2C**). *TrendCatcher* assigned trajectory pattern types to all dynamic genes, and provided hierarchical pie charts to visualize the composition of trajectory patterns (**Supplemental Figure 2A**).

To systematically assess and visualize the dynamic programming of the top biological pathways associated with SARS-CoV-2 infection, *TrendCatcher* generated a *TimeHeatmap*.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The *TimeHeatmap* function of *TrendCatcher* visualizes shifts in pathways by displaying the mean-fold change of individual DDEGs in a given pathway at defined time points while also depicting the number of dynamic genes and the percentage of dynamic genes within that pathway (**Figure 2D**). This quantifies the pathway level shifts over time, and allows for the magnitude of pathway change to be visualized. Together with the number and fraction of dynamic genes in a given pathway, to gauge the relative importance of the pathway during the time course.

During initial infection (Day 0 to Day2), pathways related to innate immune response and interferon pathways were highly upregulated. Examples are upregulation of pathways such as defense response to virus and regulation of innate immune response increased with an average log2 fold-change of 2.57 and 1.76 within the first day. Mucosal immune response, antimicrobial humoral response and killing of cells of other organisms were activated during the later stage of infection (Day 4 to Day 7), with an average log2 fold-change around 2. On the other hand, mitochondrial ATP synthesis coupled electron transport and protein targeting to ER, on the other hand, were gradually down-regulated until Day 14. Dynamic gene signatures from type I interferon signaling pathway and mitochondrial ATP synthesis coupled electron transport were shown using traditional heatmaps in **Supplemental Figure 2B,C**. The temporal analysis of this non-human primate dataset highlights the rapidity of interferon signaling activation btu also underscores that this initial burst of immune activation is transient and followed by a gradual downregulation of the antiviral interferon responses during the first week following infection.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Increased generation of immunoglobulin synthesis in plasma B cells as early as Day 1 of symptom onset in patients diagnosed with SARS-CoV-2 infection

Next, we analyzed the longitudinal gene expression profiles of peripheral blood mononuclear cells (PBMCs) obtained from patients diagnosed with a SARS-CoV-2 infection who were admitted to the hospital, but predominantly had uncomplicated disease progression with 4 out 5 patients showing only mild symptoms (14) (**Figure 3A**). The study performed single cell RNA-seq analysis on PBMCs, thus allowing for a cell-type specific analysis of gene expression shifts in distinct B cell subtypes, T cell subtypes and NK cells. We adopted the cell-type labels from the original study and generated pseudo-bulk RNA-seq datasets for each cell type in order to quantify changes in DDEGs for a specific cell type. We also adopted the time annotation using stages, by binning the disease processes of COVID-19 patients from symptom onset to discharge into stages 0, 1, 2, 3 and 4. We defined Stage 0 as the time point of samples obtained from healthy controls and stages 1, 2, 3 and 4 represented Day 1 to Day 16 from symptom onset. We only applied *TrendCatcher* to cell types containing more than 1,000 cells at any given stage to ensure the robustness of the results.

TrendCatcher identified 400 DDEGs in memory B cells, 213 DDEGs in naive B cells, 1,413 DDEGs in plasma B cells, 398 DDEGs in $CD4^+$ T cells, 645 DDEGs in $CD8^+$ T cells, 423 DDEGs in MAIT, 1,161 DDEGs in naive T cells and 667 DDEGs in NK cells. The *TimeHeatmap* of plasma B cells visualized the most dynamic biological pathways. Importantly, this temporal analysis of single cell RNA-seq data showed how rapidly plasma B cells ramp up the upregulation Fc-gamma receptor signaling and immunoglobulin synthesis as early as stage 1 (which corresponds to Day 1 of symptom onset) (**Figure 3B**). However, not all immunoglobulin genes are upregulated during the same temporal phase. As shown in (**Supplemental Figure 3**), genes involved in immunoglobulin synthesis show distinct temporal patterns. Due to the comparatively lower number of plasma B cells, the prominence of such increased immunoglobulin changes may be diluted in peripheral blood bulk RNA-seq analysis

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

or PBMC bulk RNA-seq analysis. However, temporal analysis of all peripheral blood mononuclear cell types, across the whole time course demonstrated increases in type I interferon signaling and defense responses to viruses as the most prominent changes over time (**Figure 3C**), thus mirroring the responses to SARS-CoV-2 we observed in the peripheral blood of non-human primates (**Figure 2**). To define cell-type specific temporal dynamics, we generated *TimeHeatmaps* for individual PBMC cell types and found significant upregulation of Type I interferon signaling in T cells, NK cells and memory B cells during Stage 1 but subsequent downregulation by Stage 2 of the disease in patients with mild symptoms (**Supplemental Figure 4A-C, Supplemental Figure 5A-D**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

TrendCatcher identifies early and persistent neutrophil activation as a hallmark of severe COVID-19

We then assessed whether a systematic temporal analysis of gene expression trends could be used to distinguish disease severity and prognosis of COVID-19 patients. We thus applied *TrendCatcher* to a time course human whole blood bulk RNA-seq dataset (11), which contained longitudinal whole blood transcriptomes from COVID-19 patients with mild, moderate and severe clinical outcomes. *TrendCatcher* identified 77 DDEGs from the mild group, 226 DDEGs from the moderate group, and 1,205 DDEGs from the severe group. Only 42 DDEGs were shared among these three groups (**Figure 4A**), and these were primarily associated with B cell mediated immunity, Fc-gamma receptor signaling pathways, humoral immune response and lymphocyte mediated immunity (**Figure 4B**). However, *TrendCatcher* identified 978 DDEGs uniquely shown in the severe COVID-19 patient group. Importantly, these genes were strongly enriched for neutrophil-related biological pathways, including neutrophil activation and neutrophil mediated immunity. These severe-disease associated genes also included genes found in pathways such as myeloid cell differentiation, reactive oxygen species metabolic process, and positive regulation of cytokine production (**Figure 4B**).

To systematically characterize which biological processes in whole blood RNA-seq samples were most dynamic in severe COVID-19 patients and how they progress over time, we applied the *TrendCatcher*'s *TimeHeatmap* function. As seen in the non-human primate PBMC and the human COVID-19 single cell RNA-seq datasets, we again observed the upregulation of genes involved in the response to virus, humoral immune response and type I interferon signaling pathway within Week 1 of disease onset (**Figure 4C**). Importantly, some dynamic biological responses were only enriched in the group of patients which subsequently progressed to severe COVID-19, including neutrophil activation (117 DDEGs, 23.4% of the corresponding GO term), blood coagulation (36 DDEGs, 10.5% of the corresponding GO term), and regulation of response to cytokine stimulus and respiratory burst (**Figure 4C, Supplemental**

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 6A,B). For instance, neutrophil activation gene expression increased by a mean of 1.3 log2-units (approximately 2.5-fold increase in mean gene expression) within Week 1, increased continuously until Week 4, and only very gradually decreased in surviving patients by Week 7. However, the summation of the averaged log2 fold-change from the *TimeHeatmap* was larger than zero, which indicates that the neutrophil activation may not have returned fully to baseline levels by 7 weeks. To confirm this, we applied LOESS smooth curve fitting to all neutrophil activation DDEGs identified from the three severity groups. LOESS fitting confirmed that severe COVID-19 patients showed markedly higher neutrophil activation at the early stage of infection (Week 1- Week 2), and also remained highly activated even after 7 weeks. Such a persistent neutrophil activation was not observed in either mild or moderate COVID-19 patient groups (**Figure 4D**).

We also observed severe COVID-19 patients showed evidence of greater humoral immune response gene upregulation (**Figure 4E**), as well as markedly higher upregulation blood coagulation genes (**Figure 4F**) and respiratory burst genes (**Figure 4G**), which remained upregulated even after Week 6. These were not found to be upregulated in patients with mild or moderate COVID-19 disease. These DDEGs and dynamic biological pathways may thus be suited to serve as early biomarkers that distinguish severe COVID-19 from mild and moderate COVID. All these dynamic gene signatures from neutrophil activation, blood coagulation and respiratory burst pathways were listed using traditional heatmap (**Supplemental Figure 7, Supplemental Figure 8A,B**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Early impaired IFN-I signaling in PBMC provides a hallmark of severe COVID-19

Next, to define the cell-type specific dynamic gene signatures and biological processes, we used *TrendCatcher* to analyze a human PBMC single cell RNA-seq dataset time course dataset in which patients were categorized as having either moderate or severe COVID-19 (15). Importantly, as this dataset only contained mRNA from peripheral blood mononuclear cells, it lacked mRNA from neutrophils. *TrendCatcher* generated "pseudo-bulk" mRNA profiles for each cell type in order to perform the analysis of gene expression dynamics in a cell-type specific manner. *TrendCatcher* identified more dynamic shifts in almost all cell types from severe COVID-19 patients than moderate groups (**Supplemental Table 1, 2**). To identify dynamic responsive processes unique to the severe group, we compared GO enrichment for each cell type between severe and moderate. As shown in **Figure 5A**, all innate immune cells and some adaptive immune cells (including B cells, $CD8^+$ T cells) from moderate COVID-19 were highly enriched in type I interferon signaling, negative regulation of viral process and defense to virus, whereas this was not observed significantly enriched in severe COVID-19. Severe COVID-19 patients were highly dynamic in MAPK cascade, NF-κB signaling, T cell receptor signaling and positive regulation of cytokine production for both NK cells and $CD4^+$ T cells. For monocytes and DC (dendritic cells), no uniquely enriched dynamic biological processes were observed in severe COVID-19 patients versus moderate COVID-19 patients.

Furthermore, we also found the extent of the interferon signaling response to be the key distinguishing feature between moderate and severe COVID-19. To quantitatively compare the trajectory differentiation of the type I interferon signaling pathway, we performed the LOESS smooth fitting to the DDEGs identified from this pathway. As shown in **Figure 5B**, there is a profound separation between these two groups, with early strong activation of IFN-I signaling in NK cells, monocytes, B cells and $CD8^+$ T cells in moderate COVID-19 patients but not in severe COVID-19 patients. Patients with moderate COVID-19 showed activation of type I interferon signaling pathways whereas patients with severe COVID-19 had a blunted activation of type I interferon signaling. This is also shown in the cell-type specific

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

TimeHeatmap. As shown in **Figure C,D**, although NK cells from both moderate and severe COVID-19 patients demonstrated activation of type I interferon response within the first week, moderate COVID-19 exhibited a stronger activation than severe group, with average 1.62 log2 fold-change compared to 0.87 log2 fold-change. In $CD8^+$ T cells, only moderate COVID-19 groups were observed to have a strong type I interferon response within the first week. On the other hand, $CD8^+$ T cells in patients who would go on to develop severe COVID-19, showed upregulation of cell proliferation and cell differentiation genes instead. Together, these data suggest that a robust early Type I interferon response in PBMCs is associated with reduced severity of COVID-19.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

TrendCatcher identifies metabolic gene expression shifts in NK cells as a hallmark response to COVID-19 vaccination

We next applied *TrendCatcher* to a longitudinal human PBMC single cell RNA-seq vaccination dataset (16), to provide insights into how the immune system physiologically responds to mRNA vaccines over time. The vaccination study collected single-cell PBMCs from 56 healthy volunteers vaccinated with the Pfizer-BioNtech at Day 0, 1, 2, 7, 21, 22, 28, and 42 after the vaccination. We processed each patient's scRNA-seq individually. We clustered cells using the Seurat algorithm (17) and annotated the cell types using SingleR (18). Then we remove cell types containing less than 1,000 cells for each time point across all samples. As **Figure 6A** shown, it is the UMAP of one patient's PBMCs single cell data at Day 0.

TrendCatcher identified 650 DDEGs in NK cells, 450 in B cells, 23 in $CD8^+$ T cells, 62 in monocytes and only 6 in $CD4^+$ T cells. This indicates NK cells exhibit a strong dynamic gene expression shift after vaccination. After comparing the GO enrichment analysis across cell types, we found NK cells gene expression shifts were enriched in metabolic processes in response to the Pfizer-BioNtech SARS-CoV-2 mRNA vaccine, such as regulation of cellular amino acid metabolism and ATP metabolism (**Figure 6B**). As shown in **Figure 6C**, a *TimeHeatmap* of NK cells shows 54 DDEGs, which account for nearly half of the ATP metabolic process pathway genes and exhibited a mean of 0.44 log2 fold upregulation after the first dose of vaccination. These findings indicate the reprogramming of metabolism in NK cells may be an indicator of an intact vaccine response. Another hallmark of the vaccine response was the activation of interferon pathways, including type I interferon signaling pathway and response to interferon-gamma (**Supplemental Figure 9A**). Importantly, the *TimeHeatmap* demonstrates that these gene expression shifts were very transient, usually decreasing by Day 2 or Day 7, and were again upregulated when the subjects received the booster vaccine at Day 21. Interferon pathways showed an average increase of 0.71 and 1.42 log2 fold-change. In adaptive immune cells, we also observed strong type I interferon signaling

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

pathways activated in both B cells and $CD8^+$ T cells (**Supplemental Figure 9B,C**). Additionally, B cells demonstrated upregulation of antigen processing and protein synthesis (**Supplemental Figure 9B**), which were upregulated only for the first day, and then began decreasing. It is noteworthy that early upregulation of interferon pathways in the PBMCs of vaccine recipients mirrors that seen in patients with moderate COVID-19 infection, consistent with the notion that early and transient interferon upregulation is a hallmark of a healthy immune response to the SARS-CoV-2 infection.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

Temporal analysis of gene expression is gaining importance in the analysis of complex dynamic processes such as disease progression. Besides gene dynamic pattern characterization, time-course gene expression data are also used to infer regulatory and signaling relationships among genes (19, 20). Integrating with other different types of measurements, such as pathology and infection over time helps disentangle the complex dynamic processes and possible underlying mediators (21). Thus, accurate identification of dynamic differentially expressed genes (DDEGs) in a time course RNA-seq or single cell RNA-seq study can help identify time-dependent disease mechanisms, adaptive and maladaptive molecular signatures as well as potential biomarkers that may be associated with disease severity.

