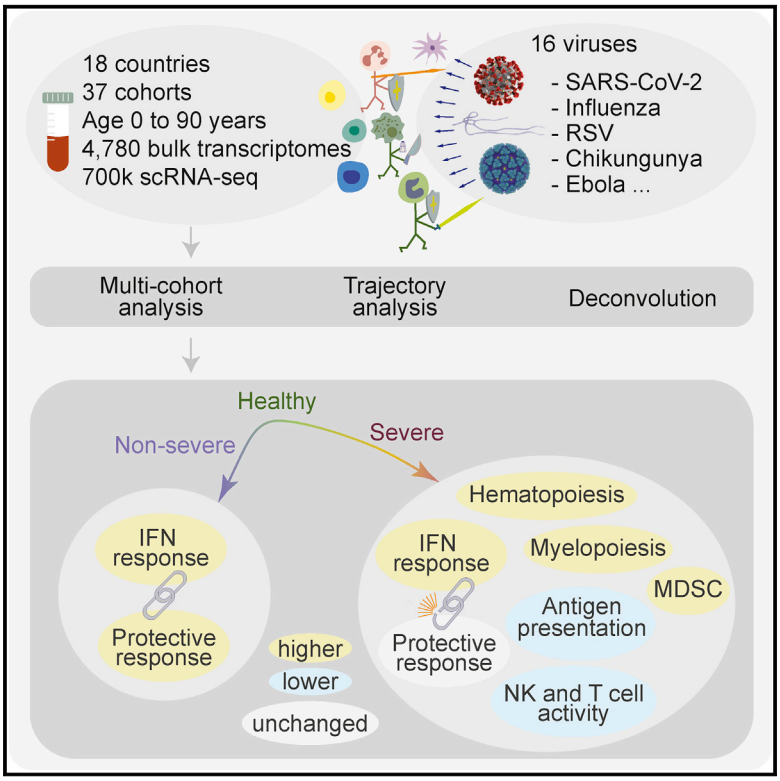
Article

Immunity

Multi-cohort analysis of host immune response identifies conserved protective and detrimental modules associated with severity across viruses

Graphical abstract



Authors

Hong Zheng, Aditya M. Rao, Denis Dermadi, ..., Evangelos J. Giamarellos-Bourboulis, James R. Heath, Purvesh Khatri

Correspondence

pkhatri@stanford.edu

In brief

Viral infections induce a conserved host response distinct from bacterial infections, but whether this conserved response distinguishes severity is unclear. Zheng et al. analyzed >5000 bulk and single-cell transcriptome profiles from patients infected with one of 16 viruses, including SARS-CoV-2, Ebola, and chikungunya. They identified protective and detrimental host response modules that distinguish patients with mild or severe outcomes.

Highlights

- Conserved host response distinguishes non-severe and severe viral infections
- Increased hematopoiesis, myelopoiesis, and myeloid-derived suppression with severity
- Protective and detrimental modules define trajectories for mild and severe outcomes
- Interferon response is decoupled from the protective response in severe outcomes

Zheng et al., 2021, Immunity *54*, 753–768
April 13, 2021
https://doi.org/10.1016/j.immuni.2021.03.002

CellPress

Immunity

CellPress
OPEN ACCESS

Article

Multi-cohort analysis of host immune response identifies conserved protective and detrimental modules associated with severity across viruses

Hong Zheng,[1,2,10] Aditya M. Rao,[1,3,10] Denis Dermadi,[1,2,10] Jiaying Toh,[1,3,10] Lara Murphy Jones,[1,2,4,10] Michele Donato,[1,2,10] Yiran Liu,[1,5] Yapeng Su,[6] Cheng L. Dai,[6] Sergey A. Kornilov,[6] Minas Karagiannis,[7] Theodoros Marantos,[7] Yehudit Hasin-Brumshtein,[8] Yudong D. He,[8] Evangelos J. Giamarellos-Bourboulis,[7] James R. Heath,[6,9] and Purvesh Khatri[1,2,11,*]

[1]Institute for Immunity, Transplantation and Infection, School of Medicine, Stanford University, CA 94305, USA
[2]Center for Biomedical Informatics Research, Department of Medicine, School of Medicine, Stanford University, CA 94305, USA
[3]Immunology program, Stanford University, CA 94305, USA
[4]Division of Critical Care Medicine, Department of Pediatrics, School of Medicine, Stanford University, CA 94305, USA
[5]Cancer Biology program, Stanford University, CA 94305, USA
[6]Institute for Systems Biology, Seattle, WA, USA
[7]4th Department of Internal Medicine, National and Kapodistrian University of Athens, Medical School, 124 62 Athens, Greece
[8]Inflammatix, Inc. Burlingame, CA, USA
[9]Department of Bioengineering, University of Washington, Seattle, WA 98195
[10]These authors contributed equally
[11]Lead contact
*Correspondence: pkhatri@stanford.edu
https://doi.org/10.1016/j.immuni.2021.03.002

SUMMARY

Viral infections induce a conserved host response distinct from bacterial infections. We hypothesized that the conserved response is associated with disease severity and is distinct between patients with different outcomes. To test this, we integrated 4,780 blood transcriptome profiles from patients aged 0 to 90 years infected with one of 16 viruses, including SARS-CoV-2, Ebola, chikungunya, and influenza, across 34 cohorts from 18 countries, and single-cell RNA sequencing profiles of 702,970 immune cells from 289 samples across three cohorts. Severe viral infection was associated with increased hematopoiesis, myelopoiesis, and myeloid-derived suppressor cells. We identified protective and detrimental gene modules that defined distinct trajectories associated with mild versus severe outcomes. The interferon response was decoupled from the protective host response in patients with severe outcomes. These findings were consistent, irrespective of age and virus, and provide insights to accelerate the development of diagnostics and host-directed therapies to improve global pandemic preparedness.

INTRODUCTION

Outbreaks of infectious diseases globally have been increasing steadily over the last 40 years (Christiansen, 2018). The first two decades of the 21st century have been marked by seven outbreaks of novel viruses, including severe acute respiratory syndrome coronavirus (SARS-CoV-1), H1N1 influenza, Middle East Respiratory Syndrome Coronavirus (MERS-CoV), chikungunya, Ebola, Zika, and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Four of these outbreaks resulted in pandemics in the last decade (Morens and Fauci, 2020). With each outbreak, a typical approach has been to pursue a pathogen-specific strategy. With novel viruses, when our biological understanding of the causative agent is poor, the acquisition of sufficient knowledge to manage the disease is time-consuming and expensive. Adopting a pathogen-agnostic strategy, such as through the identification of an underlying conserved host response across patient populations, could greatly accelerate the development of diagnostics and therapies to manage future emerging outbreaks. For instance, most approved antiviral drugs are effective against a small number of viruses, and highly susceptible to resistance (Bekerman and Einav, 2015). In contrast, identifying conserved host biology, such as host proteins required by multiple viruses, could be used to develop broad-spectrum antivirals. Similarly, a conserved host response to viral infections could be used to develop diagnostics and prognostics. Several studies have repeatedly demonstrated the utility of the host immune response to pathogens to accurately diagnose the presence and type of infections (Andres-Terre et al., 2015; Sweeney et al., 2015; 2016b; Mayhew et al., 2020). We have previously identified a conserved host response to distinguish bacterial and viral infections (Andres-Terre et al., 2015; Sweeney et al., 2015; 2016b). We have

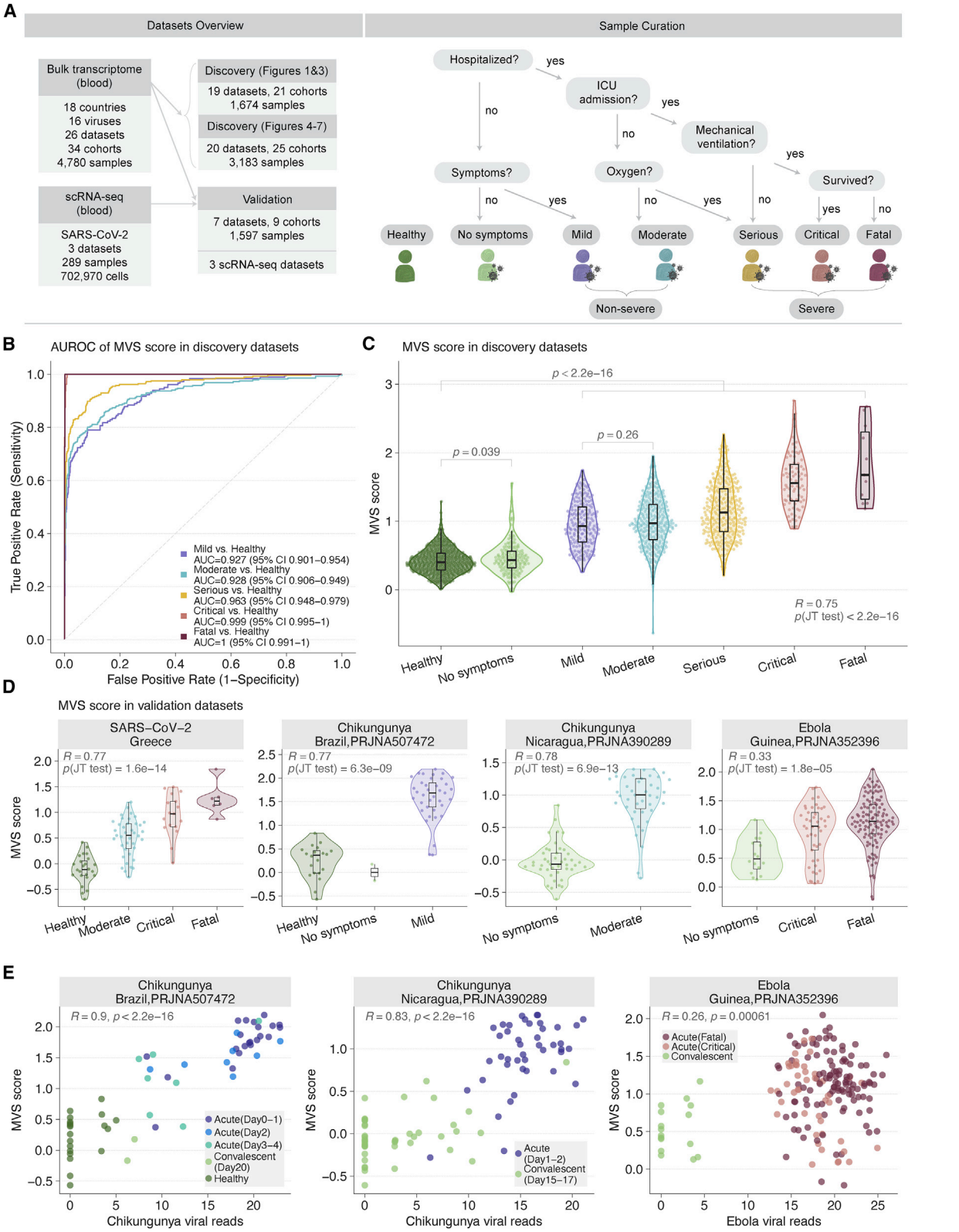


Figure 1. Conserved host response to viral infection, represented by the MVS, is associated with severity

(A) Datasets used for analysis divided into discovery and validation (left) and criteria for assigning viral infection severity categories to samples (right). "No symptoms" includes individuals with asymptomatic viral infection or convalescents.

(legend continued on next page)

also demonstrated that the conserved host immune response to infection is detected earlier than symptom onset (Andres-Terre et al., 2015; Sweeney et al., 2016a; Warsinske et al., 2018; Gupta et al., 2020; Turner et al., 2020).

Here, we hypothesized that our previously described conserved host response signature to respiratory viral infections, called the Meta-Virus Signature (MVS) (Andres-Terre et al., 2015), is also conserved in viral infections that cause severe disease, including Ebola, SARS-CoV-2, and others, and it could be used to identify common genes associated with detrimental and protective host immune responses, irrespective of the virus. We tested these hypotheses by integrating 34 independent cohorts comprising 4,780 blood transcriptome profiles and single-cell RNA-seq profiles of 702,970 immune cells from 289 samples from healthy controls (HCs) and patients with acute viral infection. We found that the MVS is (1) present in SARS-CoV-2, Ebola, chikungunya, influenza, and other viruses, (2) correlated with severity, and (3) predominantly expressed in myeloid cells. Using a patient trajectory differentiation method, we found that patients with mild or severe viral infection follow different trajectories comprised of four gene modules corresponding to protective and detrimental host immune responses. We defined the severe-or-mild (SoM) score that accurately distinguished patients with non-severe and severe outcomes. By leveraging the biological, clinical, and technical heterogeneity across data, we provide strong evidence of a conserved host immune response to acute viral infection, irrespective of the virus. Further analysis of these conserved host response modules could lead to the development of diagnostics, prognostics, and host-directed therapies for a broad spectrum of viruses that could facilitate risk stratification and targeted treatment of patients during the current pandemic and in novel outbreaks that will inevitably arise in the future.

RESULTS

Data collection, curation, and preprocessing

We searched the public repositories for blood transcriptome profiles from patients with viral infection (STAR Methods). After excluding datasets used to discover the MVS previously, we identified 26 datasets composed of 4,780 samples from patients across 18 countries infected with at least one of 16 viruses (Figure 1A, Table S1, and Data S1). Overall, these datasets included a broad spectrum of biological, clinical, and technical heterogeneity represented by blood samples profiled from children and adults infected with a virus using either microarray or RNA sequencing. We assigned a standardized severity category to each of the 4,780 samples (Figure 1A and STAR Methods). Briefly, we divided non-hospitalized samples into "no symptoms" or "mild," and hospitalized patients into "moderate," "serious," "critical," and "fatal" categories based on the level of care required and outcomes as described in the original publications (Figure 1A and STAR Methods). We also defined two broader categories: "non-severe," encompassing patients with mild and moderate viral infection, and "severe," encompassing patients with serious, critical, and fatal viral infection (Figure 1A). For cohorts that lacked sample-level severity data, we assigned the same severity category to each sample based on the cohort description.

MVS represents a conserved host response to viral infection and is associated with severity

To test our hypothesis that a conserved host response to viral infection is associated with severity, we co-normalized 1,674 blood transcriptomes (663 HCs, 167 asymptomatic/convalescent, 181 mild, 286 moderate, 286 serious, 80 critical, and 11 fatal) from 19 independent datasets, the majority of which were infected with adenovirus, influenza, human rhinovirus (HRV), or respiratory syncytial virus (RSV), using COCONUT (Figure 1A, Table S1; STAR Methods) (Sweeney et al., 2016b). The MVS score accurately distinguished patients with viral infection from HCs across all datasets (Figure 1B, Figure S1A) and correlated with severity ($r = 0.75$, $p < 2.2e\text{-}16$; Figure 1C). The MVS score was higher in all infected patients compared to HCs ($p < 2.2e\text{-}16$), regardless of symptoms, severity, and virus (Figure 1C). In asymptomatic or convalescent patients, the MVS score was marginally higher than in HCs ($p = 0.039$), but not different between patients with mild versus moderate severity ($p = 0.26$). Across all datasets, the MVS score was correlated with viral infection severity ($0.43 \leq R \leq 0.93$), regardless of virus, geography, or age (Figure S1B). In 405 samples from patients infected with SARS-CoV-2, Ebola, or chikungunya across 4 datasets, the MVS score was correlated with severity ($0.33 \leq R \leq 0.78$; $p \leq 1.8e\text{-}05$; Figure 1D) and distinguished patients with viral infection from HCs (Figure S1C). In three independent datasets of blood samples from patients with either chikungunya or Ebola infection, profiled using RNA-seq, we detected sequencing reads from the corresponding viral RNA (STAR Methods). In each of these three datasets, the MVS score significantly correlated with the number of viral reads detected in blood ($p \leq 6.1e\text{-}4$; Figure 1E). Further, in each dataset, both the number of viral reads in blood and the MVS score decreased as patients progressed from acute infection to convalescence.

Collectively, our results show that a conserved host response to viral infection, represented by the MVS score, is correlated with the severity and the number of viral reads detected in blood samples, irrespective of biological, clinical, or technical heterogeneity or the infecting virus.

Myeloid cells are the primary source of the MVS

Next, we investigated whether the MVS score is associated with specific immune cell types. We integrated three single-cell

(B) ROC curves for distinguishing patients with viral infection of varying severity from HCs using the MVS score (1,674 samples in 19 datasets).
(C) Distribution of the MVS scores across the severity of viral infection (1,674 samples in 19 datasets). Each point represents a blood sample. Jonckheere-Terpstra (JT) trend test was used to assess the significance of the trend of the MVS score over severity. p values using Mann–Whitney U test for the comparison of MVS scores in two groups.
(D) Validation of correlation between the MVS score severity of viral infection in 4 independent RNA-seq datasets from patients with SARS-CoV-2, chikungunya, or Ebola infection.
(E) Positive correlation between the MVS score and the number of viral reads detected in blood samples RNA-seq datasets. Each point represents a sample. The x axis represents the number of viral reads; the y axis represents the MVS score for each sample. p values were computed using Pearson correlation test. See also Figure S1.

RNA-seq (scRNA-seq) datasets consisting of 702,970 immune cells from 289 PBMC samples (258 SARS-CoV-2, 27 healthy, 2 influenza, 2 RSV) from 170 individuals across three independent cohorts (Seattle, Atlanta, Stanford) (Table S2) (Arunachalam et al., 2020; Su et al., 2020; Wilk et al., 2020). The Seattle Cohort profiled 557,240 immune cells from 258 PBMC samples of HCs and patients with SARS-CoV-2 infection (16 healthy, 94 asymptomatic, 8 mild, 37 moderate, 78 serious, 21 critical, 3 fatal) using CITE-seq. The patients with SARS-CoV-2 infection in the Seattle Cohort were profiled at two time points: (1) near the time of a positive clinical diagnosis and (2) a few days later. The Atlanta Cohort profiled 76,929 immune cells from 18 PBMC samples of HCs and patients infected with one of 3 viruses (5 healthy, 1 moderate influenza, 1 serious influenza, 2 serious RSV, 2 convalescent SARS-CoV-2, 3 moderate SARS-CoV-2, 3 serious SARS-CoV-2, 1 fatal SARS-CoV-2) using CITE-seq. Finally, the Stanford Cohort profiled 68,801 immune cells from 13 PBMC samples of HCs and patients with SARS-CoV-2 infection (6 healthy, 1 moderate, 3 serious, 2 critical, 1 fatal) using Seq-Well. Collectively, these three cohorts included clinical, biological, and technical heterogeneity at a single-cell level.

We integrated the three scRNA-seq cohorts using Seurat (Satija et al., 2015) (Figures 2A–2D, Figures S2A–S2C). Immune cells across the three cohorts clustered into myeloid cells (monocytes, myeloid dendritic cells, granulocytes, etc.), T and NK cells, and B cells (Figures 2A and 2B, Figure S2A). The MVS score was substantially higher in myeloid cells from hospitalized patients with viral infection (Figures 2C–2E, Figures S2B and S2C) and positively correlated with the severity of viral infection in myeloid cells ($R = 0.28$, $p = 2.4e-06$), which was driven by CD14+ monocytes ($R = 0.45$, $p = 2.7e-14$) compared to CD16+ monocytes ($R = 0.25$, $p = 2.4e-05$) (Figure 2F). Further, proportions of myeloid cells increased with severity ($R = 0.46$, $p = 2.0e-15$), which was also driven by CD14+ monocytes ($R = 0.54$, $p < 2.2e-16$). Proportions of CD16+ monocytes decreased with increasing severity of viral infection ($R = -0.31$, $p = 2.2e-07$) (Figure 2G). The MVS score in myeloid cells at the single-cell level and proportions of myeloid cells were positively correlated ($R = 0.34$, $p = 4.8e-09$), which was also driven by CD14+ monocytes ($R = 0.47$, $p < 2.2e-16$) (Figure 2H). Together, these results showed that in response to a viral infection, proportions of and the conserved host response to viral infection at a single-cell level in CD14+ monocytes increase with severity.

Next, we performed *in silico* cellular deconvolution of blood transcriptome profiles of 4357 patient samples from 32 independent cohorts using immunoStates (Bongen et al., 2018; Chowdhury et al., 2018; Scott et al., 2019; Vallania et al., 2018) (Table S3) to investigate if these changes in CD14+ and CD16+ monocytes were also observed at the bulk transcriptome level. We performed three multi-cohort analyses to compare changes in estimated immune cell proportions in (1) non-severe viral infections compared to HCs, (2) severe viral infections compared to HCs, and (3) severe compared to non-severe viral infections.

Similar to scRNA-seq analysis, proportions of total monocytes were significantly higher in patients with non-severe viral infection compared to HCs (ES = 1.10, FDR = 4.33e-13), but not in those with severe viral infection (Figures 2I–2J, Table S4). The proportion of CD14+ monocytes increased significantly in patients with non-severe (ES = 1.12, FDR = 1.30e-21) and severe (ES = 0.9, FDR = 6.56e-10) viral infection compared to HCs, but were not different between patients with non-severe or severe viral infection (Figures 2I–2K, Table S4). In line with the scRNA-seq data, proportions of CD16+ monocytes were significantly lower in patients with severe viral infection compared to HCs (ES = −1.16, FDR = 5.13e-08) and those with non-severe viral infection (ES = −0.88, FDR = 1.73e-17) (Figures 2J–2K, Table S4), but were unchanged in non-severe patients compared to HCs (Figure 2I, Table S4).

Cellular deconvolution analysis also found that the proportions of neutrophils were significantly higher in patients with severe viral infection compared to HCs (ES = 1.24, FDR = 4.12e-16) and those with non-severe viral infection (ES = 0.99, FDR = 4.33e-07) (Figure S2D, Table S4). These cells are likely low-density immature granulocytes typically found in patients with sepsis.

Collectively, our integrated analyses of 4646 samples across 35 independent cohorts using scRNA-seq and *in silico* deconvolution of bulk transcriptome profiles showed that the conserved host response to viral infections is predominantly from myeloid cells, where proportions of CD14+ monocytes increased and CD16+ monocytes decreased with increased severity of viral infection.

MVS identifies distinct clusters of patients with non-severe and severe viral infection

As expected, low dimensional visualization of 1674 co-normalized samples using UMAP showed that HCs were distinct from patients with viral infection irrespective of the infecting virus (Figure 3A). The MVS score increased along the first UMAP component (UMAP1; Figure 3B), while patients with mild viral infection were clustered separately from those with severe viral infection along the second UMAP component (UMAP2; Figure 3C). We further validated the robustness of the localization of samples by severity observed in UMAP by mapping 8 independent cohorts consisting of 2,604 samples from patients with one of 4 viral infections to the same low dimensional space (Figures 3D and 3E). All but one of these 8 cohorts were challenge studies, where 129 healthy individuals were inoculated with influenza (1,465 samples from 70 subjects), RSV (419 samples from 20 subjects), or HRV (634 samples from 39 subjects). Each of the infected subjects in these challenge studies had asymptomatic or mild infection. When mapped to the UMAP space created using the 1,674 samples, samples from the challenge studies clustered with mild viral infections (Figure 3D). In contrast, patients with critical and fatal SARS-CoV-2 infection mapped to the region enriched for patients with critical and fatal viral infection, whereas patients with moderate SARS-CoV-2 infection mapped to the region enriched for patients with non-severe viral infection (Figure 3E), again demonstrating that the host response to viral infection is conserved and associated with severity, irrespective of the virus. Together, this observation suggested that a distinct subset of genes in the MVS may be differentially associated with the severity of viral infection.

Hospitalized patients with viral infection follow a different trajectory from non-hospitalized patients with viral infection

Based on the UMAP of samples, we hypothesized that patients with mild and severe viral infection follow different trajectories.

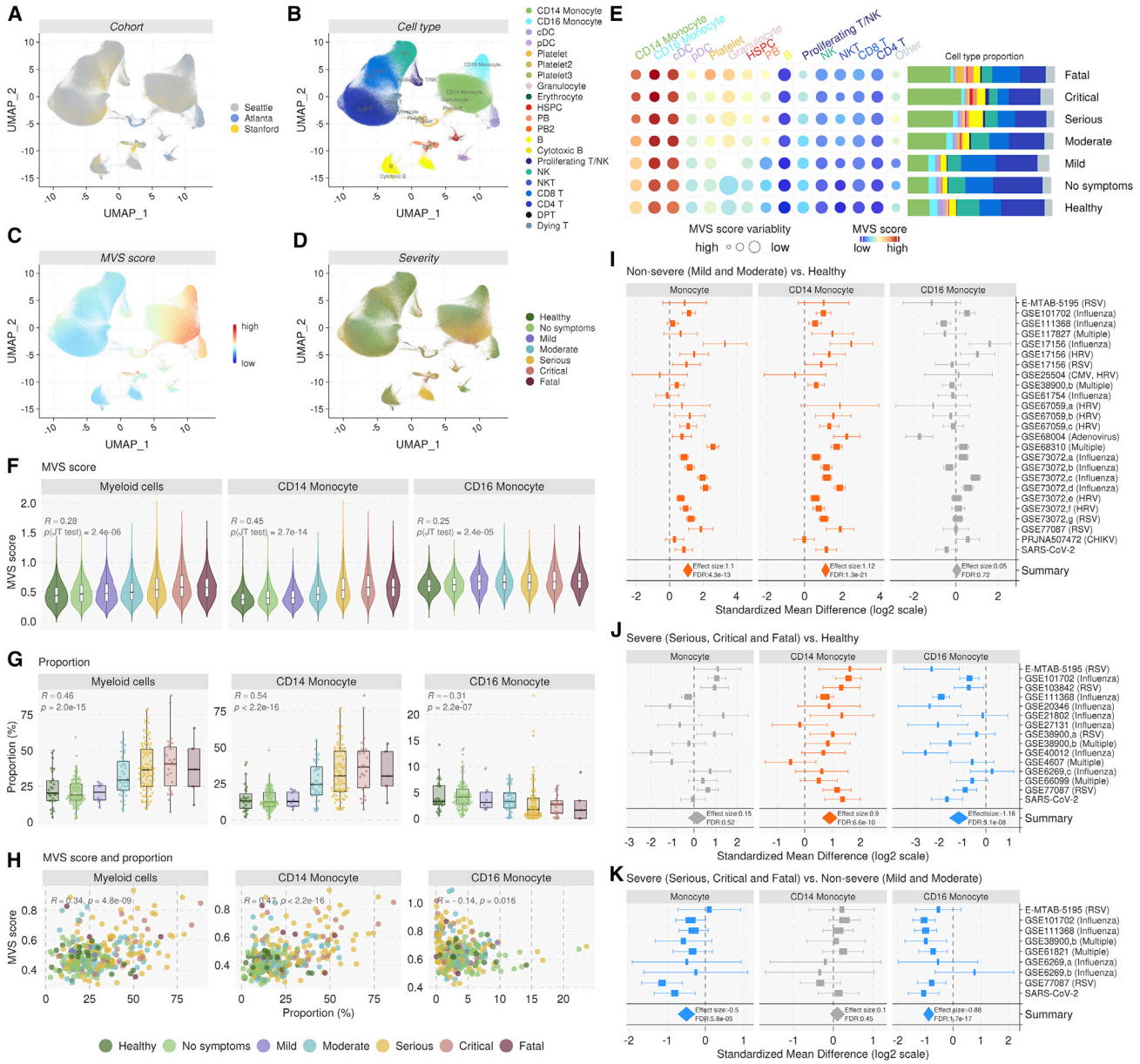


Figure 2. Single-cell RNA-seq identifies monocytes as the primary source of the MVS

(A–D) UMAP visualization of 702,970 immune cells from 289 PBMC samples in three independent scRNA-seq datasets colored by (A) cohort, (B) cell type, (C) MVS score, and (D) severity of viral infection.

(E) Circle map depicting the average MVS score in each cell type in each severity category, where color represents the average MVS score in each cell type, and size is proportional to the variability of MVS score in the cell type with larger size representing lower variability. The barplot shows the mean proportion of each cell type in each severity category.

(F) MVS score in in myeloid cells, CD14+, and CD16+ monocytes. p values were computed using JT trend test.

(G) Proportion of myeloid cells, CD14+, and CD16+ monocytes in each sample. p values were computed using JT trend test.

(H) Correlation between the MVS score and proportion of myeloid cells, CD14+ and CD16+ monocytes. p values were computed using Pearson correlation test.

(I–K) Changes in proportions of total, CD14+, and CD16+ monocytes estimated using *in silico* deconvolution of bulk transcriptome profiles. Forest plots for change in proportions of total, CD14+, and CD16+ monocytes in (I) non-severe patients versus HCs, (J) severe patients versus HCs, and (K) severe versus non-severe patients. The x axes represent standardized mean difference between two groups, computed as Hedges' *g*, in log2 scale. The size of a rectangle is proportional to the standard error of mean difference in the study. Whiskers represent the 95% confidence interval. The diamonds represent overall, combined mean difference for a given cell type in a given comparison. Width of the diamonds represents the 95% confidence interval of overall mean difference. See also Figure S2.