In recent years, tools have been implemented to characterize time course RNA-seq data; however, these tools were focused on either bulk RNA-seq datasets (3, 6, 7) and few infer trajectories from scRNA-seq (22, 23). Compared to methods that analyze time course data without considering the sequential nature of time points, modeling time as a continuous function avoids a relative loss of statistical testing power, especially when many multiple time points were studied. There is also a need for methods which combine DDEGs identification with the visualization of dynamic pathway shifts.

In this study, we developed an open-source R-software package to perform temporal analysis in longitudinal bulk RNA-seq and scRNA-seq datasets. *TrendCatcher* can identify DDEGs and infer the trajectories for each DDEG. By combining a time-window sliding strategy on inferred gene trajectories and leveraging annotated biological pathway databases, *TrendCatcher* can infer and visualize the most dynamic biological pathways in response to the external stimuli. Furthermore, by utilizing the *TimeHeatmap* function, *TrendCatcher* can help researchers identify the magnitude and dynamic nature of pathways shifts. Using simulated datasets for benchmarking, *TrendCatcher* achieved higher accuracy (AUC = 0.9) compared to other

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

commonly used methods for the analysis of temporal gene expression datasets, when analyzing three or more time points. The advantage of using *TrendCatcher* was especially apparent in the setting of biphasic or multiphasic temporal trajectories, in which gene expression levels can change their trend of upregulation or downregulation during the time course.

Despite the extraordinary success of rapidly developed and deployed mRNA vaccines against SARS-CoV-2, the ongoing COVID-19 pandemic remains a major global health problem in part due to the emergence of newer highly contagious SARS-CoV-2 variants of concern as well as vaccine hesitancy. This requires the identification of novel mechanistic targets, especially in vulnerable patients who have a high risk of developing severe COVID-19. One of the key pathogenic factors driving COVID-19 severity is the profound immune dysregulation observed in patients with severe COVID-19 that can result in respiratory failure (24-27). Human immune responses to SARS-CoV-2 infection are highly dynamic and time dependent, requiring upregulation as well as downregulation of distinct immune signaling pathways at the appropriate times to ensure optimal host defense (24-27). Understanding the dynamics of the COVID-19 immune response could form the basis of developing therapies that are appropriate for a given time window (28). Personalized medicine or precision medicine are gaining traction by developing therapies tailored to patients based on their genotype and phenotype, but personalization likely also requires tailoring therapies based on the temporal phase of a disease. These studies highlight the need for time-course or longitudinal analyses of the host responses to the SARS-CoV-2 infection.

In this study, we applied *TrendCatcher* to systematically analyze sequential blood samples from either non-human primates infected with SARS-CoV-2, human patients with COVID-19 of varying severity or human subjects who received a SARS-CoV-2 mRNA vaccine. *TrendCatcher* identified dynamic gene expression and biological pathway signatures for (a) SARS-CoV-2 infection progression over time, (b) severe COVID-19 vs. moderate or mild

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

COVID-19 and (c) COVID-19 vaccine recipients versus control. *TrendCatcher* established response to virus, humoral immune response and type I interferon signaling pathway activation across peripheral blood cell types in mild, moderate and severe COVID-19. However, we found temporal patterns of gene expression shifts that were unique in the severe COVID-19 patients. Severe COVID-19 was associated with marked activation of neutrophils and upregulation of coagulation pathways as well as blunted type I interferon signaling as early as week 1 in the peripheral blood of patients. Importantly, severe COVID-19 was associated with persistent activation of neutrophils and genes regulating coagulation for as long as 6 weeks, underscoring that the importance of the temporal analysis by *TrendCatcher* which identified hallmarks of severe COVID-19 in peripheral blood samples.

Our findings complement recent studies which implicate neutrophils in the excessive inflammation, coagulopathy, immunothrombosis and organ damage which characterize severe COVID-19 (29-33). Neutrophils are particularly active in highly vascularized organs, such as lungs and kidneys, which are prime targets of SARS-CoV-2 induced injury in COVID-19 (34). Dysregulated neutrophil responses, such as prolonged activation may cause damage to vessels and underlying parenchymal (35-37). Studies also show that activated neutrophils through Toll-like receptors, chemokine- and cytokine receptors can stimulate the neutrophil extracellular trap (NET) formation (29). And recent studies showed the disbalance between NET formation and degradation can trigger immunothrombosis and tissue injury (38, 39). Excessive activation of macrophages and adaptive immune cells in severe COVID-19 which can form vicious cycles of positive feedback circuits has also been demonstrated (40, 41), but less is known about how early neutrophils are activated in disease progression because many of the studies on neutrophils in clinical COVID-19 or animal models of severe SARS-CoV-2 focused on the late stages of the disease as well as the neutrophils in the lung tissue. In our study, we used a systematic temporal analysis and observed that profound neutrophil activation in the peripheral blood, which was predominant in the severe COVID-19 group,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

occurs as early as week one after diagnosis or symptoms persists even after six or seven weeks in surviving patients.

Our observation of early upregulation of coagulation genes in the whole blood transcriptomes of severe COVID-19 patients also points to another feature that is associated with severe COVID-19, thrombotic and embolic complications such as strokes (13, 42, 43). Recent studies have found the procoagulant changes in endothelial cells underly the coagulopathy in severe COVID-19 (44). Endothelial dysfunction in COVID-19 patients may be exacerbated through inflammatory cytokines and neutrophil extracellular traps, thus pointing to interactions between circulating or recently recruited neutrophils and the vessel wall endothelial cells that are in contact with circulating immune cells as drivers of such coagulopathic manifestations (45-47). We believe the analysis of the RNA-seq and scRNA-seq data support neutrophil activation and upregulation of coagulation genes as defining features of severe COVID-19, highlighting their role as early biomarkers to provide prognostic information and thus optimally treat patients who have a higher likelihood of progressing to severe disease. Importantly, our results also raise questions about the role of neutrophils and of coagulation in what is referred to as “long COVID” or “post-acute sequelae of COVID” (48-50), because we observed that the activation persists for as long as six to seven weeks after the initial infection.

Our temporal analysis showed that severe COVID-19 single-cell PBMCs were characterized by impaired type I interferon signaling at the onset of infection (Week 1), compared to the moderate COVID-19 group. This impaired type I interferon signaling was identified in innate cells, such as NK cells and monocytes, and adaptive cells, such as B cell and $CD8^+$ T cells. Type I interferons, which are essential for antiviral immunity (51, 52), have been shown to be upregulated in COVID-19 (53, 54). Other studies also suggested that impaired type I interferon signaling may promote severe COVID-19 and that interferon therapy could be used as therapy in severe COVID-19 (55). However, since these studies were not designed for longitudinal analyses, the timing of when to intervene on IFN signaling was not clear. Traditional analyses

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of gene expression data are not suited for the identification of temporal windows and thus do not address whether impaired IFN production is present at the onset of infection, whether it was delayed, or whether IFN production is exhausted after an initial activation (54). *TrendCatcher* addressed this question by showing the dynamic timing of type I IFN from both the moderate group and severe group.

In our COVID-19 vaccination single cell PBMC temporal data analysis, we also identified prominent metabolic shifts in NK cells. Cellular metabolism is recognized as an important factor that can determine the fate and function of lymphocytes (56, 57). Certain metabolic pathways have been shown to have direct immunoregulatory roles (56). Activated NK cells engage in a robust metabolic response, which is required for normal effector functions (58), and our data suggests that assessing NK metabolic shifts may be an intriguing alternative to assessing vaccine responsiveness, which does not rely on antibody levels which can be highly variable in laboratory assays.

In conclusion**,** we have developed the novel *TrendCatcher* R package platform designed for time course RNA-seq data analysis which identifies and visualizes distinct dynamic transcriptional programs. When applied to real whole blood bulk RNA-seq time course datasets from COVID-19 patients, we observed that patients with severe COVID-19 showed gene expression profiles consistent with profound neutrophil activation and coagulopathy early during the progression of the disease (starting from the first week of symptom onset). This indicates that neutrophil activation may be a key early biomarker and possible mediator of severe COVID-19. The distinct kinetics of gene expression shifts identified by *TrendCatcher* in COVID-19 patients may allow for the identification of therapeutic targets during distinct temporal phases.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Author contributions

X.W. and J.R. designed the study, X.W. performed the analyses and wrote the first draft of the manuscript, J.R. co-wrote the initial draft of the manuscript. M.S. and Y.D. provided critical revisions of the manuscript.

Acknowledgments

The studies were supported by NIH grants R01-HL149300 (to JR), P01-HL151327 (to JR), and T32-HL139439 (to MS).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Materials and Methods

TrendCatcher framework

The main components of the *TrendCatcher* framework are shown in **Figure 1**. *TrendCatcher* requires two main inputs: the raw count table C of a temporal study with a dimension of $m \times n$, where m denotes the number of genes and n denotes the number of samples, and a user-defined baseline time variable T, such as "0 hour". Since samples may have different sequencing depths and batch effect, *TrendCatcher* integrates with limma (5) and provides preprocessing steps, such as batch correction and normalization. For single-cell RNA sequencing datasets, *TrendCatcher* extracts cells for each cell type annotated in the meta data slot of Seurat object and convert it into cell-type specific "pseudobulk" time course RNA library. Based on a user-specified threshold, relatively low abundant genes are removed from the count table, and reads are normalized and batch effects are removed. *TrendCatcher*'s core algorithm is composed of five main steps: (a) baseline fluctuation confidence interval estimation, (b) model dynamic longitudinal count, (c) time point dynamic p-value calculation, (d) gene-wise dynamic p-value calculation, and (e) break point screening and gene-wise dynamic pattern assignment. Mathematical details will be expanded in the following sections. For the output of *TrendCatcher*, there are mainly two components: a master table and a set of functions for versatile visualization purposes. The master table contains all the dynamic details of each single gene, including its dynamic p-value, its break point location time, and its dynamic trajectory pattern. In addition to the master table, *TrendCatcher* produces five main types of visualizations: (a) a figure showing the observed counts and fitted splines of each gene, (b) genes trajectories from each trajectory pattern group (c) a hierarchical pie chart that represents trajectory pattern composition, (d) a *TimeHeatmap* to infer trajectory dynamics of top dynamic biological pathways and (e) a two-sided bar plot to show the top most positively and negatively changed (averaged accumulative log2FC) biological pathways.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Baseline fluctuation confidence interval estimation

We assumed that the observed number of RNA-seq reads count from the baseline time (e.g., t=0 hour) $X_{i,t_{baseline}}$ was generated from a negative binomial distribution (Equation 1):

$$X_{i,t_{baseline}} \sim NB(\mu_{i,t_{baseline}}, \varphi_i) \tag{1}$$

Where gene $i = 1, \dots, n$ and $\mu_{i,t_{baseline}}$ is the mean count of gene i at baseline time, and φ_i is the dispersion factor. First, the dispersion factor φ_i was pre-estimated as a constant hyperparameter for each gene with DESeq2 (3), as shown in Equation 2. Here, $\sigma_i(t)$ is the variance at time t.

$$\sigma_i(t)^2 = \mu_i(t) + \varphi_i * \mu_i(t)^2 \tag{2}$$

Then, $\mu_{i,t_{baseline}}$ was estimated using maximum likelihood from Equation 3.

$$\hat{\mu}_{i,t_{baseline}} = argmax\ L_{NB}(u, \varphi) = argmax\ \frac{\Gamma(\varphi + x)}{x!\,\Gamma(\varphi)} \left(\frac{\mu}{\mu + \varphi}\right)^x \left(\frac{\varphi}{\varphi + \mu}\right)^\varphi \tag{3}$$

Based on the negative binomial distribution and the estimated mean count for baseline time, we constructed the 90% confidence interval $[X_{i,t_{baseline},0.05}, X_{i,t_{baseline},0.95}]$ as the baseline count fluctuation interval (Equation 4(a) and Equation 4(b)).

$$P\left(x_{i,t_{baseline}} \leq X_{i,t_{baseline},0.05} \mid \mu_{i,t_{baseline}}, \varphi_i\right) = 0.05 \tag{4a}$$

$$P\left(x_{i,t_{baseline}} \geq X_{i,t_{baseline},0.95} \mid \mu_{i,t_{baseline}}, \varphi_i\right) = 0.95 \tag{4b}$$

Model dynamic longitudinal count estimation

To model the time-dependent gene expression value, we applied a smoothing spline ANOVA model (59, 60) with a negative binomial family constraint to fit the reads from samples across non-baseline multiple time points. The random variable $X_{i,t(t \neq t_{baseline})}$ is assumed to follow the NB distribution in Equation 5, with a positive integer α represents number of failures before the αth success in a sequential of Bernoulli trials, and $p(t) \in (0,1)$ represents the success probability.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$X_{i,t(t\neq t_{baseline})} \sim NB(\alpha, p(t)) \tag{5}$$

The log-likelihood given a time-course observed count $x = [x_{i,t(t\neq t_{baseline})}\}_{i=1,\ldots,n;\ t=1,\ldots,T}$ is calculated as Equation 6.