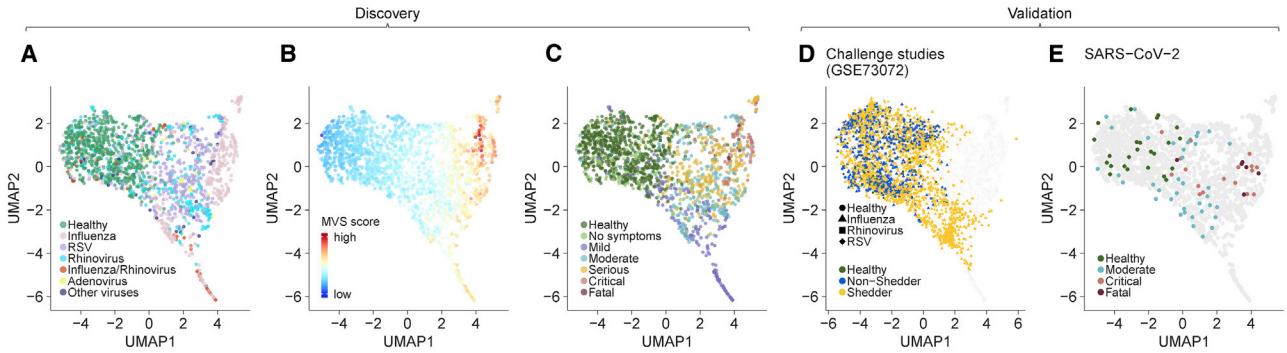


Figure 3. The MVS identifies distinct clusters of patients with non-severe and severe viral infection

(A–C) UMAP visualizations of the discovery cohorts (1,674 samples in 19 datasets) colored by (A) virus (B) MVS score, and (C) severity of viral infection.
(D and E) Projection of independent cohorts on the UMAP space obtained from the discovery cohorts: (D) seven challenge studies using influenza, RSV, or HRV in GSE73072 (2,518 samples) and (E) the SARS-CoV-2 cohort (86 samples).

Analogous to cellular differentiation analysis using single-cell profiling data, where each cell represents a snapshot along the differentiation trajectory, each sample in our analysis represents a snapshot of the host response to viral infection that spans from recognizing the presence of a virus to its elimination. To test this hypothesis, we adapted tSpace (Dermadi et al., 2020), a method for identifying cellular differentiation trajectories using single-cell data, to identify disease trajectories using bulk RNA data. We refer to the modified method as "disease space" (dSpace) (STAR Methods).

We co-normalized four of the seven randomly selected challenge studies (1,509 samples across 2 influenza, 1 HRV, and 1 RSV studies) with 1674 samples from 19 datasets using COCONUT to leverage a large number of longitudinal samples that can aid in a more accurate inference of the host response trajectories. Overall, we applied dSpace to 3,183 COCONUT co-normalized samples (1,663 HCs, 343 no symptoms, 514 mild, 286 moderate, 286 serious, 80 critical, 11 fatal) from 25 independent cohorts. We used these left-out challenge studies for validation of the inferred trajectories and to avoid possibility of introducing class imbalance because subjects in the challenge studies only had mild viral infections.

The first principal component of dSpace (dPC1) correlated with the severity of viral infection, and the second component (dPC2) distinguished hospitalized patients with viral infection from non-hospitalized patients with mild infection (Figure 4A). Importantly, participants from the influenza, RSV, and HRV challenge studies clustered almost exclusively with patients with mild infection (Figure S3A).

Next, we clustered samples using the disease space matrix, and used the resulting clusters to isolate trajectories associated with the severity of viral infection (STAR Methods). Clustering the samples using the dSpace matrix identified 20 clusters such that one category of samples dominated a cluster (Figure 4B): clusters 1–5, in which HCs and asymptomatically infected or convalescent patients accounted for >80% of samples; clusters 6–10, in which patients with mild viral infection accounted for >68% of samples, and clusters 13–20, in which hospitalized patients with moderate, serious, critical, or fatal viral infection accounted for >77% of samples (Figure 4C). Clusters 11 and 12 were heterogeneous as no one group of samples dominated them. Out of 1,509 samples, 1,507 (99.9%) from the influenza, RSV, and HRV challenge studies fell within clusters 1–12 (Figure S3B), demonstrating the robustness of the clusters defined using dSpace. We fit a principal trajectory line to the dSpace matrix, which consisted of healthy patients in the center and two divergent trajectories: one dominated by patients with mild viral infection and the other dominated by hospitalized patients with viral infection (Figure 4D, STAR Methods). Hitherto, we refer to these trajectories as "mild trajectory" and "severe trajectory," respectively. Together, trajectory analysis using dSpace showed that hospitalized patients with viral infection follow a different trajectory than those with mild infection compared to HCs, irrespective of the infecting virus.

Proportions of NK cells and the expression of NK cell-specific genes negatively correlate with the severity of viral infection

We identified 96 genes within the MVS that were significantly different between the two trajectories (Figure 5A). Using the MetaSignature database (https://metasignature.stanford.edu), we found the majority of the genes negatively correlated with the severity of viral infection are preferentially expressed in lymphocytes (T cells, B cells, and NK cells), whereas the majority of the genes positively correlated with severity are preferentially expressed in myeloid cells (granulocytes, monocytes, mDCs, and macrophages) (Figure 5B) (Haynes et al., 2017; Vallania et al., 2018).

Several NK cell-specific genes from the killer cell lectin-like receptor (KLR) family (*KLRB1, KLRG1, KLRD1*) and phosphoinositide-3-Kinase (PI3K) signaling genes (*PIK3R1*) were negatively correlated with severity (Figure 5C and Figure S4A). These genes were also lower in critical and fatal SARS-CoV-2 infections compared to HCs (Figure 5D, Figure S4B). PI3K signaling in NK cells and mutations in *PIK3R1* have been linked with human immunodeficiency and viral infections (Mace, 2018). Therefore, we hypothesized that NK cell proportions decreased with increased severity of viral infection. Deconvolution of bulk transcriptome showed the proportions of NK cells were significantly lower in patients with severe viral infections compared to HCs (ES = −0.85, FDR = 8.97e-05) and non-severe viral infections (ES = −1.03, FDR = 1.13e-06) (Figure 5E, Table S4). Further,

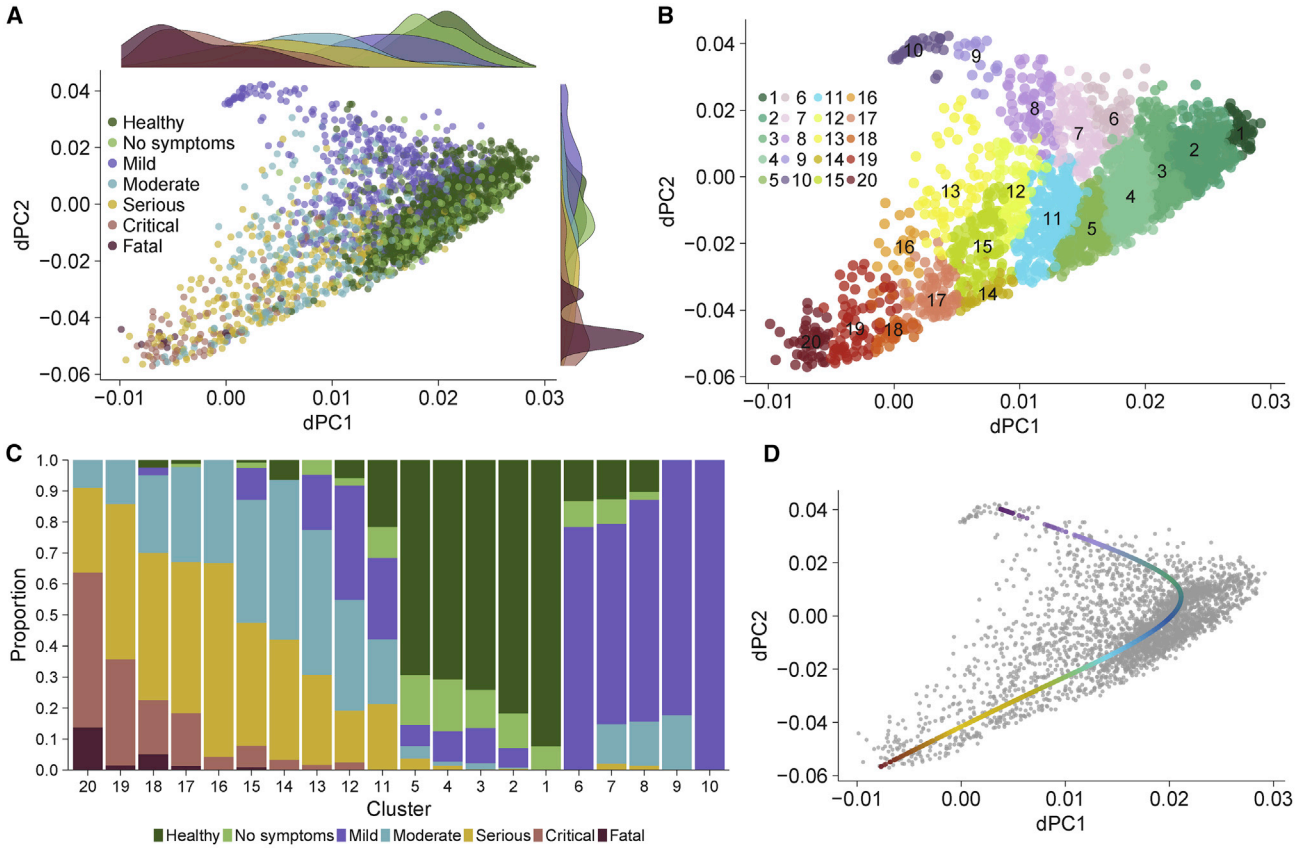


Figure 4. Patients with non-severe and severe viral infection follow divergent disease trajectories

(A) Trajectory analysis using dSpace (3,183 samples in 25 cohorts).

(B) Clustering of samples using dSpace.

(C) Proportion of samples for each severity category in each cluster.

(D) A principal line on dSpace coordinates identified by trajectory analysis. The red and purple colors of the line ends indicate the severe and non-severe trajectories, respectively. See also Figure S3.

across the three scRNA-seq cohorts, NK cell proportions were inversely correlated with severity (Figure 5F). Together, trajectory analysis using dSpace, deconvolution using immunoStates, and scRNA-seq found that the proportions of NK cells and the expression of NK cell-associated genes reduced with increased severity of viral infection, irrespective of the infecting virus.

Patients with severe viral infection show reduced antigen presentation

Trajectory analysis identified *HLA-DPB1*, a key HLA class II gene expressed in antigen presenting cells, was negatively correlated with severity, including in patients with SARS-CoV-2 infection (Figure 5A, Figures S5A and S5B). Downregulation of HLA class II in severe COVID-19 patients has been reported previously (Wilk et al., 2020). Expression of *HLA-DPB1* was lower in several antigen presenting cells (CD14 monocytes, conventional dendritic cells (cDCs), and B cells) in patients with moderate and severe viral infection in three scRNA-seq cohorts (Figures S5C and S5D) along with a reduced proportion of cDCs in severe viral infections in both scRNA-seq (Figure S5E) and bulk transcriptome deconvolution (Table S4). Together, our results showed that expression of HLA class II genes is negatively correlated with viral infection.

Myeloid-derived immune suppression is higher in patients with severe viral infection

Several differentially expressed genes between the two trajectories (Figure 5A) were preferentially expressed by immune cells of the myeloid lineage (Figure 5B). Although a subset of positively correlated genes with viral infection severity (*CAMP*, *BCAT1*, *LCN2*, *TXN*) have known proinflammatory functions in myeloid cells (Bertini et al., 1999; Bruns et al., 2015; Choi and Fujii, 2019; Eriksson et al., 2017; Papathanassiu et al., 2017; Ramos-Martínez et al., 2018) (Figure S4C), we found strong evidence of increased myeloid cell-derived immune suppression in patients with severe viral infection. Markers of polymorphonuclear myeloid-derived suppressor cells (PMN-MDSCs), *CEACAM8* (*CD66B*; Figure 5G) and *OLR1* (*LOX-1;* Figure S4C), and markers of monocytic MDSCs (M-MDSCs), *IL-4R* (Figure 5G), *ITGAM* (*CD11B;* Figure S4D), including a functional marker of MDSCs, *ARG1* (Figure S4D), were positively correlated with the severity of viral infection. *ORM1*, which drives the differentiation of monocytes to anti-inflammatory M2b macrophages (Nakamura et al., 2015), was significantly different between the two trajectories. Genes known to reduce the type I interferon (IFN) response, *GRN* and *BCL6* (Wei et al., 2019; Wu