$$L = log\ L(\alpha, p(t) \mid X = x\ )$$

$$= \sum_{t=1}^{T} \sum_{i=1}^{n} [x_{i,t(t\neq t_{baseline})}\ log\ (1 - p(t) + \alpha log p(t) + log\Gamma(\alpha + x_{i,t(t\neq t_{baseline})}) - loglog\Gamma(\alpha)] + C \tag{6}$$

Taking the logit link and model time effect, we define the logit link η

$$\eta = log\ \frac{p(t)}{1 - p(t)} \tag{7}$$

To allow for flexibility in the estimation of η , and find the best trade-off between goodness of fit and the smoothness of the spline curve, soft constraints of the form $J(\eta)$ is added to the minus log-likelihood, with the smoothing parameter $\lambda > 0$.

$$-L + \lambda\ J(\eta) \tag{8}$$

The solution to the optimization of Equation 8 leads to a smoothing fitting to the reads from samples across different non-baseline time points. The estimated mean of $\mu_{i,t(t\neq t_{baseline})}$ can be estimated using Equation 9.

$$\mu_{i,t(t\neq t_{baseline})} = \hat{\alpha} e^{\hat{\eta}(t)} = \frac{\hat{\alpha}\hat{p}(t)}{1 - \hat{p}(t)} \tag{9}$$

Gene's dynamic p-value calculation

To calculate gene's non-baseline dynamic signal significance, each gene's non-baseline estimated mean count $\mu_{i,t(t\neq t_{baseline})}$ was tested against the baseline fluctuation interval. Based on Equation 10(a) and Equation 10(b), for each gene at each single non-baseline time point, a dynamic time p-value was calculated.

$$If\ \mu_{i,t(t\neq t_{baseline})} \geq \mu_{i,t_{baseline}},\ then\ p_{i,t} = \int P(x \geq \mu_{i,t(t\neq t_{baseline})} \mid \mu_{i,t_{baseline}}, \varphi_i) \tag{10a}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$If\ \mu_{i,t(t\neq t_{baseline})} \leq \mu_{i,t_{baseline}},\ then\ p_{i,t} = \int P(x \leq \mu_{i,t(t\neq t_{baseline})} \mid \mu_{i,t_{baseline}}, \varphi_i) \tag{10b}$$

Then, we applied Fisher's combined probability test method to calculate a gene-wise dynamic p-value (Equation 11):

$$\chi^2_{i,2t} \sim -2\sum_{t=0}^{T} ln\ (p_{i,t}) \tag{11}$$

Trajectory pattern assignment

First, we connect all the significant dynamic signal time points, with a p-value threshold less than 0.05. Then we applied a break point searching strategy to capture the gene expression change trend. The definition of break point is defined using Equation 12(a) and Equation 12(b).

$$If\ \mu_{i,t} > \mu_{i,t_{next}}\ \ AND\ \ \mu_{i,t} > \mu_{i,t_{previous}}, break\ point\ type\ I. \tag{12a}$$

$$If\ \mu_{i,t} < \mu_{i,t_{next}}\ \ AND\ \ \mu_{i,t} < \mu_{i,t_{previous}}, break\ point\ type\ II. \tag{12a}$$

There are two types of break points: type I means gene up-regulated followed by a down regulation, and type II means gene expression level down regulated and then followed by an upregulation. By screening along the break point, the master-pattern and sub-pattern were assigned to each gene.

***TimeHeatmap* enrichment analysis and Two-sided bar plot**

To build the *TimeHeatmap* for visualizing the biological pathway enrichment change over time, we designed a window-sliding strategy to capture all the up-regulated or down-regulated genes within each time interval. If we denote time vector as t_j and $j \in 1, \ldots, T$, each time interval is denoted as $[t_{j-1}, t_j]$. We found N_{up} up-regulated genes and N_{down} down-regulated genes within the time window $[t_{j-1}, t_j]$, then Fisher's exact test was performed to obtain the Gene Ontology (GO) term enrichment with the corresponding time interval for up-regulated genes and down-regulated genes separately. Users can select the top most enriched

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

biological pathways (ranked by enrichment p-value) for each time interval (by default is 10). Then for each selected GO term within the corresponding time window, we calculated the averaged log 2 fold-change (Avg_log2FC) of all the DDEGs from this GO term. A series of Avg_log2FC values over time characterize the trajectory dynamics of the corresponding biological pathway, it is defined as biological pathway trajectory inference in this study. The summation of the series of Avg_log2FC estimates the averaged accumulative log 2 fold-change (GO_mean_logFC) for the corresponding GO term. *TrendCatcher* ranks biological pathways based on their dynamic magnitude inferred from the GO_mean_logFC value. Users are free to choose the top most positively and negatively changed (averaged accumulative log2FC) biological pathways. Besides GO enrichment analysis (61), *TrendCatcher* also packages Enrichr (62) biological pathway databases. Visualization is constructed using the ComplexHeatmap (63) package.

Simulated dataset

To mimic the real biological RNA-seq dataset, we only allowed ~10% of the genes to be dynamic responsive genes. In this study, we embedded 5 different types of trajectories into the temporal RNA-seq simulated datasets, including non-dynamic trajectory (~90%), monotonic trajectory (~2.5%), bi-phase trajectory (~2.5%), and multimodal trajectory (~5%) including 2 break-point and 3 break-point trajectory. Each type of trajectory was constructed by adding negative binomial distribution noise to the embedded trajectory count. For example, for monotonic trajectory, we defined the first and last time-points' RNA expression level change to be 0.5-2 log2 fold-change, then we added negative binomial distribution noise to the embedded monotonic trajectory. To check how the AUC changes as time-course study extends longer, we sampled 3, 5, 7, 9 and 11 different number of time points with evenly time intervals, and randomly sampled 3 replicates for each time point. To validate how the prediction AUC varies across different types of trajectories, we embedded ~10% of the genes to have non-dynamic trajectories, and ~90% of the genes have a specific dynamic trajectory (monotonic, bi-phase or multimodal).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Pseudobulk RNA library construction

To construct "pseudo-bulk" RNA library from single cell RNA-seq datasets, all cells for each cell type in a given sample were computationally "pooled" by summing all counts for a given gene. Since pseudobulk libraries composed of few cells are not likely modeled properly, we removed cell types containing less than 1k cells in this study. Lowly expressed genes were removed for each cell type as well, using the filterByExpr function from *edgeR* R package (4). Gene counts were transformed using the log of the counts per million (cpm) and library size was normalized using calcNormFactors function with the method = "RLE".

Gene set enrichment of DDEGs

Gene sets enrichment analysis in this study was performed using clusterProfiler (61) R package, and enrichment comparison from multiple groups were visualized using the compareCluster function.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Dataset	Organism	Number of time points	Study	Figure	Dataset Source	Reference
Bulk Peripheral Blood (Aid et al.)	Macaca mulatta	7	Covid vs. Normal	Figure 2	GSE156701	(13)
scRNA PBMC (Zhu et al.)	Homo Sapien	5	Covid vs. Normal	Figure 3	https://db.cngb.org/search/project/CNP0001102/	(14)
Bulk Peripheral Blood (Bergamaschi et al.)	Homo Sapien	8	Severe vs. Moderate	Figure 4	https://www.covid19cellatlas.org/patient/citiid/	(11)
scRNA PBMC (Liu et al.)	Homo Sapien	4	Severe vs. Moderate	Figure 5	GSE161918	(15)
scRNA PBMC (Arunachalam et al.)	Homo Sapien	8	Vaccination	Figure 6	GSE171964 and GSE169159	(16)

Table 1. Datasets presented in this study

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Innate immune cells					
NK		Monocyte		DC	
Moderate	Severe	Moderate	Severe	Moderate	Severe
708	1,179	1,479	1,845	22	17

Supplement Table 1. Number of DDEGs for each innate immune cell type identified by *TrendCatcher*. Using FDR less than 0.05.

Adaptive immune cells							
B		$CD4^+$ T		$CD8^+$ T		MAIT	
Moderate	Severe	Moderate	Severe	Moderate	Severe	Moderate	Severe
969	1,344	190	1,060	634	817	71	13

Supplement Table 2. Number of DDEGs for each adaptive immune cell type identified by *TrendCatcher*. Using FDR less than 0.05.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1. Bar-Joseph Z, et al. Studying and modelling dynamic biological processes using time-series gene expression data. *Nat Rev Genet.* 2012;13(8):552-64.

2. Hwang B, et al. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental and Molecular Medicine.* 2018;50.

3. Love MI, et al. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.

4. McCarthy DJ, et al. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 2012;40(10):4288-97.

5. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.

6. Fischer DS, et al. Impulse model-based differential expression analysis of time course sequencing data. *Nucleic Acids Res.* 2018;46(20):e119.

7. Nueda MJ, et al. Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics.* 2014;30(18):2598-602.

8. Zhou F, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet.* 2020;395(10229):1054-62.

9. Wu ZY, and McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *Jama-Journal of the American Medical Association.* 2020;323(13):1239-42.

10. Pedersen SF, and Ho YC. SARS-CoV-2: a storm is raging. *Journal of Clinical Investigation.* 2020;130(5):2202-5.

11. Bergamaschi L, et al. Longitudinal analysis reveals that delayed bystander CD8(+) T cell activation and early immune pathology distinguish severe COVID-19 from mild disease. *Immunity.* 2021;54(6):1257-+.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

12. Bernardes JP, et al. Longitudinal Multi-omics Analyses Identify Responses of Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19. *Immunity.* 2020;53(6):1296-+.
13. Aid M, et al. Vascular Disease and Thrombosis in SARS-CoV-2-Infected Rhesus Macaques. *Cell.* 2020;183(5):1354-+.
14. Zhu LN, et al. Single-Cell Sequencing of Peripheral Mononuclear Cells Reveals Distinct Immune Response Landscapes of COVID-19 and Influenza Patients. *Immunity.* 2020;53(3):685-+.
15. Liu C, et al. Time-resolved systems immunology reveals a late juncture linked to fatal COVID-19. *Cell.* 2021;184(7):1836-+.
16. Arunachalam PS, et al. Systems vaccinology of the BNT162b2 mRNA vaccine in humans. *Nature.* 2021;596(7872):410-+.
17. Hao YH, et al. Integrated analysis of multimodal single-cell data. *Cell.* 2021;184(13):3573-+.
18. Aran D, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology.* 2019;20(2):163-+.
19. Krieger G, et al. Independent evolution of transcript abundance and gene regulatory dynamics. *Genome Research.* 2020;30(7):1000-11.
20. Yuan Y, and Bar-Joseph Z. Deep learning for inferring gene relationships from single-cell expression data. *Proceedings of the National Academy of Sciences of the United States of America.* 2019;116(52):27151-8.
21. Graw F. DISEASE Deciphering the triad of infection, immunity and pathology. *Elife.* 2021;10.
22. Van den Berge K, et al. Trajectory-based differential expression analysis for single-cell sequencing data. *Nature Communications.* 2020;11(1).
23. Qiu XJ, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods.* 2017;14(10):979-+.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

24. Chen G, et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. *Journal of Clinical Investigation.* 2020;130(5):2620-9.

25. Gong J, et al. Correlation analysis between disease severity and inflammation-related parameters in patients with COVID-19: a retrospective study. *Bmc Infectious Diseases.* 2020;20(1).

26. Mehta P, et al. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet.* 2020;395(10229):1033-4.

27. Qin C, et al. Dysregulation of Immune Response in Patients With Coronavirus 2019 (COVID-19) in Wuhan, China. *Clinical Infectious Diseases.* 2020;71(15):762-8.

28. Goyal A, et al. Potency and timing of antiviral therapy as determinants of duration of SARS-CoV-2 shedding and intensity of inflammatory response. *Science Advances.* 2020;6(47).

29. Ackermann M, et al. Patients with COVID-19: in the dark-NETs of neutrophils. *Cell Death and Differentiation.* 2021;28(11):3125-39.

30. Morrissey SM, et al. A specific low-density neutrophil population correlates with hypercoagulation and disease severity in hospitalized COVID-19 patients. *Jci Insight.* 2021;6(9).

31. Cambier S, et al. Atypical response to bacterial co-infection and persistent neutrophilic broncho-alveolar inflammation distinguish critical COVID-19 from influenza. *JCI Insight.* 2021.

32. Kaiser R, et al. Self-sustaining IL-8 loops drive a prothrombotic neutrophil phenotype in severe COVID-19. *Jci Insight.* 2021;6(18).

33. Freire PP, et al. The relationship between cytokine and neutrophil gene network distinguishes SARS-CoV-2-infected patients by sex and age. *Jci Insight.* 2021;6(10).

34. Libby P, and Luscher T. COVID-19 is, in the end, an endothelial disease. *European Heart Journal.* 2020;41(32):3038-44.

35. Bardoel BW, et al. The Balancing Act of Neutrophils. *Cell Host & Microbe.* 2014;15(5):526-36.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

36. Metzemaekers M, et al. Kinetics of peripheral blood neutrophils in severe coronavirus disease 2019. *Clinical & Translational Immunology.* 2021;10(4).

37. Papayannopoulos V. Neutrophil extracellular traps in immunity and disease. *Nature Reviews Immunology.* 2018;18(2):134-47.

38. Zuo Y, et al. Autoantibodies stabilize neutrophil extracellular traps in COVID-19. *Jci Insight.* 2021;6(15).

39. Zuo Y, et al. Neutrophil extracellular traps in COVID-19. *JCI Insight.* 2020;5(11).

40. Merad M, and Martin JC. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. *Nature Reviews Immunology.* 2020;20(6):355-62.

41. Grant RA, et al. Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia. *Nature.* 2021;590(7847).

42. Ackermann M, et al. Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. *New England Journal of Medicine.* 2020;383(2):120-8.