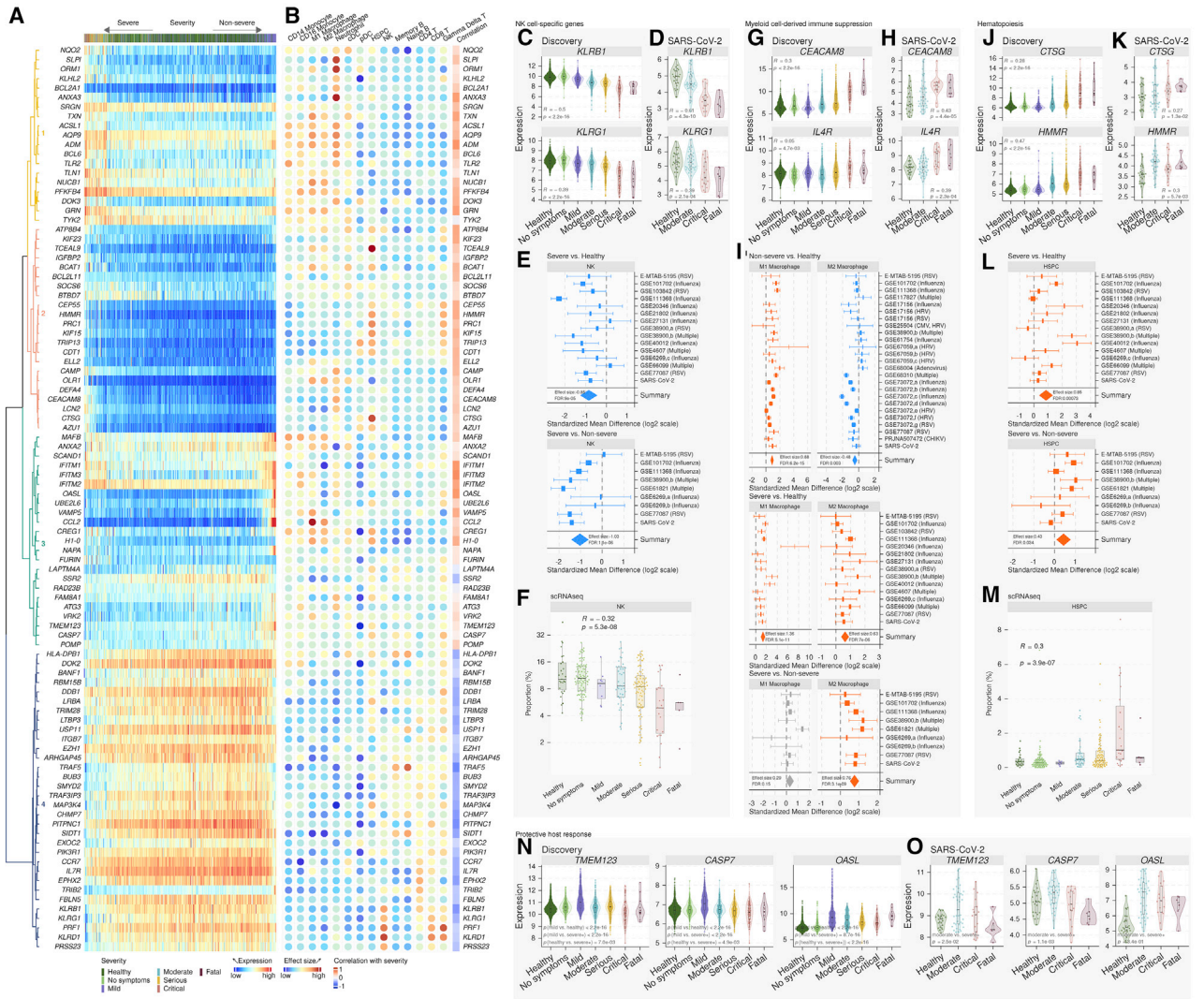


Figure 5. Immune responses from NK cells, myeloid cell-derived suppression, and hematopoiesis are associated with severity of viral infection

Violin plots: each dot represents a sample, and the y axis represents expression of the corresponding gene in a sample. p values were computed using Spearman rank correlation test or Mann–Whitney U test. Boxplots: each dot represents a sample, and the y axis represents proportion of the corresponding cell type in a sample. p values were computed using JT trend test. Forest plots: change in proportions between two groups for a given immune cell type, obtained by *in silico* deconvolution, where the x axis represents standardized mean difference (Hedges' *g*) between two groups in log2 scale. The size of a rectangle is proportional to the standard error of mean difference in the study. Whiskers represent the 95% confidence interval. The diamonds represent summary mean difference for a given cell type in a given comparison. Width of a diamond represents the 95% confidence interval of overall mean difference.

(A) Expression heatmap of the 96 trajectory-defining genes. Rows represent genes and columns represent samples, ordered by position along the disease trajectory. Colors of the dendrograms indicate gene modules.

(B) For each gene, effect size (Hedge's *g*) in a given cell type compared to all other cell types and correlation with severity.

(C and D) Expression of NK cell-specific genes in (C) the discovery (3,183 samples in 25 cohorts) and (D) a validation cohort (24 HCs, 62 SARS-CoV-2).

(E) Change in proportions of NK cells in patients with severe (top panel) and non-severe viral infection (bottom panel) compared to HCs.

(F) Proportions of NK cells along the severity of viral infection in three independent scRNA-seq cohorts.

(G and H) Expression of *CEACAM8* and *IL4R* in peripheral blood samples from patients with viral infection in (G) the discovery (3,183 samples in 25 cohorts) and (H) a validation cohort (24 HCs, 62 SARS-CoV-2).

(I) Change in proportions of pro-inflammatory macrophages (M1) and anti-inflammatory macrophages (M2) in patients with non-severe or severe viral infection compared to HCs.

(J and K) Expression of HSPC-specific genes in patients with viral infection in (J) the discovery (3,183 samples in 25 cohorts) and (K) a validation cohort (24 HCs, 62 SARS-CoV-2).

(L) Change in proportions of HSPCs patients with severe viral infection compared to HCs (top panel) and non-severe viral infection (bottom panel).

(M) Proportions of HSPCs along the severity of viral infection in three independent scRNA-seq cohorts.

(N and O) Genes with higher expression in patients with mild or moderate viral infection compared to HCs and those with severe viral infection in (N) the discovery and (O) a validation cohort. See also Figures S4 and S5.

et al., 2016), were positively correlated with severity (Figure S4C). All genes but *GRN* positively correlated with severity in the independent cohort of patients with SARS-CoV-2 infection (Figure 5H, Figure S4D). Notably, *ORM1* expression was lower in mild patients but higher in severe patients compared to HCs, including SARS-CoV-2-infected patients (Figures S4C and S4D).

Therefore, we hypothesized that proportions of pro- and anti-inflammatory macrophages would differ between patients with non-severe versus severe viral infection. Deconvolution analysis showed proportions of pro-inflammatory (M1) macrophages were higher in patients with non-severe (ES = 0.88, FDR = 6.16e-15) and severe (ES = 1.36, FDR = 5.12e-11) viral infection compared to HCs (Table S4), whereas proportions of anti-inflammatory (M2) macrophages were lower in non-severe patients (ES = −0.48, FDR = 3.00e-03), but higher in severe patients compared to HCs (ES = 0.63, FDR = 7.02e-06) and non-severe patients (ES = 0.76, FDR = 3.12e-09). Together, we found strong evidence of increased myeloid-derived immune suppression in patients with severe viral infection.

Increased hematopoiesis in patients with severe viral infection

Several differentially expressed genes between the two trajectories (*CTSG*, *PRC1*, *DEFA4, KIF15, TCEAL9, HMMR, CEP55,* and *AZU1*) were overexpressed in patients with severe viral infection, but not in those with non-severe viral infection compared to HCs (Figures 5J–5K, Figures S4E and S4F). All but one of these genes (*DEFA4*) have higher expression in circulating HSPCs (Figure 5B). Therefore, we investigated whether HSPCs were higher in patients with severe viral infection, but not in those with non-severe viral infection. Deconvolution analysis found that HSPCs were significantly higher in patients with severe viral infection compared to HCs (ES = 0.85, FDR = 7.33e-04) and compared to patients with non-severe viral infection (ES = 0.43, FDR = 3.38e-02) (Figure 5L, Table S4), but not in those with non-severe viral infection compared to HCs (Table S4). Proportions of HSPCs increased with severity in scRNA-seq across three independent cohorts of SARS-CoV-2-infected patients (Figure 5M).

Trajectory analysis identifies a protective host response associated with mild viral infections

Finally, dSpace analysis identified several genes (*CCL2*, *OASL, CASP7, TMEM123, MAFB, VRK2, UBE2L6, NAPA*) higher in patients with mild viral infection than those with severe viral infection or HCs (Figures 5N and 5O, Figures S4G and S4H). *CCL2,* a type I IFN receptor-mediated chemoattractant that promotes monocyte migration to the site of infection, and *OASL*, a type I IFN-induced gene, had higher expression in patients with mild viral infection. *CASP7* is cleaved by *CASP3* and *CASP10* and is activated upon cell death stimuli and induces apoptosis. *TMEM123* (*PORIMIN*) is a cell surface receptor that mediates oncosis, a type of cell death distinct from apoptosis characterized by a loss of cell membrane integrity without DNA fragmentation. Together, these results suggest that patients with a coordinated immune response involving monocyte recruitment, IFN response, and higher cell death have a lower risk of severe viral infection.

Protective and detrimental host response modules are associated with the severity of viral infection

Unsupervised hierarchical clustering grouped the 96 genes into four modules (Figure 5A). Module 1 and 2 were composed of genes preferentially expressed in myeloid and HSPCs and were higher in patients with severe viral infection (Figure 5B). Module 4 was composed of genes preferentially expressed in lymphoid cells (NK, T, and B cells). Genes in module 3 and 4 were higher expressed in patients with mild viral infection compared to those with severe infection (Figure 5B). These four modules broadly divided the host response genes differentially expressed between two trajectories into two categories: a detrimental host response represented by module 1 and 2 (higher in patients with severe viral infection), and a protective host response represented by module 3 and 4 (higher in patients with mild viral infection).

We selected 42 out of 96 genes with absolute effect size $^{3}1$ between the severe and mild trajectories (Table S5), resulting in 11, 13, 10, and 8 genes in modules 1, 2, 3, and 4, respectively. Module scores, defined as the geometric mean of expression of these reduced sets of genes in a given module, continued to be significantly positively (module 1, 2, and 3) and negatively (module 4) correlated with severity of viral infection ($|r|^{3}0.43$, $p < 2.23\text{-}16$; Figure 6A), which suggested that genes within each module are correlated with each other. Indeed, we found most pairs of genes within each module were positively correlated, irrespective of their infection status (Figure 6B).

We found the correlation structure within each module changed depending on the presence and severity of infection. Pairwise correlations between genes in modules 1, 2, and 4 were significantly higher in patients with severe viral infection than HCs or patients with mild viral infection ($p < 5e\text{-}05$; Figure 6C). Pairwise correlations in module 2 were significantly lower in patients with mild infection compared to HCs ($p = 3.6e\text{-}07$; Figure 6C). In contrast, pairwise correlations between genes in module 3, which included genes involved in the protective host response, were significantly higher in patients with mild infection compared to HCs and those with severe infection ($p = 5.7e\text{-}14$; Figure 6C). Together, these results show that the genes within each module are expressed in a coordinated manner depending on the infection status and severity of infection.

The protective host response modules are associated with IFN concentration in plasma proteome but decoupled from the IFN response in patients with severe viral infection

Recent reports have described higher expression of IFN-stimulated genes (ISGs) in patients with moderate SARS-CoV-2 infection than those with severe infection (Arunachalam et al., 2020). Therefore, we investigated whether this observation is generalizable to other viruses. Indeed, module 3 included three IFN-induced transmembrane (IFITM) genes (*IFITM1, IFITM2, IFITM3*), involved in the restriction of multiple viruses (Bailey et al., 2014), that were overexpressed in patients with viral infection and positively correlated with severity (Figure S6A). We also found several type I and II IFN receptors overexpressed during viral infection that positively correlated with severity, irrespective of the infecting virus (Figure S6A). In patients with mild viral infection, the distribution of correlations between IFITMs and genes in the

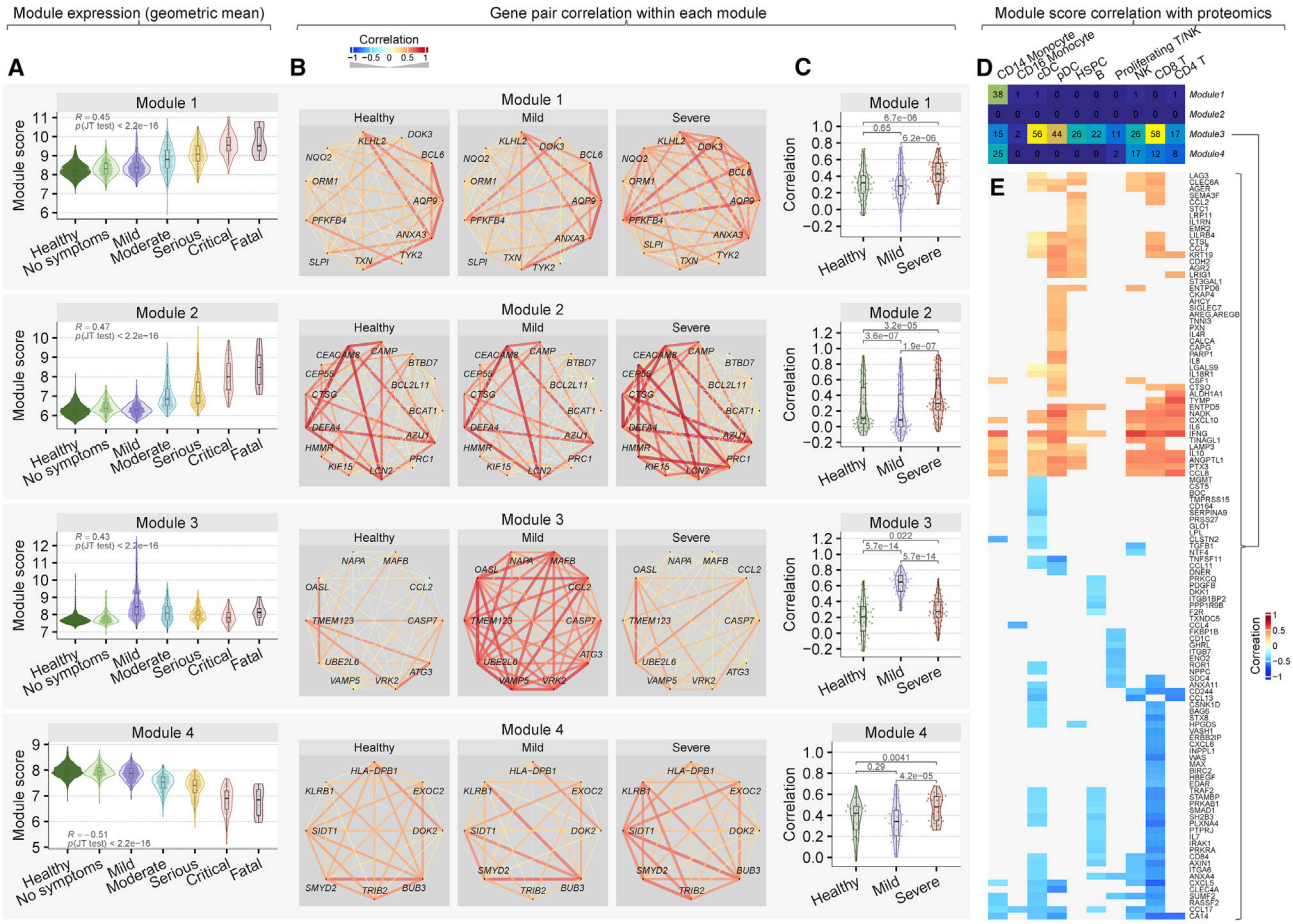


Figure 6. Coordinated protective and deleterious host response modules associated with severity of viral infection

(A) Module scores, defined as the geometric mean of expression of genes in each module in the sample, for the discovery cohorts (3,183 samples in 25 cohorts).