43. Rapkiewicz AV, et al. Megakaryocytes and platelet-fibrin thrombi characterize multi-organ thrombosis at autopsy in COVID-19: A case series. *Eclinicalmedicine.* 2020;24.

44. Schmaier AA, et al. Tie2 activation protects against prothrombotic endothelial dysfunction in COVID-19. *Jci Insight.* 2021;6(20).

45. Del Valle DM, et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nature Medicine.* 2020;26(10):1636-+.

46. Leisman DE, et al. Cytokine elevation in severe and critical COVID-19: a rapid systematic review, meta-analysis, and comparison with other inflammatory syndromes. *Lancet Respiratory Medicine.* 2020;8(12):1233-44.

47. Skendros P, et al. Complement and tissue factor-enriched neutrophil extracellular traps are key drivers in COVID-19 immunothrombosis. *Journal of Clinical Investigation.* 2020;130(11):6151-7.

48. Carfi A, et al. Persistent Symptoms in Patients After Acute COVID-19. *Jama-Journal of the American Medical Association.* 2020;324(6):603-5.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

49. Huang CL, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet.* 2021;397(10270):220-32.

50. Nalbandian A, et al. Post-acute COVID-19 syndrome. *Nature Medicine.* 2021;27(4):601-15.

51. McNab F, et al. Type I interferons in infectious disease. *Nature Reviews Immunology.* 2015;15(2):87-103.

52. Muller U, et al. FUNCTIONAL-ROLE OF TYPE-I AND TYPE-II INTERFERONS IN ANTIVIRAL DEFENSE. *Science.* 1994;264(5167):1918-21.

53. Blanco-Melo D, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell.* 2020;181(5):1036-+.

54. Hadjadj J, et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science.* 2020;369(6504):718-+.

55. Darazam IA, et al. Role of interferon therapy in severe COVID-19: the COVIFERON randomized controlled trial. *Scientific Reports.* 2021;11(1).

56. Loftus RM, and Finlay DK. Immunometabolism: Cellular Metabolism Turns Immune Regulator. *Journal of Biological Chemistry.* 2016;291(1):1-10.

57. O'Neill LAJ, et al. A guide to immunometabolism for immunologists. *Nature Reviews Immunology.* 2016;16(9):553-65.

58. O'Brien KL, and Finlay DK. Immunometabolism and natural killer cell responses. *Nature Reviews Immunology.* 2019;19(5):282-90.

59. C G. Smoothing spline ANOVA models: R package gss. *J Stat Softw.* 2014;58(5):25.

60. Metwally AA, et al. MetaLonDA: a flexible R package for identifying time intervals of differentially abundant features in metagenomic longitudinal studies. *Microbiome.* 2018;6(1):32.

61. Yu G, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284-7.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

62. Kuleshov MV, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016;44(W1):W90-7.

63. Gu ZG, et al. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32(18):2847-9.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

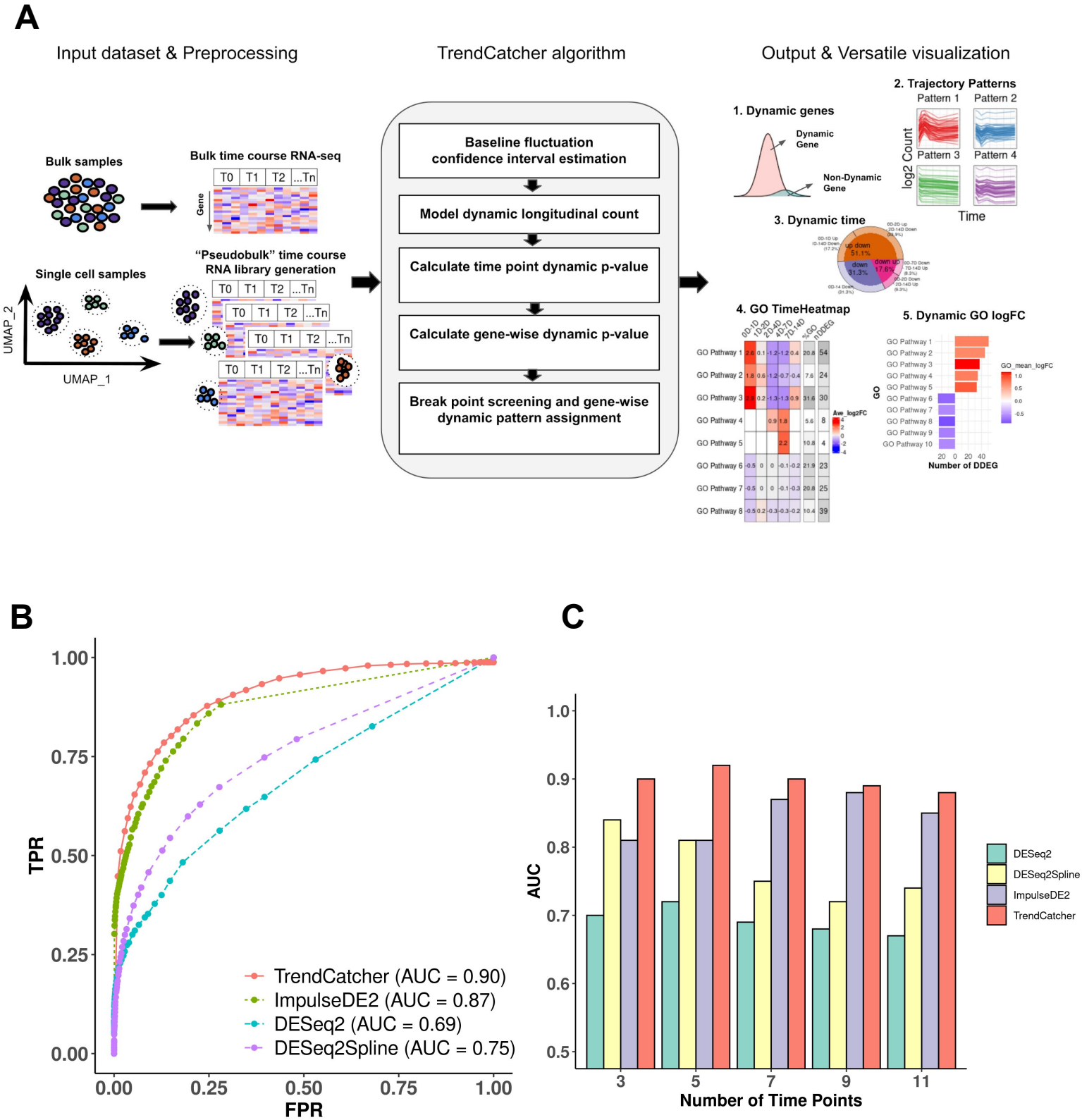
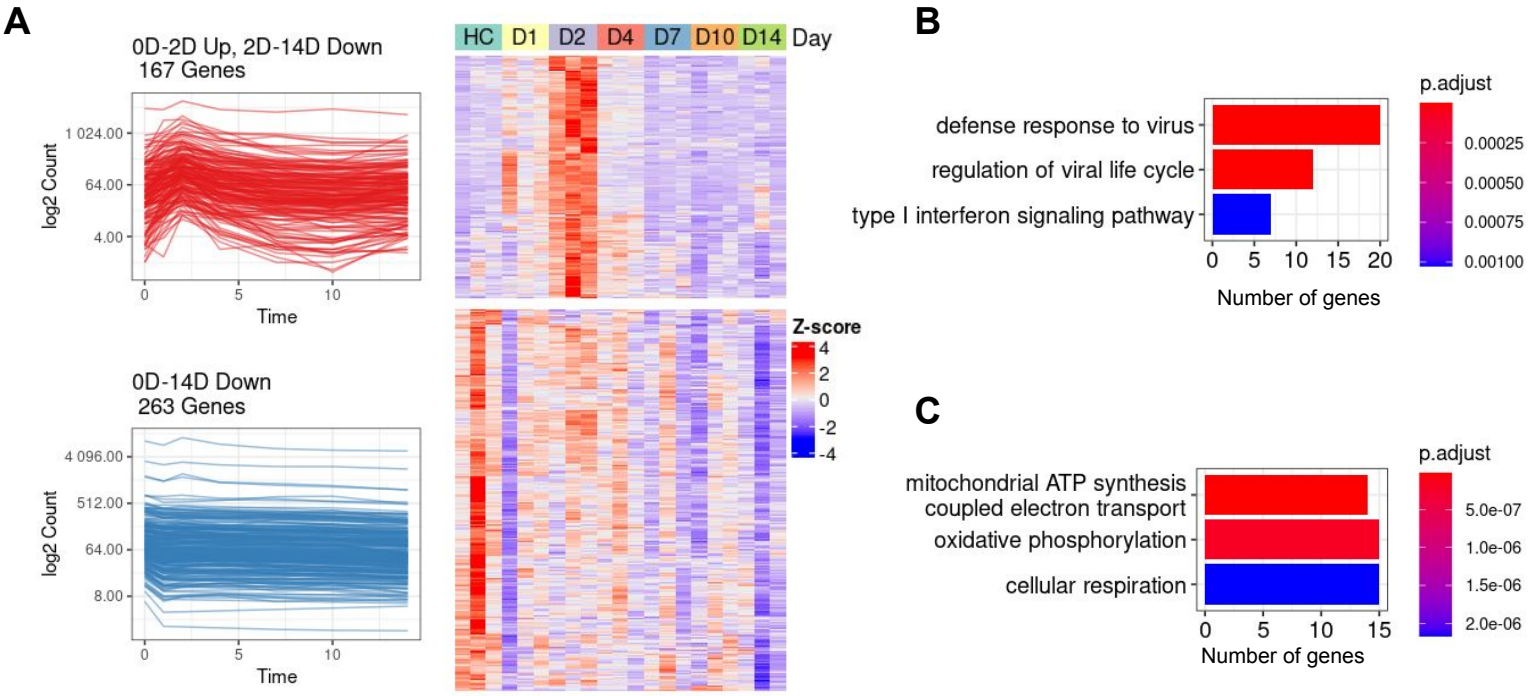


Figure 1. **Overview and benchmarking of of *TrendCatcher***. (A) *TrendCatcher* framework. *TrendCatcher* preprocess input data, which includes creating cell-type specific "pseudobulk" datasets for temporal analysis when single-cell data is used. *TrendCatcher*'s core algorithm is composed of five main steps. *TrendCatcher*'s output includes four main types of visualizations and DDEG identification (numbered 1-5). (B) *TrendCatcher* prediction ROC for a 7 time point simulated dataset compared to DESeq2, DESeq2Spline and ImpulseDE2, with mixed trajectories. (C) *TrendCatcher* prediction performance (AUC) across different number of time points, from 3 to 11 time points. *TrendCatcher*'s AUC values across time points from 3 to 11 are, 0.90, 0.92, 0.90, 0.89 and 0.88.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



D

	0D-1D	1D-2D	2D-4D	4D-7D	7D-10D	10D-14D	%GO	nDDEG
defense response to virus	2.57	0.1	-1.11	-1.18	-0.42	0.69	26.5	69
regulation of innate immune response	1.76	0.25	-1.03	-0.74	-0.32	-0.22	12.4	39
type I interferon signaling pathway	2.89	0.14	-1.33	-1.33	-0.48	1.81	32.6	31
regulation of viral life cycle	2.46	0.07	-1.11	-1.1	-0.34	1.3	27.7	41
regulation of viral genome replication	2.73	0.04	-1.14	-1.18	-0.35	1.42	34.9	29
interferon-gamma-mediated signaling pathway	2.26	0.01	-1.13	-1.06	-0.46	-0.14	23.1	21
response to interferon-beta	1.65	0.06	-0.82	-0.8	-0.26	-0.28	35.5	11
neutrophil mediated immunity	1.2		-0.55	0.56	-0.27	-0.5	9.6	48
regulation of response to cytokine stimulus	1.77	0.78	-1.02	-0.81	-0.34	-0.22	14.8	28
antimicrobial humoral response			0.95	1.86		-0.67	9.9	14
mucosal immune response				1.9			13.5	5
killing of cells of other organism				2.01	0.37	-0.68	13.4	9
mitochondrial ATP synthesis coupled electron transport	-0.26	-0.03	-0.14	-0.11	-0.13	-0.31	15.2	15
protein targeting to ER	-0.54	0.01	0.06	0.28		-0.49	19.2	23

Ave_log2FC: 2, 1, 0, -1, -2

Figure 2. Dynamic gene expression in peripheral blood following SARS-CoV-2 inoculation in a non-human primate model. (A) Analysis of the two predominant trajectory patterns in the non-human primate peripheral blood RNA-seq data from Day 0 to Day 14. The top-left figure represents 167 DDEGs following an up-down expression pattern, which peaked at Day 2 then slowly decreased until Day 14. The top-right figure represents their expression using a traditional z-score normalized heatmap. The bottom-left figure represents 263 DDEGs following a monotonic down-regulated trajectory pattern and their gene expression values were represented in the corresponding heatmap on the right. (B) Top 3 Gene ontology (GO) enrichment analysis pathways using 167 DDEGs from trajectory pattern “0D-2D Up, 2D-14D Down”. (C) Top 3 Gene ontology (GO) enrichment analysis pathways using 263 DDEGs from trajectory pattern “0D-14D Down”. (D) *TimeHeatmap* of the top 15 dynamic pathways and their dynamic time windows visualizes the temporal patterns. Each column represents a time window. The “%GO” column represents the percentage of DDEGs found in the corresponding pathway. The “nDDEG” column represents number of DDEGs found in the corresponding pathway. Color represents the averaged log fold-change of the DDEGs within each time window for the corresponding pathway.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

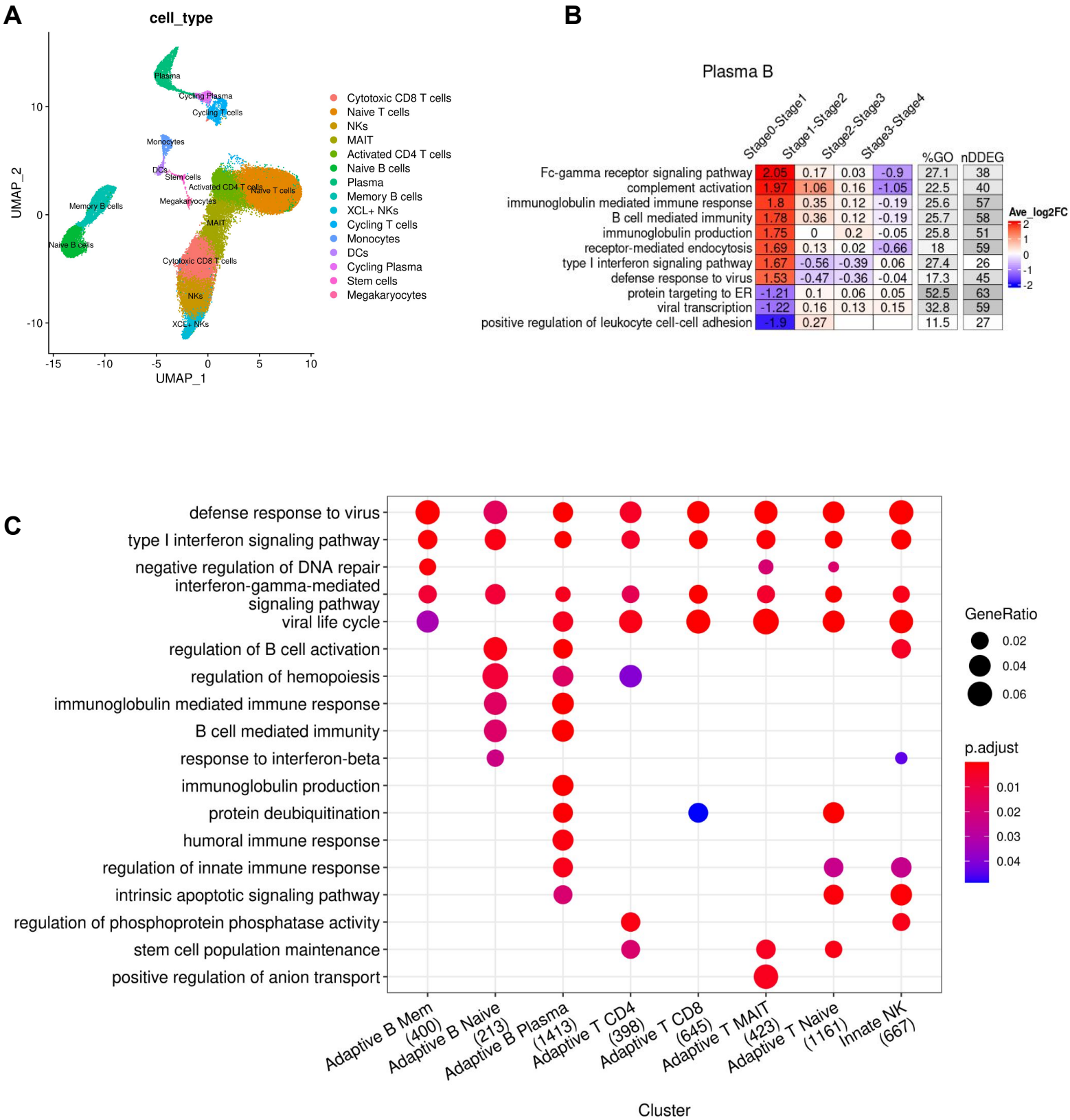
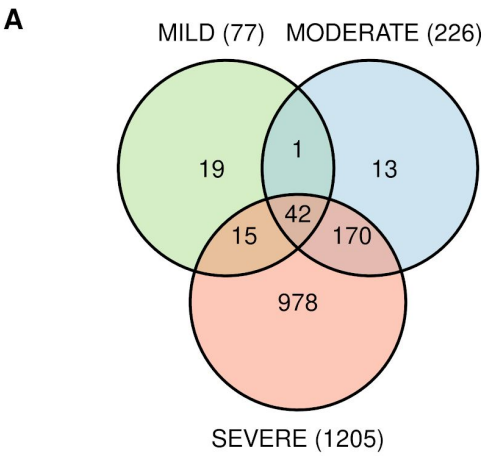
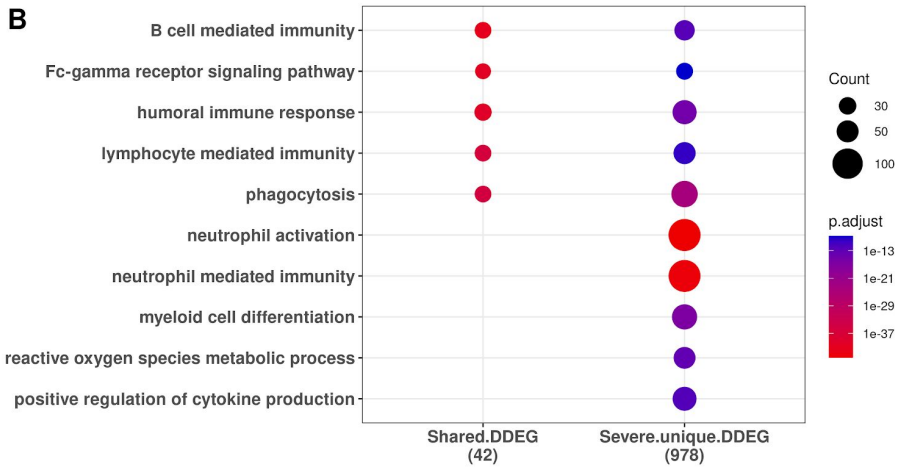


Figure 3. Cell-type specific dynamic gene expression in peripheral blood mononuclear cells following SARS-CoV-2 infection in patients. (A) UMAP visualization of single cell RNA-seq PBMC dataset (Zhu et al.) with annotated cell types from the original study. (B) *TimeHeatmap* of top dynamic biological pathway from plasma B cells. Each column represents a time window (stage). Stage 0 represents uninfected baseline. Each row represents the most dynamic biological pathways. Color represents the averaged log fold-change of the DDEGs within each time window from the corresponding pathway. (C) Top GO enrichment comparison analysis using DDEGs from each cell type. Dot size represents gene ratio. Dot color represents GO enrichment p-value. The number in the x-axis labels represent the number of DDEGs for each cell type.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.





C

Severe COVID-19

	0W-1W	1W-2W	2W-3W	3W-4W	4W-5W	5W-6W	6W-7W	%GO	nDDEG
Fc-gamma receptor signaling pathway	2.18	0.57	-0.26	-0.31	-0.18	-0.53	-0.39	43.6	61
type I interferon signaling pathway	2.02	0.35	-0.96	-1.2	-0.27	-0.29	-0.44	29.5	28
killing of cells of other organism	2	0.46	0.43	0.42	-0.95	-0.54	-0.49	23.9	16
B cell mediated immunity	1.99	0.57	-0.26	-0.35	-0.15	-0.55	-0.39	38.1	86
response to virus	1.99	0.06	-0.86	-1.03	-0.32	-0.34	-0.43	14.8	53
humoral immune response	1.98	0.52	-0.07	-0.19	-0.27	-0.56	-0.41	28.9	110
lymphocyte mediated immunity	1.96	0.54	-0.25	-0.34	-0.16	-0.53	-0.4	26.1	94
phagocytosis	1.8	0.42	-0.11	-0.18	-0.21	-0.41	-0.27	27.3	104
gas transport	1.46	0.31	0.43	0.46	0.27			25	5
respiratory burst	1.34	0.26	0.16	0.79	-0.56	-0.74	-0.7	38.9	14
regulation of response to cytokine stimulus	1.33	0.05	-0.69	-0.63	-0.29	-0.44	-0.22	14.8	28
neutrophil activation	1.32	0.32	0.28	0.02	-0.44	-0.3	-0.21	23.4	117
neutrophil mediated immunity	1.31	0.32	0.28	0.02	-0.44	-0.31	-0.21	23.2	116
reactive oxygen species metabolic process	1.23	0.28	0.32	0.16	-0.1	-0.27	-0.56	12.8	36
blood coagulation	1.22	0.24	0.22	0.29	0.19	0.01	0.3	10.5	36
myeloid cell differentiation	1.13	0.31	0.14	0.02	-0.1	-0.19	-0.4	14.6	61
autophagy of mitochondrion	0.7	0.21	0.24	0.32	0.13	0.05	0.07	10.7	8
protein targeting to ER	-0.74	-0.47		0.73	0.22	0.21	0.05	5	6
positive regulation of cell adhesion	-1.1	0.33	0.28		-0.53	0.16	0.27	4.5	19
T cell differentiation	-1.12	-0.3	-0.36	-0.43	-0.35	0.12	0.3	6.9	17
regulation of lymphocyte differentiation	-1.14							5.7	10
regulation of T cell activation	-1.16	0.29		-0.51	-0.33	-0.11	-0.12	4.9	16

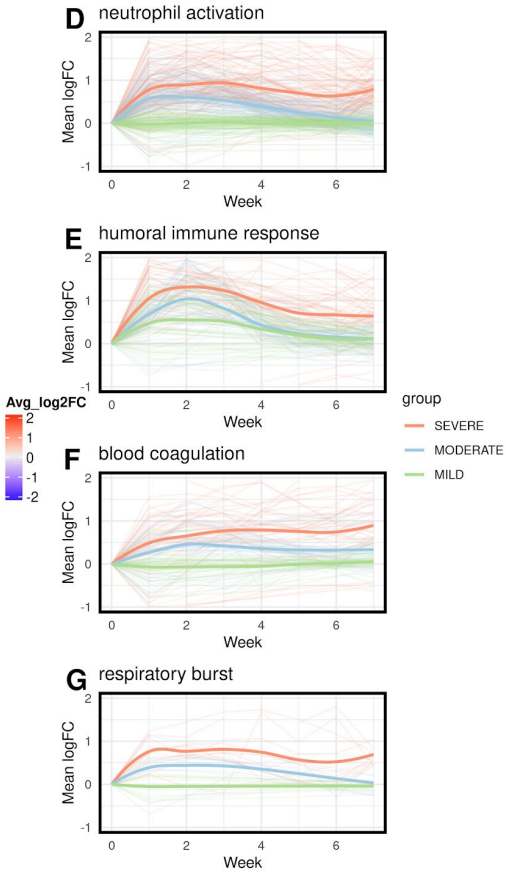
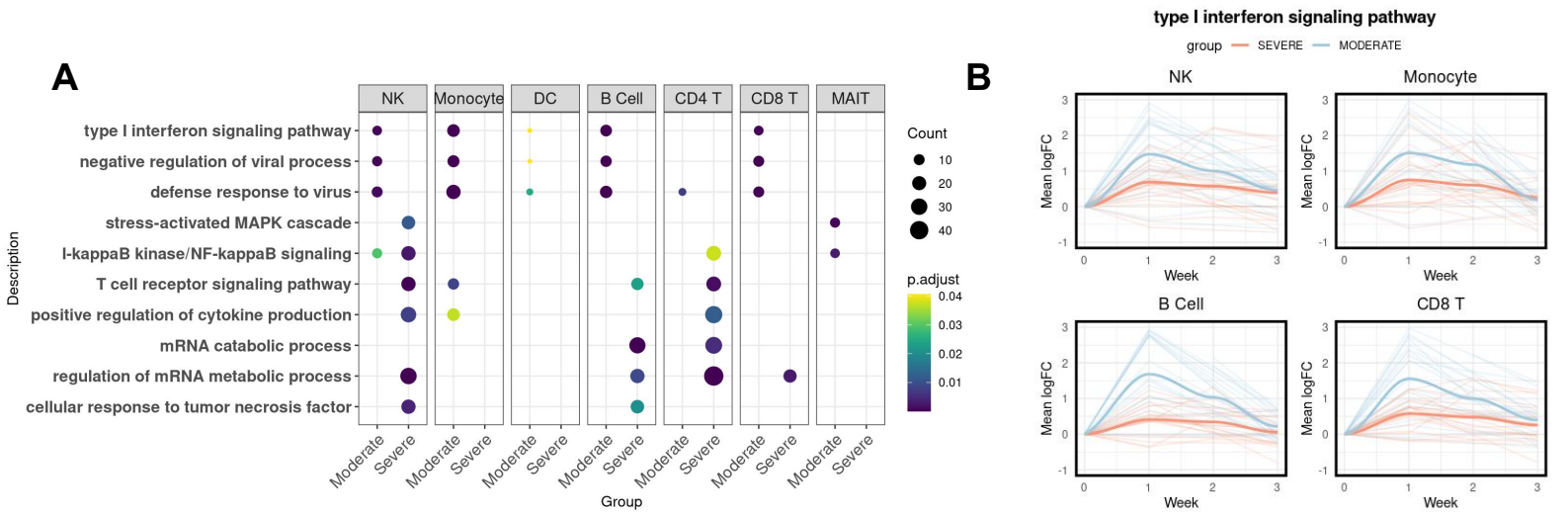