(B) Pairwise Spearman's rank correlation coefficient between genes in each module in HCs and patients with mild or severe viral infection. The width of the line indicates strength of correlation; red and blue color indicate positive and negative correlation, respectively.

(C) Each dot in the violin plots represents the correlation between a pair of genes. p values computed using Wilcoxon signed-rank test.

(D) Number of proteins correlated with detrimental and protective module scores in each immune cell type in the Seattle cohort.

(E) Heatmap of correlation between proteins in plasma samples and module 3 score in the Seattle cohort in immune cell types. See also Figure S6.

protective response module 3 was significantly higher than in patients with severe viral infection or HCs ($p \leq 1e\text{-}06$; Figure S6B). Further, the distribution of correlations between the type I and II IFN receptors and the protective response module 3 was not statistically different between HCs and patients with severe viral infection, but was significantly higher in patients with mild viral infection ($p \leq 0.03$; Figure S6C).

We used 242 samples in the Seattle cohort for which both scRNA-seq and proteomic data were available to investigate whether detrimental or protective module scores are associated with IFN concentrations in plasma samples. Module 2 scores were not associated with any protein in any immune cell types, presumably because the majority of the genes in module 2 are specifically expressed in neutrophils, which are not included in PBMCs (Figure 6D). Scores for module 1, 3, and 4 were associated with several proteins in CD14+ monocytes. Module 3 score was associated with the several proteins in each immune cell type (Figures 6D and 6E) and had statistically significant positive correlation with IFN-gamma (IFNG) concentration across all cell types except CD16+ monocytes and proliferating T/NK cells, further suggesting that patients with higher module 3 scores may have better outcomes. Together, we found that expression of the ISGs increase with the severity of viral infection, but their correlation with the protective host response does not increase in patients with severe viral infection as much as those with mild viral infection. These results also show that higher module 3 score in several immune cell types is correlated with increased IFNG. Collectively, these results show a decoupling of the protective host response from the IFN response in patients with severe viral infection, irrespective of the virus.

Host response-based module score improves classification of patients with severe and non-severe viral infection

Despite significant correlation with the severity, the MVS score cannot adequately distinguish severe and non-severe patients (Figures S7A and S7B). We hypothesized that incorporating the protective and detrimental host responses in a score would

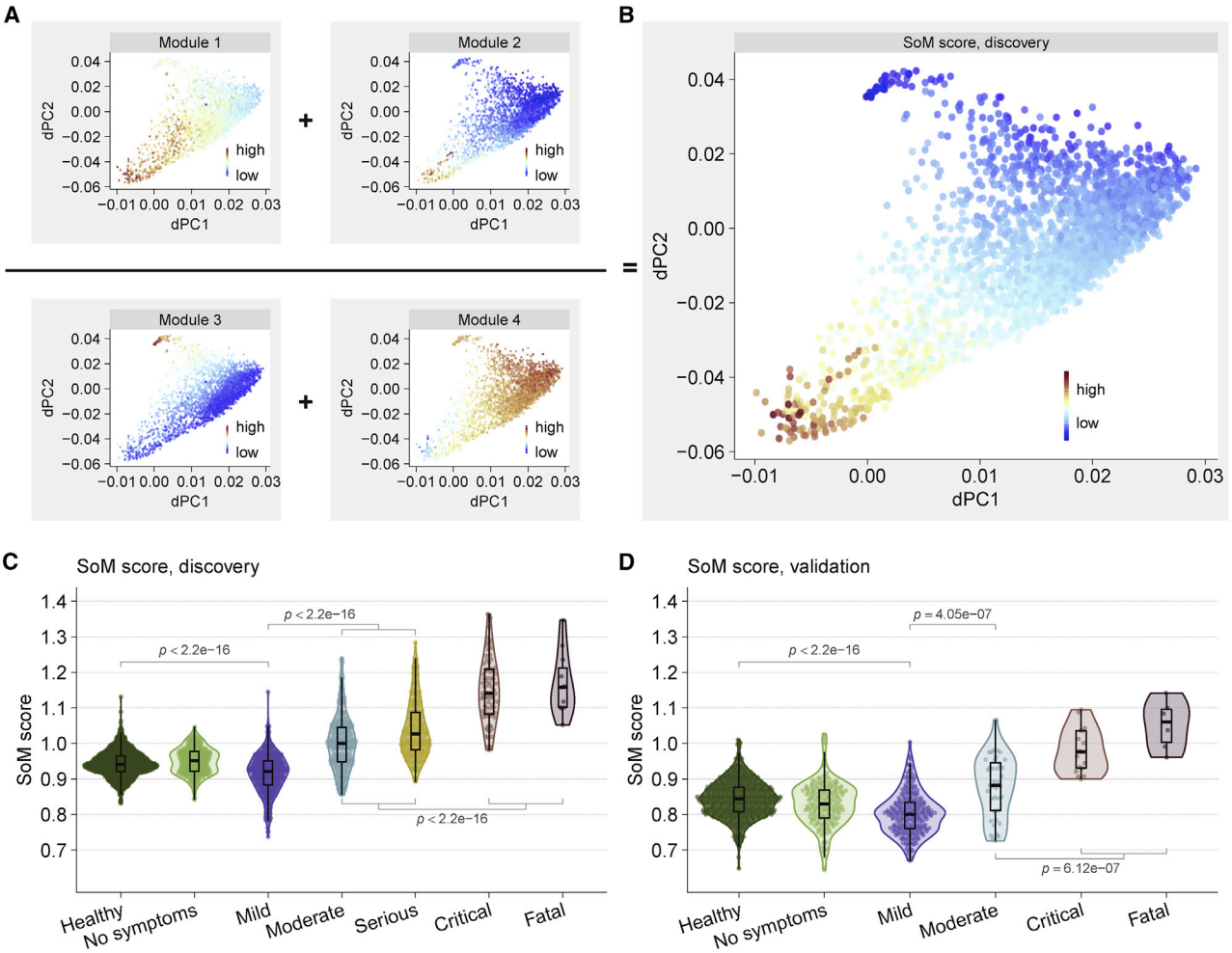


Figure 7. Host response modules improve classification of patients with severe and non-severe viral infection

(A) Module scores and (B) the SoM score across the 3,183 dSpace samples.

(C and D) The SoM score distinguishes mild and severe viral infection in the (C) discovery cohort and (D) validation cohort. Each point represents a sample. p values computed using Mann–Whitney U test. See also Figure S7.

improve discrimination between severe and non-severe viral infection. Therefore, we defined the Severe-or-Mild (SoM) score of a sample as the sum of the scores for module 1 and 2 divided by the sum of the scores for module 3 and 4 (STAR Methods). The SoM score showed a more pronounced gradient between the severe and mild trajectories than any of the individual module scores (Figure 7B). The SoM score distinguished patients with mild infection from those with severe infection with AUROC^{3}0.929 in the 3,183 samples from the discovery cohorts (Figure 7C, Figure S7C) and with AUROC > 0.98 in 5 independent validation cohorts comprised of 1,154 samples from patients infected with 4 different viruses (SARS-CoV-2, influenza, HRV, chikungunya) (Figure 7D, Figure S7D). In patients with non-severe viral infection, the SoM score distinguished those with mild infection from those with moderate infection with AUROC > 0.75 in discovery and validation datasets compared to the MVS score (AUROC < 0.63) (Figures S7A–S7D). In hospitalized patients with a viral infection, the SoM score also distinguished those with moderate infection from those with critical or fatal infection with higher accuracy than the MVS score (Figures S7A–S7D). Together, our results show that the protective and detrimental host response modules identified by trajectory analysis improve discrimination accuracy between patients with mild, moderate, and severe viral infection. Suppressing the detrimental host response modules or enhancing the protective host response modules could be therapeutic targets for host-directed broad-spectrum intervention in patients with severe viral infection.

DISCUSSION

The four viral pandemics since 2009 have underscored an urgent unmet need for creating generalizable diagnostic and therapeutic tools to enable faster deployment during infectious outbreaks. There is a strong impetus to develop virus-agnostic strategies that can be applied across multiple viruses. Here, we tested hypotheses that our previously described conserved host response to respiratory viral infections (Andres-Terre et al., 2015) represents biological mechanisms associated with

severity that are distinct between patients with non-severe and severe outcomes, irrespective of infecting virus.

Although we had identified the MVS by analyzing respiratory viruses, it is generalizable across novel viruses, including SARS-CoV-2, chikungunya, and Ebola and across ages. Out of 37 cohorts, 12 cohorts consisted of 931 samples from children (<18 years), of which 643 were from children younger than 2 years. Our results demonstrate a conserved similarity in the dysregulation of the host immune response in patients with severe outcomes, which presents several opportunities the development of diagnostic and prognostic tests, identification of drug targets for host-directed broad-spectrum antiviral therapies, and drug repurposing to improve global pandemic preparedness for the pandemics that will invariably come in the future.

The SoM score using the 42-gene signature distinguished patients with a severe outcome from those with a non-severe outcome with very high accuracy. Arguably, the 42-gene signature is not optimal for clinical translation. However, given the high pairwise correlation between genes within each module, only a small subset of genes within each module could provide the same discriminatory power, allowing identification of a parsimonious gene signature. Trajectory analysis also suggests that the SoM score has the potential to predict the severity of outcome in patients with viral infection, though it needs to be tested in additional cohorts.

High pairwise correlations between genes in module 3 in patients with mild viral infection, irrespective of virus, suggest that a highly coordinated immune response between monocyte recruitment, interferon response, and cell death is associated with protection. Our results are consistent with recent observations that ISGs are strongly induced in patients with moderate SARS-CoV-2 infection compared to those with severe SARS-CoV-2 infection (Arunachalam et al., 2020; Hadjadj et al., 2020), and generalize to patients with non-severe versus severe infection, irrespective of the virus.

The genes in module 3 were more correlated with interferon-induced transmembrane proteins (IFITMs) in patients with mild infection compared to those with severe viral infection. IFITMs are involved in restricting viruses at various stages of the life cycle, including (1) blocking host cell entry by trapping virions in endosomal vesicles, (2) inhibiting viral gene expression and protein synthesis, and (3) disrupting viral assembly (Liao et al., 2019; Zhao et al., 2019). The lower correlations between the IFITMs and the genes in module 3 suggest that the IFN-induced response is ''decoupled'' from the protective response in during severe viral infection. Understanding the mechanisms underlying this decoupling could lead to targets for host-directed antiviral therapy.

Analysis of scRNA-seq in 3 independent cohorts and *in silico* deconvolution across 32 cohorts found increased HSPCs in patients with severe viral infection, irrespective of the virus, suggesting that emergency hematopoiesis is associated with increased risk of severity, possibly as the immune response fails to adequately respond to the virus. In contrast, we have previously shown reduced proportions of HSPCs in mild viral infection (Bongen et al., 2018), which may reflect the production of myeloid cells at the expense of the lymphoid compartment to replenish myeloid cells during infection (Takizawa et al., 2012). We found increased myeloid cells and reduced lymphoid cells in both scRNA-seq and deconvolution analysis, supporting a model where human HSPCs take an active role in the immune response by differentiating into myeloid cells, in line with our previous observation (Bongen et al., 2018).

In line with recent studies (Gatti et al., 2020; Hadjadj et al., 2020; Silvin et al., 2020; Zhou et al., 2020), we found decrease in CD16+ monocytes with increased severity. This suggests that reduced CD16+ monocytes in peripheral blood, possibly due to efflux to the site of infection in response to ongoing tissue damage or dysregulated cytokine sensing, is a conserved feature of the host response in severe viral infection across viruses and may have prognostic significance.

The mechanisms underlying the differential influence of CD14+ versus CD16+ monocytes and decoupling between the IFN-response and protective response remain unknown, though mounting evidence suggests dysfunctional myeloid cells in patients with severe viral infection. Dysfunctional CD14+ monocytes characterized by low expression of HLA-DR and high expression of alarmins are associated with severe COVID-19 infection (Schulte-Schrepping et al., 2020). However, our results show that it is a generalizable feature of severe outcome in other viral infections. Low HLA-DR expression in monocytes is a well-described marker of worse outcomes in sepsis and trauma (Hoffmann et al., 2017; Monneret et al., 2006; Venet et al., 2020). We found expression of *HLA-DPB1,* a MHC class II gene, in CD14+ monocytes was inversely correlated with severity of infection, whereas expression of alarmins (e.g., S100A9) in CD14+ monocytes was positively correlated with severity. Although overproduction of IL-6 mediates the low HLA-DR expression on CD14+ monocytes of patients with severe COVID-19 (Giamarellos-Bourboulis et al., 2020; Silvin et al., 2020), anti-IL6 therapy has been unsuccessful in showing improved outcomes in COVID-19 (Stone et al., 2020), highlighting the urgent need for elucidating the underlying mechanisms.

Our correlation analysis of protective module scores with plasma proteomic data in the patients with SARS-CoV-2 infection suggest dysfunctional myeloid cells in patients with severe outcome. Protective modules were significantly correlated with IFNG in antigen-presenting cells (CD14+ monocytes, cDCs, pDCs, B cells) and cytotoxic cells (NK cells, T cells). Combined with the observation that several class II HLA genes are downregulated in patients with severe outcome, it is possible that dysfunctional antigen-presenting cells are unable to orchestrate a subsequent adaptive immune response.

It is possible that factors such as genetics, comorbidities, and initial viral load, together or individually, cause an initial overactivation of monocytes, which leads to a sequence of CD14+ overactivation and secretion of inhibitors of HLA genes, further leading to a cascade of defective antigen presentation and T cell differentiation that fail to control ongoing infection. This may ultimately lead to the characteristic profile of severe disease: overactive dysfunctional myeloid and deficient lymphoid compartments amidst immunosuppression and a cytokine storm.

We found increased proportions of PMN- and M-MDSCs, and anti-inflammatory macrophages along with higher expression of their phenotypic and functional markers in patients with severe viral infection, but not in those with mild viral infection, irrespective of the virus. These results suggest that lower MDSCs in the early phase of infection is protective and provide strong

evidence that, although increased PMN- and M-MDSCs may limit hyperinflammation during active viral infection, they may lead to a detrimental amplification of immunosuppression, irrespective of the virus. The modulation of monocyte responses, as reflected by gene expression, is compatible with the detrimental role of monocytes/macrophages in severe SARS-CoV-2 infection associated with respiratory dysfunction (Giamarellos-Bourboulis et al., 2020).