Figure 4. Temporal analysis of whole blood RNA-seq data in patients grouped according to disease severity. (A) Venn diagram of DDEGs identified from three COVID-19 severity groups. (B) Top GO enrichment from shared DDEGs across three groups compared with top GO enrichment from severe group. Dot size represents number of DDEGs enriched from the corresponding biological pathway. (C) *TimeHeatmap* of the top dynamic pathways from the severe group. Each column represents a time window, with W depicting week. (D-E) LOESS curve fitting of DDEGs identified in the severe COVID-19 group of the (D) neutrophil activation pathway, (E) humoral immune response pathway, (F) blood coagulation pathway and (G) respiratory burst pathway. Red curves represent the severe group, blue curves represent the moderate group, and green curves represent the mild group.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



C

Moderate NK

	0W-1W	1W-2W	2W-3W	%GO	nDDEG
type I interferon signaling pathway	1.62	-0.48	-0.53	23.2	22
negative regulation of viral process	1.44	-0.53	-0.55	19.1	17
defense response to virus	1.4	-0.45	-0.46	13.5	35
intrinsic apoptotic signaling pathway	0.68	0.09	0	11.3	32
I-kappaB kinase/NF-kappaB signaling	0.56	0.27	0.13	9.6	27
RNA splicing	0.49	0.13	0.12	10.8	52
response to unfolded protein	0.49	0.15	0.11	8.6	16
regulation of mRNA metabolic process	0.45	0.14	0.14	9.3	31

Ave_log2F: 2, 1, 0, -1, -2

D

Severe NK

	0W-1W	1W-2W	2W-3W	%GO	nDDEG
type I interferon signaling pathway	0.87	-0.31	-0.22	18.9	18
response to virus	0.7	0.06	-0.03	14.8	53
negative regulation of lymphocyte activation	0.65	-0.25	-0.25	14.4	22
regulation of apoptotic signaling pathway	0.61	-0.29	-0.23	11.8	41
I-kappaB kinase/NF-kappaB signaling	0.48	0.1	0.06	15.6	44
RNA splicing	0.37	0.2	0.15	14.8	71
leukocyte cell-cell adhesion	-0.65	-0.31	-0.27	7.9	29
negative regulation of cell activation		-0.23	-0.24	9.5	19

Ave_log2FC: 2, 1, 0, -1, -2

E

Moderate CD8 T

	0W-1W	1W-2W	2W-3W	%GO	nDDEG
type I interferon signaling pathway	1.97	-0.62	-0.63	13.7	13
negative regulation of viral genome replication	1.91	-0.77	-0.79	16.7	9
response to interferon-alpha	1.83	-0.55	-0.53	23.8	5
defense response to virus	1.55	-0.62	-0.61	10.4	27
RNA splicing	0.46	0.14	0.07	13.1	63
proteasomal protein catabolic process	0.46	0.14	0.12	5.6	27
intrinsic apoptotic signaling pathway	0.26	0.13	-0.02	9.9	28
signal transduction by p53 class mediator	-0.53	0.14	0.14	7.6	20

Ave_log2FC: 2, 1, 0, -1, -2

F

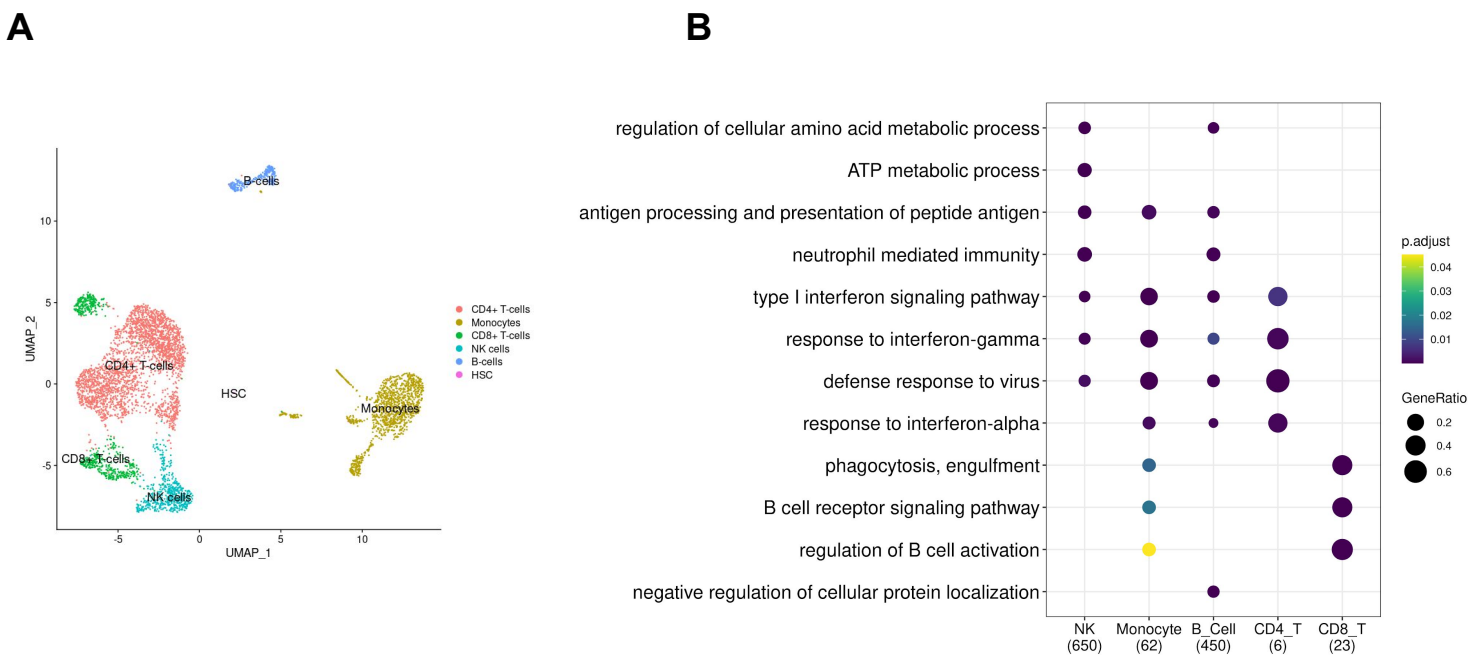
Severe CD8 T

	0W-1W	1W-2W	2W-3W	%GO	nDDEG
regulation of alpha-beta T cell activation	0.89	-0.19	-0.18	12.9	13
G1/S transition of mitotic cell cycle	0.69	0.07	0.06	10.5	29
ribonucleoprotein complex biogenesis	0.46	0.15	0.14	10.6	50
RNA splicing	0.41	0.14	0.14	16.4	79
signal transduction by p53 class mediator	0.16	0.03	0.04	12.2	32

Ave_log2FC: 2, 1, 0, -1, -2

Figure 5. Temporal analysis of single cell RNA-seq data of PBMCs from patients with either moderate or severe COVID-19. (A) Dot plot showing GO enrichment comparison between severe COVID-19 and moderate COVID-19 for each cell type. The size of the dot represents gene count. (B) LOESS curve fitting on DDEGs identified from the type I interferon pathway using TrendCatcher from moderate COVID-19 and severe COVID-19. Blue color indicates moderate group and red color indicates severe group. (C) *TimeHeatmap* of NK cells from moderate COVID-19. (D) *TimeHeatmap* of NK cells from severe COVID-19. (E) *TimeHeatmap* of $CD8^+$T cells from moderate COVID-19. (F) *TimeHeatmap* of $CD8^+$T cells from severe COVID-19. Each column represents a time window with W depicting week.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



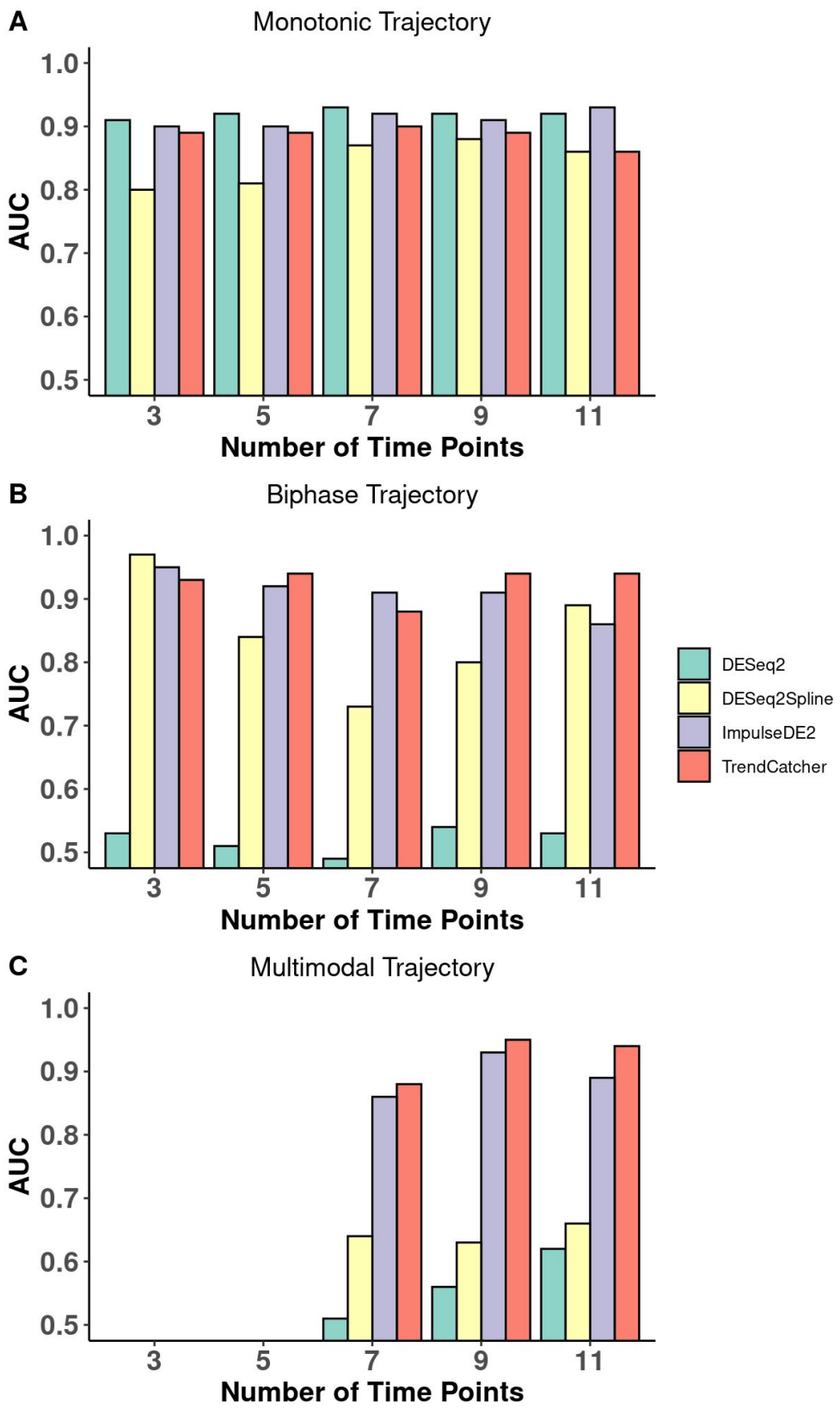
C

	0D-1D	1D-2D	2D-7D	7D-21D	21D-22D	22D-28D	28D-42D	%GO	nDDEG
type I interferon signaling pathway	0.6	0.2	0.05	0.27	0.11	-0.06	-0.54	16.8	16
regulation of cell cycle phase transition	0.59	0	0.02	-0.16	0	-0.03	-0.19	10.5	50
response to interferon-gamma	0.58	0.02	0.08	0.25	0.22	0.14	-0.47	10.2	20
interferon-gamma-mediated signaling pathway	0.55	0.03	0.1	0.31	0.26	0.13	-0.63	13.2	12
Fc receptor signaling pathway	0.49	0	0.02	-0.01	0.01	0.02	-0.38	18	43
regulation of innate immune response	0.47	-0.01	-0.03	0.02	0.05	-0.01	-0.34	14	44
antigen processing and presentation of peptide antigen via MHC class I	0.46	-0.01	-0.02	-0.04	0.01	-0.03	-0.26	34.3	34
regulation of cellular amino acid metabolic process	0.46	-0.01	-0.02	-0.05	0	-0.03	-0.25	43.5	27
ATP metabolic process	0.44	-0.01	0.02	-0.08	-0.01	-0.04	-0.18	17.3	54
oxidative phosphorylation	0.42	0	0.01	-0.09	-0.02	-0.04	-0.17	27.7	41