Among their immunosuppressive roles, MDSCs are known to suppress NK cell activity through arginase and ROS/RNS (Schrijver et al., 2019). Our trajectory and *in silico* deconvolution analyses, and scRNA-seq data found several NK cell-specific genes and the proportions of NK cells were negatively correlated with the severity of viral infection. We have previously shown that healthy individuals with lower expression of *KLRD1* are more likely to be infected when challenged (Bongen et al., 2018). A negative correlation between expression of *KLRD1* and the severity of viral infection, including SARS-CoV-2, further emphasizes that *KLRD1*-expressing NK cells may play a protective role following infection, irrespective of the infecting virus.

Taken together, our analyses offer a systems view of the immune state during viral infection and factors that mediate and predict progression to mild or severe outcomes, despite the heterogeneity and regardless of the infecting virus. We identified host response modules that could lead to new intervention strategies, including diagnostics for predicting patients at higher risk of severe outcomes, and broad-spectrum host-directed therapies for improved pandemic preparedness.

Limitations of Study

Our analysis has two potential limitations. First, COCONUT co-normalization may have removed variability in immune responses of HCs of different ages and from different geographic regions. However, our results demonstrate that the effect of acute viral infection on the host immune response is substantially larger to overcome this heterogeneity. Second, it is possible that the MVS may be biased toward respiratory viral infections because it was discovered using respiratory viruses. However, the MVS score is also higher in patients with chikungunya or Ebola infection, suggesting that the MVS may be conserved across non-respiratory viruses too. It is also possible that the MVS only represents conserved host response to acute viral infection, and that chronic viruses such as hepatitis B or C, HIV, CMV, EBV, etc. may not evoke the same host response, which should be a focus of future studies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Dataset collection and preprocessing
  - Standardized severity assignment
  - Viral challenge studies
  - COCONUT co-normalization
  - MVS genes and score
  - RNA sequencing analysis
  - Detection of viral reads in RNA-seq data
  - Analysis of single-cell RNA-seq data
  - In silico cellular deconvolution using immunoStates and multi-cohort analysis of estimated cellular proportions
  - Trajectory inference analysis
  - Statistical analysis of trajectories identified with dSpace
  - Calculation of the SoM score
  - Correlation of cell proportions, MVS score, and SoM score with severity
  - Module score association with plasma proteomics

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.immuni.2021.03.002.

ACKNOWLEDGMENTS

P.K. is funded in part by the Bill and Melinda Gates Foundation (OPP1113682); the National Institute of Allergy and Infectious Diseases (NIAID) grants 1U19AI109662, U19AI057229, and 5R01AI125197; Department of Defense contracts W81XWH-18-1-0253 and W81XWH1910235; and the Ralph & Marian Falk Medical Research Trust. A.M.R. is funded by National Science Foundation Graduate Research Fellowship and Stanford Graduate Fellowship. Y.L. is funded by the Knight-Hennessy Scholars Program and National Science Foundation Graduate Research Fellowship. J.T. is funded by National Science Scholarship (PhD) from the Agency of Science, Technology, and Research (A*STAR), Singapore.

AUTHOR CONTRIBUTIONS

P.K. conceived and supervised the study. H.Z., A.M.R., D.D.B., J.T., L.M.J., and M.D. collected, annotated, processed, and analyzed data. P.K., H.Z., A.M.R., D.D.B., J.T., L.M.J., M.D., and Y.L. interpreted analyses results and wrote the manuscript. M.K., T.M., and E.J.G.B. enrolled patients with SARS-CoV-2 infection in Greece. Y.H.B. and Y.D.H. profiled and processed RNA sequencing data from the patients with SARS-CoV-2 infection in Greece. Y.S. and J.R.H. profiled single-cell RNA-seq of PBMCs from patients with SARS-CoV-2 infection in the Seattle Cohort. C.L.D. and S.A.K. analyzed proteomic data from patients with SARS-CoV-2 infection in the Seattle cohort.

DECLARATION OF INTERESTS

E.J.G.B. has received honoraria from Abbott CH, Angelini Italy, bioMérieux Inc, InflaRx GmbH, MSD Greece, and XBiotech Inc.; independent educational grants from AbbVie, Abbott, Astellas Pharma Europe, AxisShield, bioMérieux Inc, InflaRx GmbH, ThermoFisher Brahms GmbH, and XBiotech Inc; and funding from the FrameWork 7 program HemoSpec (granted to the National and Kapodistrian University of Athens), the Horizon2020 Marie-Curie Project European Sepsis Academy (granted to the National and Kapodistrian University of Athens), and the Horizon 2020 European Grant ImmunoSep (granted to the Hellenic Institute for the Study of Sepsis). P.K. is a shareholder and a consultant to Inflammatix, Inc. Y.H.B. and Y.D.H. are employees of, and stockholders in, Inflammatix, Inc.

Received: September 23, 2020
Revised: December 3, 2020
Accepted: March 1, 2021
Published: March 24, 2021

REFERENCES

Andres-Terre, M., McGuire, H.M., Pouliot, Y., Bongen, E., Sweeney, T.E., Tato, C.M., and Khatri, P. (2015). Integrated, Multi-cohort Analysis Identifies Conserved Transcriptional Signatures across Multiple Respiratory Viruses. Immunity *43*, 1199–1211.

Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat. Immunol. *20*, 163–172.

Arunachalam, P.S., Wimmers, F., Mok, C.K.P., Perera, R.A.P.M., Scott, M., Hagan, T., Sigal, N., Feng, Y., Bristow, L., Tak-Yin Tsang, O., et al. (2020). Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. Science *369*, 1210–1220.

Bailey, C.C., Zhong, G., Huang, I.-C., and Farzan, M. (2014). IFITM-Family Proteins: The Cell's First Line of Antiviral Defense. Annu. Rev. Virol. *1*, 261–283.

Bekerman, E., and Einav, S. (2015). Infectious disease. Combating emerging viral threats. Science *348*, 282–283.

Berdal, J.-E., Mollnes, T.E., Wæhre, T., Olstad, O.K., Halvorsen, B., Ueland, T., Laake, J.H., Furuseth, M.T., Maagaard, A., Kjekshus, H., et al. (2011). Excessive innate immune response and mutant D222G/N in severe A (H1N1) pandemic influenza. Journal of Infection *63*, 308–316.

Bermejo-Martin, J.F., Martin-Loeches, I., Rello, J., Antón, A., Almansa, R., Xu, L., Lopez-Campos, G., Pumarola, T., Ran, L., Ramirez, P., et al. (2010). Host adaptive immunity deficiency in severe pandemic influenza. Critical Care *14*, R167.

Bertini, R., Howard, O.M., Dong, H.F., Oppenheim, J.J., Bizzarri, C., Sergi, R., Caselli, G., Pagliei, S., Romines, B., Wilshire, J.A., et al. (1999). Thioredoxin, a redox enzyme released in infection and inflammation, is a unique chemoattractant for neutrophils, monocytes, and T cells. J. Exp. Med. *189*, 1783–1789.

Bongen, E., Vallania, F., Utz, P.J., and Khatri, P. (2018). KLRD1-expressing natural killer cells predict influenza susceptibility. Genome Med. *10*, 45.

Bruns, H., Büttner, M., Fabri, M., Mougiakakos, D., Bittenbring, J.T., Hoffmann, M.H., Beier, F., Pasemann, S., Jitschin, R., Hofmann, A.D., et al. (2015). Vitamin D-dependent induction of cathelicidin in human macrophages results in cytotoxicity against high-grade B cell lymphoma. Science Translational Medicine *7*, 282ra47–282ra47.

Choi, J.W., and Fujii, T. (2019). The Prevalence of Low Plasma Neutrophil Gelatinase-Associated Lipocalin Level in Systemic Inflammation and its Relationship with Proinflammatory Cytokines, Procalcitonin, Nutritional Status, and Leukocyte Profiles. Clin. Lab. *65*, https://doi.org/10.7754/Clin.Lab.2019.181235.

Christiansen, J. (2018). Global Infections by the Numbers. Sci. Am. *318*, 48–49.

Davenport, E.E., Antrobus, R.D., Lillie, P.J., Gilbert, S., and Knight, J.C. (2014). Transcriptomic profiling facilitates classification of response to influenza challenge. Journal of Molecular Medicine *93*, 105–114.

de Steenhuijsen Piters, W.A.A., Heinonen, S., Harsat, R., Bunsow, E., Smith, B., Suarez-Arrabal, M.-C., Chaussabel, D., Cohen, D.M., Sanders, E.A.M., Ramilo, O., et al. (2016). Nasopharyngeal Microbiota, Host Transcriptome, and Disease Severity in Children with Respiratory Syncytial Virus Infection. American Journal of Respiratory and Critical Care Medicine *194*, 1104–1115.

Dermadi, D., Bscheider, M., Bjegovic, K., Lazarus, N.H., Szade, A., Hadeiba, H., and Butcher, E.C. (2020). Exploration of Cell Development Pathways through High-Dimensional Single Cell Analysis in Trajectory Space. iScience *23*, 100842.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dunning, J., Blankley, S., Hoang, L.T., Cox, M., Graham, C.M., James, P.L., Bloom, C.I., Chaussabel, D., Banchereau, J., Brett, S.J., et al. (2018). Progression of whole-blood transcriptional signatures from interferon-induced to neutrophil-associated patterns in severe influenza. Nature Immunology *19*, 625–635.

Efron, B., and Tibshirani, R. (2002). Empirical bayes methods and false discovery rates for microarrays. Genet. Epidemiol. *23*, 70–86.

Eriksson, J., Gidlöf, A., Eriksson, M., Larsson, E., Brattström, O., and Oldner, A. (2017). Thioredoxin a novel biomarker of post-injury sepsis. Free Radic. Biol. Med. *104*, 138–143.

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T.F., and Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics *28*, 2678–2679.

Gatti, A., Radrizzani, D., Viganò, P., Mazzone, A., and Brando, B. (2020). Decrease of Non-Classical and Intermediate Monocyte Subsets in Severe Acute SARS-CoV-2 Infection. Cytometry A *97*, 887–890.

Giamarellos-Bourboulis, E.J., Netea, M.G., Rovina, N., Akinosoglou, K., Antoniadou, A., Antonakos, N., Damoraki, G., Gkavogianni, T., Adami, M.-E., Katsaounou, P., et al. (2020). Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. Cell Host Microbe *27*, 992–1000.e3.

Gupta, R.K., Turner, C.T., Venturini, C., Esmail, H., Rangaka, M.X., Copas, A., Lipman, M., Abubakar, I., and Noursadeghi, M. (2020). Concise whole blood transcriptional signatures for incipient tuberculosis: a systematic review and patient-level pooled meta-analysis. Lancet Respir. Med. *8*, 395–406.

Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Péré, H., Charbit, B., Bondet, V., Chenevier-Gobeaux, C., et al. (2020). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. Science *369*, 718–724.

Haynes, W.A., Vallania, F., Liu, C., Bongen, E., Tomczak, A., Andres-Terrè, M., Lofgren, S., Tam, A., Deisseroth, C.A., Li, M.D., et al. (2017). EMPOWERING MULTI-COHORT GENE EXPRESSION ANALYSIS TO INCREASE REPRODUCIBILITY. Pac. Symp. Biocomput. *22*, 144–153.

Heinonen, S., Jartti, T., Garcia, C., Oliva, S., Smitherman, C., Anguiano, E., de Steenhuijsen Piters, W.A.A., Vourinen, T., Ruuskanen, O., Dimo, B., et al. (2016). Rhinovirus Detection in Symptomatic and Asymptomatic Children: Value of Host Transcriptome Analysis. American Journal of Respiratory and Critical Care Medicine *193*, 772–782.

Hoang, L.T., Tolfvenstam, T., Ooi, E.E., Khor, C.C., Naim, A.N.M., Ho, E.X.P., Ong, S.H., Wertheim, H.F., Fox, A., Van Vihn Nguyen, C., et al. (2014). Patient-based transcriptome-wide analysis identify interferon and ubiquination pathways as potential predictors of influenza A disease severity. PLoS One *3*, e111640.

Hoffmann, S., Harms, H., Ulm, L., Nabavi, D.G., Mackert, B.-M., Schmehl, I., Jungehulsing, G.J., Montaner, J., Bustamante, A., Hermans, M., et al.; PREDICT Investigators (2017). Stroke-induced immunodepression and dysphagia independently predict stroke-associated pneumonia - The PREDICT study. J. Cereb. Blood Flow Metab. *37*, 3671–3682.

Jaggi, P., Meijas, A., Xu, Z., Yin, H., Moore-Clingenpeel, M., Smith, B., Burns, J.C., Tremoulet, A.H., Jordan-Villegas, A., Chaussabel, D., et al. (2018). Whole blood transcriptional profiles as a prognostic tool in complete and incomplete Kawasaki Disease. PLoS ONE *13*, e0197858.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127.

Jong, V.L., Ahout, I.M.L., van den Ham, H.-J., Jans, J., Zaaraoui-Boutahar, F., Zomer, A., Simonetti, E., Bijil, M.A., Brand, H.K., van Ijcken, W.F.J., et al. (2016). Transcriptome assists prognosis of disease severity in respiratory syncytial virus infected infants. Scientific Reports *6*, 36603.

Liao, Y., Goraya, M.U., Yuan, X., Zhang, B., Chiu, S.-H., and Chen, J.-L. (2019). Functional Involvement of Interferon-Inducible Transmembrane Proteins in Antiviral Immunity. Front. Microbiol. *10*, 1097.

Liu, T.-Y., Burke, T., Park, L.P., Woods, C.W., Zaas, A.K., Ginsburg, G.S., and Hero, A.O. (2016). An individualized predictor of health and disease using paired reference and target samples. BMC Bioinformatics *17*, 47.

Liu, X., Speranza, E., Munoz-Fontela, C., Haldenby, S., Rickett, N.Y., Garcia-Dorival, I., Fang, Y., Hall, Y., Zekeng, E.-G., Ludtke, A., et al. (2017). Transcriptomic signatures differentiate survival from fatal outcomes in humans infected with Ebola virus. Genome Biology *18*, 1–17.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550–21.

Mace, E.M. (2018). Phosphoinositide-3-Kinase Signaling in Human Natural Killer Cells: New Insights from Primary Immunodeficiency. Front. Immunol. *9*, 445.

Martiin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12.

Mayhew, M.B., Buturovic, L., Luethy, R., Midic, U., Moore, A.R., Roque, J.A., Shaller, B.D., Asuni, T., Rawling, D., Remmel, M., et al. (2020). A generalizable 29-mRNA neural-network classifier for acute bacterial and viral infections. Nat. Commun. *11*, 1177.

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software *3*, 861.

Mejias, A., Dimo, B., Suarez, N.M., Garcia, C., Suarez-Arrabal, M.C., Jartti, T., Blankenship, D., Jordan-Villegas, A., Ardura, M.I., Xu, Z., et al. (2013). Whole blood gene expression profiles to assess pathogenesis and disease severity in infants with respiratory syncytial virus infection. PLOS Medicine *10*, e1001549.

Michlmayr, D., Pak, T.R., Rahman, A.H., Ad David Amir, El, Kim, E.Y., Schulze, S.K., Suprun, M., Stewart, M.G., Thhomas, G.P., Balmaseda, A., et al. (2018). Comprehensive innate immune profiling of chikungunya virus infection in pediatric cases. Molecular Systems Biology *14*, 491.

Monneret, G., Lepape, A., Voirin, N., Bohé, J., Venet, F., Debard, A.-L., Thizy, H., Bienvenu, J., Gueyffier, F., and Vanhems, P. (2006). Persisting low monocyte human leukocyte antigen-DR expression predicts mortality in septic shock. Intensive Care Med. *32*, 1175–1183.

Morens, D.M., and Fauci, A.S. (2020). Emerging Pandemic Diseases: How We Got To COVID-19. Cell *182*, 1077–1092.

Nakamura, K., Ito, I., Kobayashi, M., Herndon, D.N., and Suzuki, F. (2015). Orosomucoid 1 drives opportunistic infections through the polarization of monocytes to the M2b phenotype. Cytokine *73*, 8–15.

Papathanassiu, A.E., Ko, J.-H., Imprialou, M., Bagnati, M., Srivastava, P.K., Vu, H.A., Cucchi, D., McAdoo, S.P., Ananieva, E.A., Mauro, C., and Behmoaras, J. (2017). BCAT1 controls metabolic reprogramming in activated human macrophages and is associated with inflammatory diseases. Nat. Commun. *8*, 16040–13.

Parnell, G.P., McLean, A.S., Booth, D.R., Armstrong, N.J., Nalos, M., Huang, S.J., Manak, J., Tang, W., Tam, O.-Y., Chan, S., et al. (2012). A distinct influenza infection signature in the blood transcriptome of patients with severe community-acquired pneumonia. Critical Care *16*, R157.

Parnell, G., McLean, A., Booth, D., Huang, S., Nalos, M., and Tang, B. (2011). Aberrant cell cycle and apoptotic changes characterise severe influenza A infection–a meta-analysis of genomic signatures in circulating leukocytes. PLoS One *6*, e17186.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

Ramilo, O., Allman, W., Chung, W., Meijas, A., Ardura, M., Glaser, C., Wittkowski, K.M., Piqueras, B., Banchereau, J., Palucka, A.K., et al. (2007). Gene expression patterns in blood leukocytes discriminate patients with acute infections. Blood *109*, 2066–2077.

Ramos-Martínez, E., López-Vancell, M.R., Fernández de Córdova-Aguirre, J.C., Rojas-Serrano, J., Chavarría, A., Velasco-Medina, A., and Velázquez-Sámano, G. (2018). Reduction of respiratory infections in asthma patients supplemented with vitamin D is related to increased serum IL-10 and IFNγ levels and cathelicidin expression. Cytokine *108*, 239–246.

Chowdhury, R., Vallania, F., Yang, Q., Lopez Angel, C.J., Darboe, F., Penn-Nicholson, A., Rozot, V., Nemes, E., Malherbe, S.T., Ronacher, K., et al. (2018). A multi-cohort study of the immune factors associated with M. tuberculosis infection outcomes. Nature *560*, 644–648.

Rodriguez-Fernandez, R., Tapia, L.I., Yang, C.-F., Torres, J.P., Chavez-Bueno, S., Garcia, C., Jaramillo, L.M., Moore-Clingenpeel, M., Jafri, H.S., Peeples, M.E., et al. (2017). Respiratory Syncytial Virus Genotypes, Host Immune Profiles, and Disease Severity in Young Children Hospitalized With Bronchiolitis. Journal of Infectious Diseases *217*, 24–34.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Schrijver, I.T., Théroude, C., and Roger, T. (2019). Myeloid-Derived Suppressor Cells in Sepsis. Front. Immunol. *10*, 327.

Schulte-Schrepping, J., Reusch, N., Paclik, D., Baßler, K., Schlickeiser, S., Zhang, B., Krämer, B., Krammer, T., Brumhard, S., Bonaguro, L., et al.; Deutsche COVID-19 OMICS Initiative (DeCOI) (2020). Severe COVID-19 is marked by a dysregulated myeloid cell compartment. Cell *182*, 1419–1440.e23.

Scott, M.K.D., Quinn, K., Li, Q., Carroll, R., Warsinske, H., Vallania, F., Chen, S., Carns, M.A., Aren, K., Sun, J., et al. (2019). Increased monocyte count as a cellular biomarker for poor outcomes in fibrotic diseases: a retrospective, multicentre cohort study. The Lancet Resp Med, 1–12.

Silvin, A., Chapuis, N., Dunsmore, G., Goubet, A.-G., Dubuisson, A., Derosa, L., Almire, C., Hénon, C., Kosmider, O., Droin, N., et al. (2020). Elevated calprotectin and abnormal myeloid cell subsets discriminate severe from mild COVID-19. Cell *182*, 1401–1418.e18.

Smith, C.L., Dickinson, P., Forster, T., Craigon, M., Ross, A., Khondoker, M.R., France, R., Ivens, A., Lynn, D.J., Orme, J., et al. (2014). Identification of a human neonatal immune-metabolic network associated with bacterial infection. Nature Communications *5*, 4649.

Soares-Schanoski, A., Baptista Cruz, N., de Castro-Jorge, L.A., de Carvalho, R.V.H., Santos, C.A.D., Rós, N.D., Oliveira, Ú., Costa, D.D., Santos, C.L.S.D., Cunha, M.D.P., et al. (2019). Systems analysis of subjects acutely infected with the Chikungunya virus. PLoS Pathogen *15*, e1007880.

Soneson, C., and Robinson, M.D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. Nature Publishing Group, 1–15.

Srivastava, A., Malik, L., Smith, T., Sudbery, I., and Patro, R. (2019). Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. Genome Biol. *20*, 65.

Stone, J.H., Frigault, M.J., Serling-Boyd, N.J., Fernandes, A.D., Harvey, L., Foulkes, A.S., Horick, N.K., Healy, B.C., Shah, R., Bensaci, A.M., et al. (2020). Efficacy of Tocilizumab in Patients Hospitalized with Covid-19. New Engl J Med *383*, 2333–2344.

Su, Y., Chen, D., Lausted, C., Yuan, D., Choi, J., Dai, C., Voillet, V., Scherler, K., Troisch, P., Duvvuri, V.R., et al. (2020). Multiomic Immunophenotyping of COVID-19 Patients Reveals Early Infection Trajectories. bioRxiv, 2020.07.27.224063.

Sweeney, T.E., Shidham, A., Wong, H.R., and Khatri, P. (2015). A comprehensive time-course–based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. Science Translational Medicine *7*, 287ra71–287ra71.

Sweeney, T.E., Braviak, L., Tato, C.M., and Khatri, P. (2016a). Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. Lancet Respir. Med. *4*, 213–224.

Sweeney, T.E., Wong, H.R., and Khatri, P. (2016b). Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. Science Translational Medicine *8*, 346ra91–346ra91.

Takizawa, H., Boettcher, S., and Manz, M.G. (2012). Demand-adapted regulation of early hematopoiesis in infection and inflammation. Blood *119*, 2991–3002.

Tang, B.M., Shojaei, M., Meyers, A., Ho, J., Ball, T.B., Keynan, Y., Pisipati, A., Kumar, A., Eisen, D.P., et al. (2019). Neutrophils-related host factors associated with severe disease and fatality in patients with influenza infection. Nature Communications *10*, 3422.

Thair, S.A., He, Y.D., Hasin-Brumshtein, Y., Sakaram, S., Pandya, R., Toh, J., Rawling, D., Remmel, M., Coyle, S., Dalekos, G.N., et al. (2020). Transcriptomic similarities and differences in host response between SARS-CoV-2 and other viral infections. iScience *24*, 101947.

Turner, C.T., Gupta, R.K., Tsaliki, E., Roe, J.K., Mondal, P., Nyawo, G.R., Palmer, Z., Miller, R.F., Reeve, B.W., Theron, G., and Noursadeghi, M. (2020). Blood transcriptional biomarkers for active pulmonary tuberculosis in a high-burden setting: a prospective, observational, diagnostic accuracy study. Lancet Respir. Med. *8*, 407–419.

Vallania, F., Tam, A., Lofgren, S., Schaffert, S., Azad, T.D., Bongen, E., Haynes, W., Alsup, M., Alonso, M., Davis, M., et al. (2018). Leveraging heterogeneity across multiple datasets increases cell-mixture deconvolution accuracy and reduces biological and technical biases. Nat. Commun. *9*, 4735.

Venet, F., Demaret, J., Gossez, M., and Monneret, G. (2020). Myeloid cells in sepsis-acquired immunodeficiency. Ann. N Y Acad. Sci.

Wang, Y., Dennehy, P.H., Keyserling, H.L., Tang, K., Gentsch, J.R., Glass, R.I., and Jiang, B. (2007). Rotavirus infection alters peripheral T-cell homeostasis in children with acute diarrhea. Jornal of Virology *81*, 3904–3912.

Warsinske, H.C., Rao, A.M., Moreira, F.M.F., Santos, P.C.P., Liu, A.B., Scott, M., Malherbe, S.T., Ronacher, K., Walzl, G., Winter, J., et al. (2018). Assessment of Validity of a Blood-Based 3-Gene Signature Score for Progression and Diagnosis of Tuberculosis, Disease Severity, and Treatment Response. JAMA Netw. Open *1*, e183779–e13.

Wei, F., Jiang, Z., Sun, H., Pu, J., Sun, Y., Wang, M., Tong, Q., Bi, Y., Ma, X., Gao, G.F., and Liu, J. (2019). Induction of PGRN by influenza virus inhibits the antiviral immune responses through downregulation of type I interferons signaling. PLoS Pathog. *15*, e1008062.

Wilk, A.J., Rustagi, A., Zhao, N.Q., Roque, J., Martínez-Colón, G.J., McKechnie, J.L., Ivison, G.T., Ranganath, T., Vergara, R., Hollis, T., et al. (2020). A single-cell atlas of the peripheral immune response in patients with severe COVID-19. Nat. Med. *26*, 1070–1076.

Wong, H.R., Cvijanovich, N.Z., Hall, M., Allen, G.L., Thomas, N.J., Freishtat, R.J., Anas, N., Meyer, K., Checchia, P.A., Lin, R., et al. (2012). Interleukin-27 is a novel candidate diagnostic biomarker for bacterial infection in critically ill children. Critical Care *16*, R213–R218.

Wong, H.R., Cvijanovich, N., Wheeler, D.S., Bigham, M.T., Monaco, M., Odoms, K., Macias, W.L., and Williams, M.D. (2008). Interleukin-8 as a stratification tool for interventional trials involving pediatric septic shock. American Journal of Respiratory and Critical Care Medicine *178*, 276–282.

Wong, H.R., Shanley, T.P., Sakthivel, B., Cvijanovich, N., Lin, R., Allen, G.L., Thomas, N.J., Doctor, A., Kalyanaraman, M., Tofil, N.M., et al. (2007). Genome-level expression profiles in pediatric septic shock indicate a role for altered zinc homeostasis in poor outcome. Physiological Genomics *30*, 146–155.

Wu, J., and Irizarry, R. (2020). gcrma: Background Adjustment Using Sequence Information. (Bioconductor).

Wu, T., Ji, Y., Moseman, E.A., Xu, H.C., Manglani, M., Kirby, M., Anderson, S.M., Handon, R., Kenyon, E., Elkahloun, A., et al. (2016). The TCF1-Bcl6 axis counteracts type I interferon to repress exhaustion and maintain T cell stemness. Science Immunology *1*, eaai8593–eaai8593.

Yu, J., Peterson, D.R., Baran, A.M., Bhattacharya, S., Wylie, T.N., Falsey, A.R., Mariani, T.J., and Storch, G.A. (2019). Host Gene Expression in Nose and Blood for the Diagnosis of Viral Respiratory Infection. Journal of Infectious Diseases *219*, 1151–1161.

Zaas, A.K., Chen, M., Varkey, J., Veldman, T., Hero, A.O., Lucas, J., Huang, Y., Turner, R., Gilbert, A., Lambkin-WIlliams, R., et al. (2009). Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. Cell Host & Microbe *6*, 207–217.

Zhai, Y., Franco, L.M., Atmar, R.L., Quarles, J.M., Arden, N., Bucasas, K.L., Wells, J.M., Nino, D., Wang, X., Zapata, G.E., et al. (2015). Host transcriptional response to influenza and other acute respiratory viral infections - a prospective cohort study. PLoS Pathogen *11*, e1004869.

Zhao, X., Li, J., Winkler, C.A., An, P., and Guo, J.-T. (2019). IFITM Genes, Variants, and Their Roles in the Control and Pathogenesis of Viral Infections. Front. Microbiol. *9*, 3228.