Ave_log2FC: 1, 0.5, 0, -0.5, -1

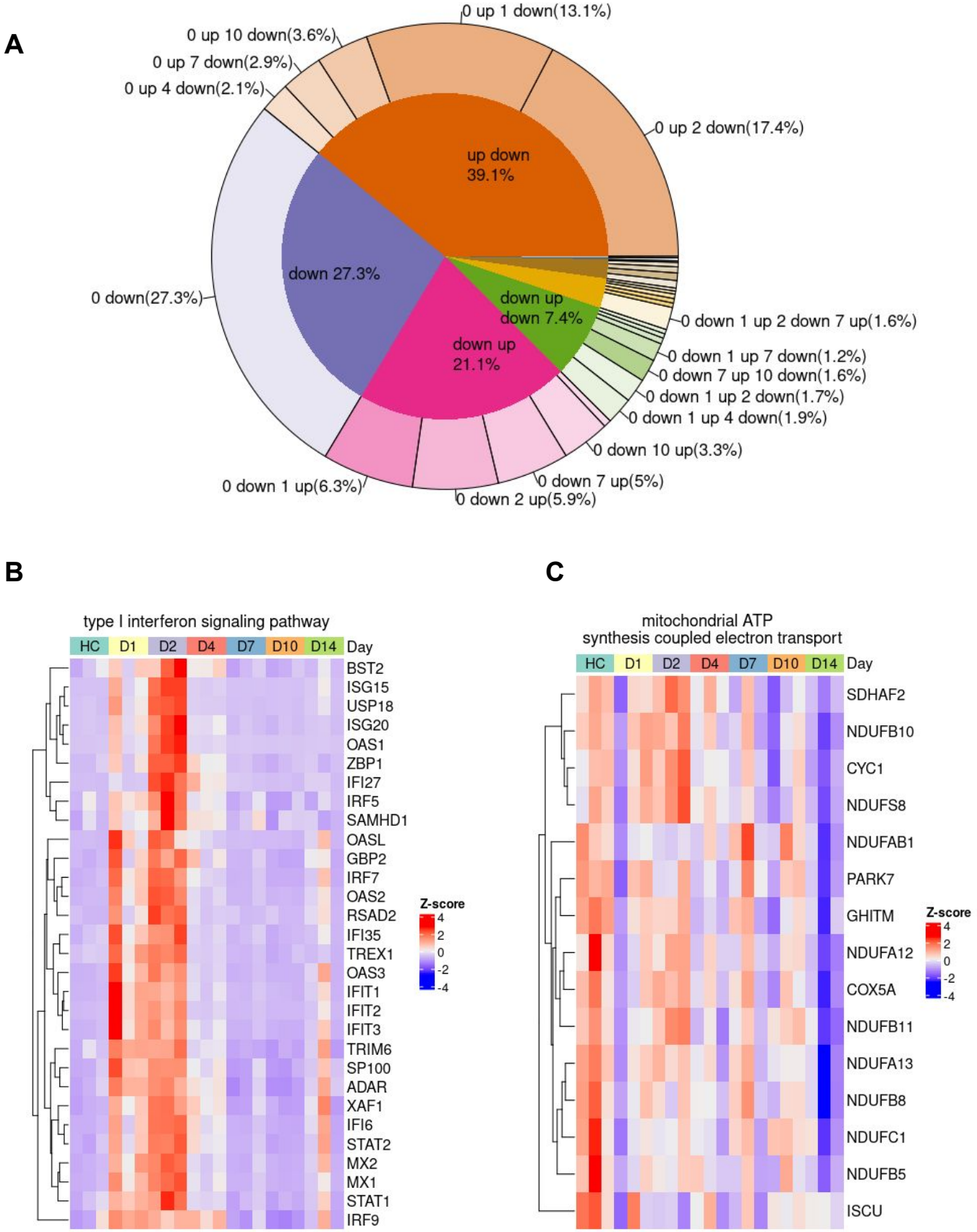
Figure 6. Temporal analysis of PBMC single cell RNA-seq data from human subjects receiving the SARS-CoV-2 mRNA vaccine. (A) UMAP of the single-cell transcriptional profile of patient 2047 at Day 0. Cell types were auto-annotated by SingleR. (B) Dot plot of comparison of the top GO terms enriched from cell-type specific DDEGs. *TrendCatcher* identified 650, 62, 450, 6 and 23 DDEGs from NK cells, monocytes, B cells, $CD4^+$ T cells and $CD8^+$ T cells. Size of the dot represents gene ratio from the enrichment analysis corresponding to each cell type. (C) T*imeHeatmap* of NK cells. Each column represents a time window measured by days. The second dose was administered at Day 21.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



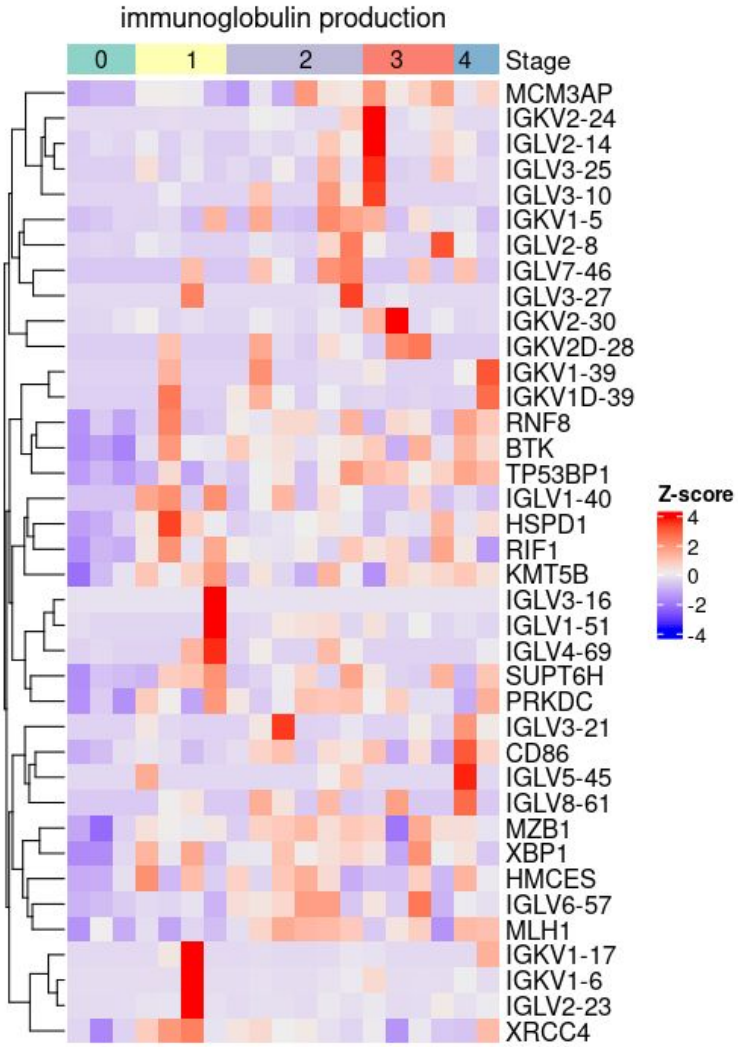
Supplementary Figure 1, related to Figure 1. TrendCatcher prediction performance on different types of trajectories. (A) Prediction performance of *TrendCatcher* across varying numbers of time points for monotonic trajectories. (B) Prediction performance of *TrendCatcher* across varying numbers of time points for biphasic trajectories. (C) Prediction performance of *TrendCatcher* across varying numbers of time points for multimodal trajectories. DESeq2 is shown in green, DESeq2Spline is shown in yellow, ImpulseDE2 is shown in purple, *TrendCatcher* is shown in red.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Supplementary Figure 2, related to Figure 2. Trajectory pattern composition of DDEGS and dynamic gene signatures from highlighted pathways. (A) Hierarchical pie chart shows the composition of trajectory patterns of DDEGs identified in a non-human primate bulk peripheral blood mRNA (Aid et al.). (B) Dynamic gene signatures for SARS-CoV-2 infection from type I interferon signaling pathway. (C) Dynamic gene signatures response for SARS-CoV-2 infection from mitochondrial ATP synthesis coupled electron transport pathway. Color represents normalized z-score for gene expression.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Supplementary Figure 3, related to Figure 3. Dynamic gene signatures of immunoglobulin pathway. Heatmap showing B cell's 38 DDEGs identified from immunoglobulin production process. Color represents normalized z-score of gene expression.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A

Memory B

	Stage0-Stage1	Stage1-Stage2	Stage2-Stage3	Stage3-Stage4	%GO	nDDEG
type I interferon signaling pathway	2.08	-0.63	-0.46	-0.35	10.5	10
defense response to virus	1.8	-0.8	-0.36	-0.32	8.8	23
viral genome replication	1.57	-0.85	-0.45		9.3	12
regulation of chromosome organization	0.92	0.04	0.15	0.27	8.1	22
protein targeting to ER	-1.27	0.23	0.11	-0.02	40.8	49

Ave_log2FC: 2, 1, 0, -1, -2

B

Naive B

	Stage0-Stage1	Stage1-Stage2	Stage2-Stage3	Stage3-Stage4	%GO	nDDEG
type I interferon signaling pathway	2.55	-0.83	-0.51	-0.27	8.4	8
defense response to virus	2.06	-0.74	-0.48	-0.32	4.2	11
response to interferon-beta	1.67	-0.5	-0.27	-0.35	12.9	4
regulation of B cell activation	0.97	0.39	0.29	-0.5	5.7	11
protein targeting to ER	-1.58	0.34	0.15	-0.02	19.2	23
viral transcription	-1.62	0.37	0.16	-0.06	13.9	25

Ave_log2FC: 2, 1, 0, -1, -2

C

NK

	Stage0-Stage1	Stage1-Stage2	Stage2-Stage3	Stage3-Stage4	%GO	nDDEG
negative regulation of viral genome replication	2.04	-0.85	-0.34	0.24	18.5	10
type I interferon signaling pathway	1.91	-0.6	-0.33	-0.03	21.1	20
response to interferon-alpha	1.7	-0.53	-0.29	-0.3	33.3	7
defense response to virus	1.69	-0.59	-0.28	0.02	14.6	38
interferon-gamma-mediated signaling pathway	1.35	-0.31	-0.12	-0.2	13.2	12
regulation of chromatin organization	1.05	-0.11	0.04	0.18	14.1	26
histone modification	1.04	-0.24	-0.1	0.05	8.5	38
RNA splicing	0.92	-0.04	0.02	0.07	11.9	57
protein targeting to ER	-1.17	0.16	0.11	0.12	34.2	41
viral transcription	-1.24	0.17	0.1	0.08	28.9	52

Ave_log2FC: 2, 1, 0, -1, -2

Supplementary Figure 4, related to Figure 3. TimeHeatmap of B cells and NK cells. (A) *TimeHeatmap* of Memory B cells. (B) *TimeHeatmap* of Naive B cells. (C) *TimeHeatmap* of NK cells.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A

Naive T

	Stage0-Stage1	Stage1-Stage2	Stage2-Stage3	Stage3-Stage4	%GO	nDDEG
type I interferon signaling pathway	1.57	-0.65	-0.35	-0.19	24.2	23
defense response to virus	1.35	-0.59	-0.24	-0.15	17.7	46
regulation of chromatin organization	0.93	0.01	-0.02	0.15	21.6	40
histone modification	0.87	-0.02	-0.03	0.19	15.4	69
RNA splicing	0.79	0	0	0.1	21.6	104
protein targeting to ER	-1.28	0.23	0.13	0.03	53.3	64
viral transcription	-1.33	0.22	0.13	0.01	42.8	77

Ave_log2FC: 2, 1, 0, -1, -2

B

CD4 T

	Stage0-Stage1	Stage1-Stage2	Stage2-Stage3	Stage3-Stage4	%GO	nDDEG
type I interferon signaling pathway	2.33	-0.73	-0.48	-0.36	9.5	9
defense response to virus	1.82	-0.75	-0.42	-0.27	6.2	16
viral genome replication	1.52	-0.55	-0.36	-0.33	7	9
regulation of chromosome organization	0.94	0	0	0.08	9.2	25
histone modification	0.89	-0.19	-0.19	-0.17	5.4	24
protein targeting to ER	-1.31	0.22	0.14	0.12	39.2	47
viral transcription	-1.34	0.22	0.14	0.08	30	54

Ave_log2FC: 2, 1, 0, -1, -2

C

CD8 T

	Stage0-Stage1	Stage1-Stage2	Stage2-Stage3	Stage3-Stage4	%GO	nDDEG
type I interferon signaling pathway	2.06	-0.7	-0.49	-0.31	16.8	16
regulation of viral life cycle	1.85	-0.82	-0.45	-0.4	9.5	14
viral genome replication	1.73	-0.6	-0.34	-0.28	12.4	16
defense response to virus	1.68	-0.6	-0.34	-0.04	11.2	29
regulation of chromosome organization	0.89	0.05	0.03	0.21	11.7	32
interferon-gamma-mediated signaling pathway	-0.02	-0.41	0.15	-0.37	17.6	16
protein targeting to ER	-1.23	0.17	0.13	0.08	41.7	50
T cell activation	-1.62	0.25	0.21	-0.52	6.5	31

Ave_log2FC: 2, 1, 0, -1, -2

D

MAIT

	Stage0-Stage1	Stage1-Stage2	Stage2-Stage3	Stage3-Stage4	%GO	nDDEG
type I interferon signaling pathway	2.3	-0.86	-0.53	-0.23	11.6	11
defense response to virus	1.79	-0.88	-0.38	-0.26	7.7	20
chromatin remodeling	1	-0.05	-0.02	0.09	9.6	20
protein targeting to ER	-1.25	0.16	0.12	0.13	20.8	25
viral transcription	-1.33	0.18	0.14	0.15	16.1	29

Ave_log2FC: 2, 1, 0, -1, -2

Supplementary Figure 5, related to Figure 3. TimeHeatmap of T cells. (A) *TimeHeatmap* of Naive T cells. (B) *TimeHeatmap* of CD4$^+$ T cells. (C) *TimeHeatmap* of CD8$^+$ T cells. (D) *TimeHeatmap* of MAIT cells.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A

Mild COVID-19

	0W-1W	1W-2W	2W-3W	3W-4W	4W-5W	5W-6W	6W-7W	%GO	nDDEG
Fc-gamma receptor signaling pathway	1.71	0.49	-0.03	-0.74	-0.6	-0.31	-0.25	18.6	26
B cell mediated immunity	1.68	0.6	-0.06	-0.73	-0.59	-0.3	-0.25	13.3	30
lymphocyte mediated immunity	1.68	0.6	-0.06	-0.73	-0.59	-0.3	-0.25	8.3	30
phagocytosis	1.68	0.6	-0.21	-0.65	-0.59	-0.3	-0.25	7.9	30
humoral immune response	1.63	0.56	0.25	-0.85	-0.58	-0.29	-0.24	8.4	32
protein targeting to ER			0.73	-1.82	1		0.53	4.2	5

Avg_log2FC: 2, 1, 0, -1, -2

B

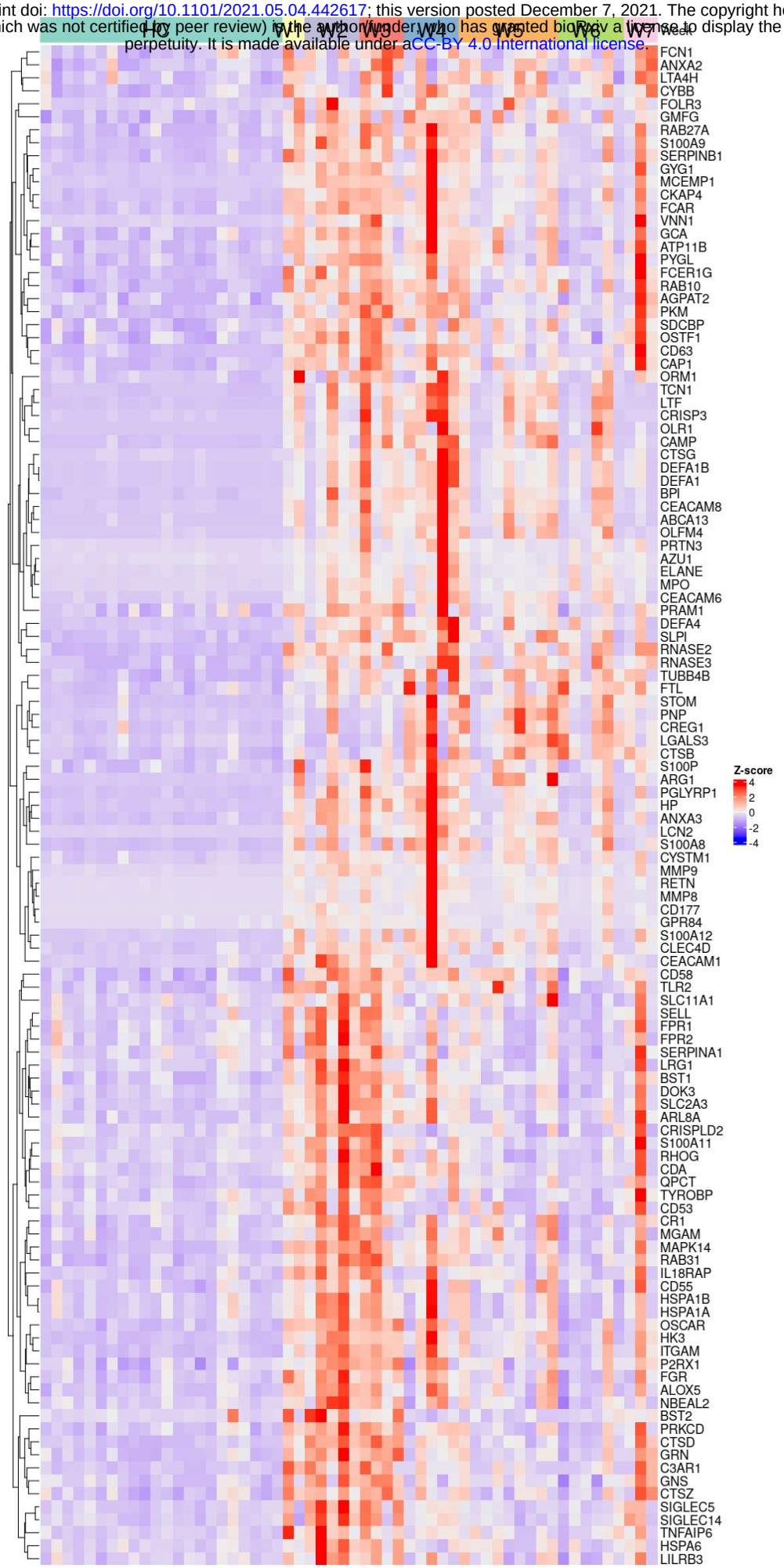
Moderate COVID-19

	0W-1W	1W-2W	2W-3W	3W-4W	4W-5W	5W-6W	6W-7W	%GO	nDDEG
type I interferon signaling pathway	2.34	-0.63	-0.57	-0.52	-0.47	-0.42	-0.4	26.3	25
humoral immune response	1.53	1.14	-0.87	-0.99	-0.51	-0.27	-0.28	15.5	59
lymphocyte mediated immunity	1.48	1.29	-0.86	-1.06	-0.54	-0.24	-0.22	13.9	50
B cell mediated immunity	1.47	1.32	-0.87	-1.09	-0.54	-0.23	-0.22	20.8	47
Fc-gamma receptor signaling pathway	1.43	1.42	-0.89	-1.12	-0.54	-0.21	-0.2	28.6	40
phagocytosis	1.39	1.23	-0.87	-1.04	-0.52	-0.23	-0.2	13.4	51

Avg_log2FC: 4, 2, 0, -2, -4

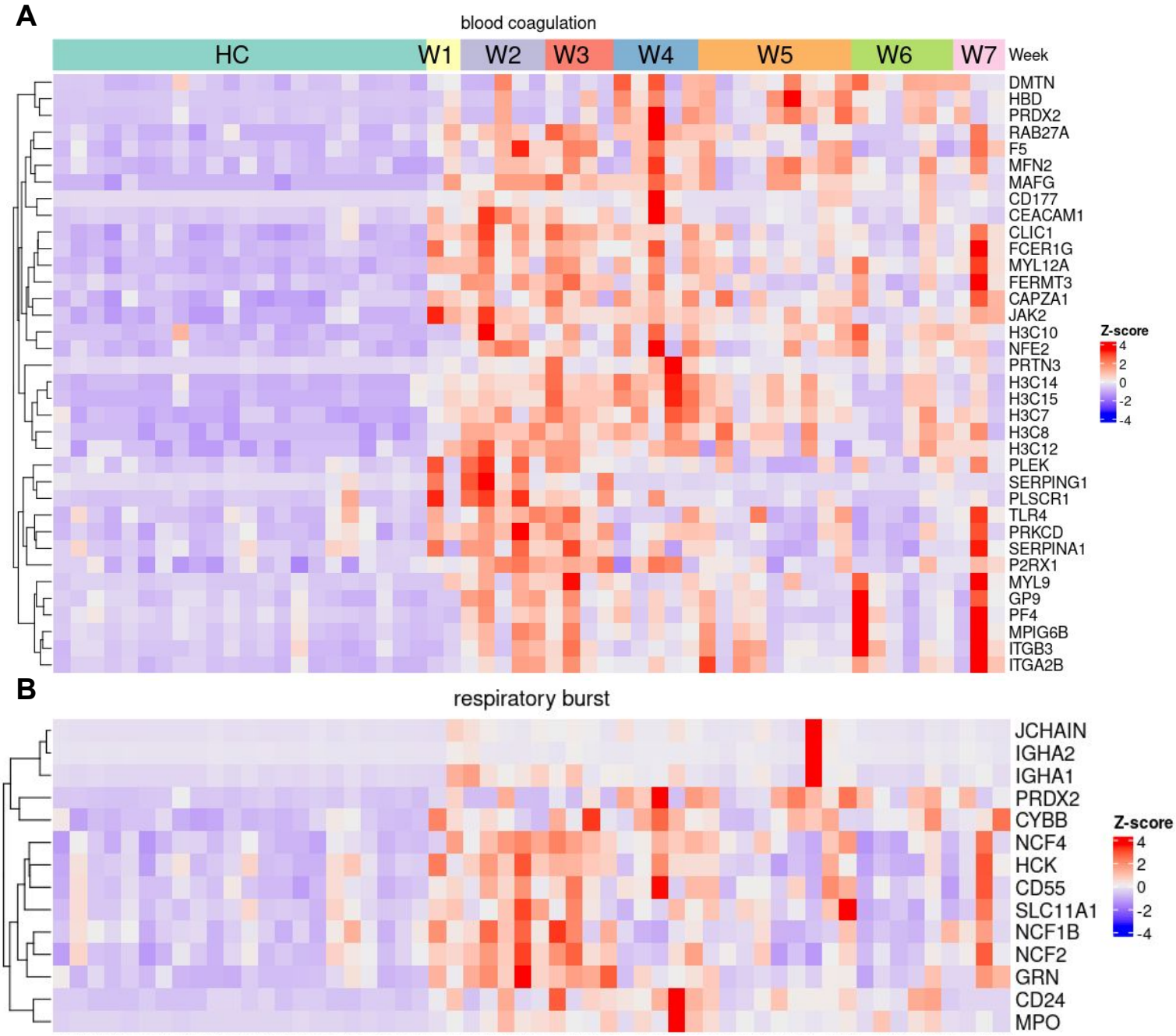
Supplementary Figure 6, related to Figure 4. TimeHeatmap of mild and moderate COVID-19. (A) *TimeHeatmap* of top dynamic GO terms found in severe COVID-19 shown in mild group. (B) *TimeHeatmap* of top dynamic GO terms found in severe COVID-19 shown in moderate group.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Supplementary Figure 7, related to Figure 4. Dynamic gene signatures of neutrophil activation from severe COVID-19. Heatmap of severe COVID-19 DDEGs identified from the neutrophil activation pathway. The heatmap is ordered by week following infection with healthy controls (HC) on the left. Color represents Z-score normalized expression values. Each column represents one sample.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Supplementary Figure 8, related to Figure 4. Dynamic gene signatures of blood coagulation and respiratory burst pathways from severe COVID-19. (A) Severe COVID-19 DDEGs identified from the blood coagulation pathway. (B) Severe COVID-19 DDEGs identified from respiratory burst pathway. The heatmap is ordered by week following infection with healthy controls (HC) on the left. Color represents Z-score normalized expression values.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.04.442617; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A

Monocytes

	0D-1D	1D-2D	2D-7D	7D-21D	21D-22D	22D-28D	28D-42D	%GO	nDDEG
humoral immune response	2.21	-0.19	-0.29	-0.83	0.93	-0.39	-1.57	2.6	10
regulation of immune effector process	1.66	0.12	-0.17	0.08	0.48	-0.2	-1.12	2.8	13
defense response to virus	1.27	0.07	-0.03	0.31	1.25	-0.3	-1.47	5.8	15
cellular response to interferon-gamma	1.02	0.03	0.07	0.55	1.42	-0.26	-1.55	7.3	13
type I interferon signaling pathway	1	0.27	-0.01	0.31	0.71	-0.2	-1.16	14.7	14
response to interferon-gamma	0.99	0.14	0.07	0.5	1.42	-0.27	-1.49	7.6	15
regulation of innate immune response	0.98	0.14	0.05	0.61	0.96	-0.13	-1.54	3.2	10
antigen processing and presentation of exogenous peptide antigen via MHC class I	0.73	0.09	0.1	0.47	1.11	-0.31	-1.29	7.5	6

Ave_log2FC: 2, 1, 0, -1, -2

B

B cells

	0D-1D	1D-2D	2D-7D	7D-21D	21D-22D	22D-28D	28D-42D	%GO	nDDEG
type I interferon signaling pathway	1.16	-0.19	-0.48	0.15	0.31	-0.2	-0.66	18.9	18
negative regulation of cellular protein localization	0.97	-0.08	-0.16	-0.19	-0.01	-0.09	-0.38	14.3	16
entry into host	1.52	-0.37	-0.22	-0.24	-0.01	-0.14	-0.2	11.1	17
antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	0.82	-0.21	-0.11	-0.21	0.11	-0.15	-0.25	16	12
response to endoplasmic reticulum stress	0.92	-0.24	-0.12	-0.14	-0.02	-0.11	-0.19	8.1	24
regulation of cellular amino acid metabolic process	0.79	-0.16	-0.1	-0.18	-0.01	-0.13	-0.2	17.7	11
establishment of planar polarity	0.83	-0.12	-0.1	-0.2	-0.01	-0.12	-0.19	12.2	15
Wnt signaling pathway, planar cell polarity pathway	0.85	-0.13	-0.1	-0.21	-0.01	-0.12	-0.19	13	14
hematopoietic progenitor cell differentiation	0.89	-0.11	-0.18	-0.23	-0.01	-0.12	-0.2	10.6	17
interferon-gamma-mediated signaling pathway	1.07			1.07	0.89	-0.18	-1.09	11	10

Ave_log2FC: 2, 1, 0, -1, -2

C

CD8$^+$ T cells

	0D-1D	1D-2D	2D-7D	7D-21D	21D-22D	22D-28D	28D-42D	%GO	nDDEG
type I interferon signaling pathway	1.1			0.4		2.05		2.1	2
defense response to virus	0.92			0.61	0.29	2.05		1.5	4
response to interferon-gamma	0.65			0.72	0.08			1.5	3
regulation of B cell activation		-1.64	-0.76	-1.56	-0.14	-0.72	-1.56	5.2	10
membrane invagination		-1.41	-0.76	-1.7	-0.14	-0.72	-1.83	6.3	9

Ave_log2FC: 2, 1, 0, -1, -2

Supplementary Figure 9, related to Figure 6. TimeHeatmap of monocytes, B cells and CD8$^+$ T cells(A) *TimeHeatmap* of monocytes. (B) *TimeHeatmap* of B cells. (C) *TimeHeatmap* of CD8$^+$T cells. Day 1 is the first dose of vaccination, Day 21 is the second dose of vaccination.