Zhou, Z., Ren, L., Zhang, L., Zhong, J., Xiao, Y., Jia, Z., Guo, L., Yang, J., Wang, C., Jiang, S., et al. (2020). Heightened Innate Immune Responses in the Respiratory Tract of COVID-19 Patients. Cell Host Microbe *27*, 883–890.e2.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
RNA-seq dataset	Liu et al., 2017	accession# PRJNA352396
RNA-seq dataset	Soares-Schanoski et al., 2019	accession# PRJNA507472
RNA-seq dataset	Michlmayr et al., 2018	accession# PRJNA390289
RNA-seq dataset	Thair et al., 2020	accession# GSE152641
Microarray dataset	de Steenhuijsen Piters et al., 2016	accession# GSE77087
Microarray dataset	Liu et al., 2016	accession# GSE73072
Microarray dataset	Zhai et al., 2015	accession# GSE68310
Microarray dataset	Jaggi et al., 2018	accession# GSE68004
Microarray dataset	Heinonen et al., 2016	accession# GSE67059
Microarray dataset	Sweeney et al., 2015	accession# GSE66099
Microarray dataset	Wong et al., 2007, 2008, 2012	accession# GSE4607
Microarray dataset	Ramilo et al., 2007	accession# GSE6269
Microarray dataset	Hoang et al., 2014	accession# GSE61821
Microarray dataset	Davenport et al., 2014	accession# GSE61754
Microarray dataset	Parnell et al., 2012	accession# GSE40012
Microarray dataset	Mejias et al., 2013	accession# GSE38900
Microarray dataset	Wang et al., 2007	accession# GSE2729
Microarray dataset	Berdal et al., 2011	accession# GSE27131
Microarray dataset	Smith et al., 2014	accession# GSE25504
Microarray dataset	Bermejo-Martin et al., 2010	accession# GSE21802
Microarray dataset	Parnell et al., 2011	accession# GSE20346
Microarray dataset	Zaas et al., 2009	accession# GSE17156
Microarray dataset	Yu et al., 2019	accession# GSE117827
Microarray dataset	Dunning et al., 2018	accession# GSE111368
Microarray dataset	Rodriguez-Fernandez et al., 2017	accession# GSE103842
Microarray dataset	Tang et al., 2019	accession# GSE101702
Microarray dataset	Jong et al., 2016	accession# E-MTAB-5195
scRNA-seq dataset	Su et al., 2020	accession# E-MTAB-9357
scRNA-seq dataset	Arunachalam et al., 2020	accession# GSE155673
scRNA-seq dataset	Wilk et al., 2020	accession# GSE150728
Software and Algorithms		
R	R Core Team (2020)	www.r-project.org
gcRMA	Wu and Irizarry, 2020	https://www.bioconductor.org/packages/release/bioc/html/gcrma.html
MetaIntegrator	Haynes et al., 2017	https://cran.r-project.org/web/packages/MetaIntegrator/index.html
COCONUT	Sweeney et al., 2016b	https://cran.r-project.org/web/packages/COCONUT/index.html
Qualimap	García-Alcalde et al., 2012	http://qualimap.conesalab.org/
Salmon	Patro et al., 2017	https://combine-lab.github.io/salmon/
Tximport	Soneson and Robinson, 2018	https://bioconductor.org/packages/release/bioc/html/tximport.html
DESeq2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Alevin	Srivastava et al., 2019	https://combine-lab.github.io/salmon/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
UMAP	McInnes et al., 2018	https://cran.r-project.org/web/packages/umap/index.html
Seurat	Satija et al., 2015	https://satijalab.org/seurat/
Trim Galore	Martiin, 2011	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
SingleR	Aran et al., 2019	https://bioconductor.org/packages/release/bioc/html/SingleR.html
STAR	Dobin et al., 2013	https://github.com/alexdobin/STAR
immunoStates	Vallania et al., 2018	https://cran.r-project.org/web/packages/MetaIntegrator/index.html
tSpace	Dermadi et al., 2020	https://github.com/hylasD/tSpace

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources, software, and data should be directed to and will be fulfilled by the Lead Contact, Purvesh Khatri (pkhatri@stanford.edu).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

This study did not generate any unique datasets or code. All datasets, software, and algorithms used in this study are publicly available and listed in the Key Resource table.

QUANTIFICATION AND STATISTICAL ANALYSIS

Dataset collection and preprocessing

We downloaded 26 gene expression datasets from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO), Sequence Read Archive (SRA), ArrayExpress, and European Nucleotide Archive (ENA), consisting of 4,780 samples from 34 independent cohorts derived from whole blood or peripheral blood mononuclear cells (PBMCs) (Table S1). We excluded all datasets used to discover the MVS previously to ensure all cohorts analyzed in the current study were independent. We defined a cohort as a comparable group of individuals within a dataset, where each dataset has a unique GEO identifier and may contain multiple cohorts. For example, the dataset GSE73072 contains seven cohorts of individuals challenged with one of three viruses. The samples in these datasets represented the biological and clinical heterogeneity observed in the real-world patient population, including HCs and patients infected with 16 different viruses with severity ranging from asymptomatic to fatal viral infection over a broad age range (0-90 years) (Figure 1A and Table S1). The samples were from patients enrolled across 18 different countries representing diverse genetic backgrounds of patients and viruses. We included technical heterogeneity in our analysis as these datasets were profiled using microarray and RNA sequencing (RNA-seq) from different manufacturers.

We renormalized all microarray datasets using standard methods when raw data were available from the GEO. We applied GC robust multiarray average (gcRMA) to arrays with mismatch probes for Affymetrix arrays. We used normal-exponential background correction and quantile normalization for Illumina, Agilent, GE, and other commercial arrays. We did not renormalize custom arrays and used preprocessed data as made publicly available by the study authors. We mapped microarray probes in each dataset to Entrez Gene identifiers (IDs) to facilitate integrated analysis. If a probe matched more than one gene, we expanded the expression data for that probe to add one record for each gene. When multiple probes mapped to the same gene within a dataset, we applied a fixed-effect model. Within a dataset, cohorts assayed with different microarray types were treated as independent.

Standardized severity assignment

For each dataset, we used the sample phenotypes as defined in the original publication. We manually assigned a severity category to each sample based on the cohort description for each dataset in the original publication as follows: (1) HCs – asymptomatic, uninfected healthy individuals, (2) asymptomatic or convalescents – afebrile asymptomatic individuals who tested positive for a virus or those fully recovered from a viral infection with completely resolved symptoms, (3) mild – symptomatic individuals with viral infection that were either managed as outpatient or discharged from the emergency department (ED), (4) moderate – symptomatic individuals with viral infection who were admitted to the general wards and did not require supplemental oxygen, (5) serious - symptomatic individuals with viral infection who were described as ‘severe’ by original authors, admitted to general wards with supplemental

oxygen, or admitted to the intensive care unit (ICU) without requiring mechanical ventilation or inotropic support, (6) critical - symptomatic individuals with viral infection who were on mechanical ventilation in the ICU or were diagnosed with acute respiratory distress syndrome (ARDS), septic shock, or multiorgan dysfunction syndrome (MODS), and (7) fatal – patients with viral infection who died in the ICU.

For datasets that did not provide sample-level severity data (GSE101702, GSE38900, GSE103842, GSE66099, GSE77087), we assigned severity categories as follows. We categorized all samples in a dataset as ''moderate'' when either (1) > 70% of patients were admitted to the general wards as opposed to discharged from the ED, (2) < 20% of patients admitted to the general wards required supplemental oxygen, or (3) patients were admitted to the general wards and categorized as 'mild' or 'moderate' by the original authors. We categorized all samples in a dataset as ''severe'' when > 20% of patients had either (1) been admitted to the general wards and categorized as 'severe' by original authors, (2) required supplemental oxygen, or (3) required ICU admission without mechanical ventilation.

Viral challenge studies

GSE73072 included seven viral challenge studies that determined the infection status of a subject through reverse transcription PCR (RT-PCR) for a given virus (H1N1, H3N2, RSV, HRV) in longitudinally collected nasopharyngeal samples. In these studies, we assigned all baseline pre-challenge samples and subjects who never shed virus, as determined by RT-PCR, to the 'healthy' category. We assigned samples from infected subjects, defined as those who had virus detected in any of their nasopharyngeal samples, to one of three categories: (1) before infection - blood samples collected after challenge but before a virus was detected in a nasopharyngeal sample, (2) after infection - blood samples collected after the last nasopharyngeal sample in which a virus was detected, and (3) during infection - blood samples collected between the first and last nasopharyngeal sample in which a virus was detected.

COCONUT co-normalization

We used Combat CONormalization Using conTrols (COCONUT) for between-dataset normalization (Sweeney et al., 2016b). COCONUT allows for co-normalization of gene expression data without bias toward sample diagnosis by applying a modified version of the ComBat empirical Bayes normalization method (Johnson et al., 2007), which assumes a similar distribution between control samples. Briefly, healthy samples from each cohort undergo ComBat co-normalization without covariates, and the ComBat estimated parameters are computed for the healthy samples in each dataset. By applying these parameters to the non-healthy samples, all datasets keep the same background distribution while retaining the same relative distance between healthy and disease samples, which preserves the biological variability between the two groups within a dataset. We have previously shown that when COCONUT co-normalization is applied, housekeeping genes remain invariant across both conditions and cohorts, and each gene retains the same distribution across conditions within each dataset (Sweeney et al., 2016b).

MVS genes and score

We did not derive a *de novo* gene signature to represent the conserved host response to viral infection. Instead, we used our previously described 396-gene signature from peripheral blood (Andres-Terre et al., 2015). As previously described, we defined the MVS score of a sample as the difference between the geometric mean of the overexpressed genes and the geometric mean of the under-expressed genes in the MVS (Andres-Terre et al., 2015). Out of 396 genes in the MVS, 251 genes (111 over- and 140 under-expressed) were measured across all datasets. We used the 251 gene subset of the MVS in our analyses as the MVS score using either 251-gene or 396-gene signatures were highly correlated (data not shown).

We measured the correlation of the MVS score with viral infection severity using Spearman's rank correlation coefficient. We used the Mann–Whitney U test (Wilcoxon rank-sum test) to compare MVS scores between two groups. We tested the trend of the MVS score along viral infection severity categories using the Jonckheere-Terpstra trend test.

RNA sequencing analysis

We obtained the raw reads for the Ebola (PRJNA352396) and chikungunya (PRJNA507472 and PRJNA390289) cohorts from the European Nucleotide Archive (ENA). We obtained the raw reads of the SARS-CoV-2 cohort from Inflammatix. We assessed trimmed Illumina adaptors and removed reads that were too short after adaptor trimming (less than 20 nt) with Trim Galore (v0.6.5). We then mapped the cleaned reads to the human genome (hg38) using STAR (v2.7.3) (Dobin et al., 2013). We performed additional quality control by checking the mapped reads with Qualimap (v.2.2.2) (García-Alcalde et al., 2012). To quantify gene expression, we obtained human transcriptome sequences from GENCODE site (v32), processed the cleaned reads with Salmon (1.2.1) (Patro et al., 2017) to get transcript-level expression, and summarized to gene-level expression using Tximport (v1.16.0) (Soneson and Robinson, 2018). Finally, we applied the variance stabilizing transformation from DESeq2 (v1.26.0) (Love et al., 2014) to normalize gene expression for downstream analysis and visualization.

Detection of viral reads in RNA-seq data

We obtained genome sequences of 501 human viruses from the NCBI virus database (accessed on April 19, 2020). We concatenated viral sequences with the list human transcriptome sequences and built a decoy-aware index using Salmon. We mapped the reads to the concatenated index using Salmon with the selective-alignment algorithm, which, together with the decoy-aware index, mitigates potential spurious mapping of reads arising from unannotated human genomic loci and reduces false positives. We extracted reads

mapped to viral genomes and filtered them to remove secondary alignments and paired-end reads with only one mate mapped. We also checked the reads with NCBI Nucleotide BLAST to ensure viral origin. We normalized the viral read counts by the total number of sequencing reads of each sample. We measured the correlation between the MVS score and viral read counts using Pearson correlation coefficient.

Analysis of single-cell RNA-seq data

We downloaded scRNA-seq data of the Stanford cohort from NCBI GEO (Wilk et al., 2020). We also obtained the scRNA-seq data for the Atlanta cohort (Arunachalam et al., 2020) and the Seattle cohort (Su et al., 2020). We processed raw scRNA-seq data with Alevin (v1.2.1) (Srivastava et al., 2019) to get the read count matrices. We performed quality control, normalization, dimension reduction, UMAP projection, and Shared Nearest Neighbors clustering on the three datasets with Seurat (Satija et al., 2015). Then we applied the Seurat integration workflow to integrate the three datasets using Reciprocal PCA analysis. Cell type was annotated with SingleR inference (Aran et al., 2019), cell type markers, and annotations from the original publications.

In silico cellular deconvolution using immunoStates and multi-cohort analysis of estimated cellular proportions

We performed *in silico* cellular deconvolution using immunoStates as a basis matrix with support vector regression to estimate proportions of 25 immune cell subsets in each sample (Vallania et al., 2018). To investigate changes in the immune cell proportions between patients with different severity of viral infection, we conducted three multi-cohort analyses using MetaIntegrator R package (Haynes et al., 2017) between samples from the following categories: 1) subjects with non-severe viral infection (severity categories 'mild' and 'moderate') versus HCs, 2) subjects with severe viral infection (severity categories 'serious', 'critical', and 'fatal') versus HCs, and 3) subjects with severe viral infection versus subjects with non-severe viral infection. We combined effect sizes across studies using a random-effects inverse variance model. For each meta-analysis, we calculated the change in proportions for each immune cell type between groups in each cohort as the Hedges' g effect size (ES). We corrected p values for multiple hypotheses testing using the Benjamini-Hochberg correction to obtain the false discovery rate (FDR). We used a threshold of FDR < 20% and representation in a minimum of 5 studies in conjunction with leave-one-out analysis to identify immune cell types with increased or decreased proportions between groups. Individual samples that met the following criteria were excluded: non-viral infection, non- HCs, and one sample from PRJNA252396 (SRR4888654) which had the same expression value for all 317 genes. Datasets with less than two samples in each of the compared groups were excluded from meta-analysis. Individual samples that met the following criteria were excluded: non-viral infection, non-healthy control, and one sample from PRJNA252396 (SRR4888654) which had the same expression value for all 317 genes. Datasets with less than two samples in each of the compared groups were excluded from meta-analysis.

Trajectory inference analysis

We co-normalized 1674 samples from 21 cohorts in 19 datasets with 1509 samples from four independent challenge studies using COCONUT. Each challenge study inoculated healthy volunteers with one of four viruses (HRV, RSV, H1N1, and H3N2). We adapted tSpace, a method for identifying cellular differentiation trajectories using scRNA-seq data (Dermadi et al., 2020), to identify disease trajectories using bulk transcriptome profiles. We refer to the adaption to bulk transcriptome data as disease space (dSpace), although the core method remains identical to tSpace. The tSpace algorithm has three steps: (1) calculation of a set of sub-graphs, (2) calculation of the trajectory space matrix across the sub-graphs, and (3) visualization. In the first step, we calculated a set of sub-graphs keeping L out of K nearest neighbors in a KNN graph. The user defines the number of sub-graphs (G), neighborhood size (K), and how many nearest neighbors are preserved in the sub-graphs (L). The second step computes a *trajectory space* distance matrix using a modified Dijkstra algorithm that implements waypoints (WP) to exponentially weigh and refine distances. The final *trajectory space* matrix is a dense matrix in which each sample is a row, and calculated trajectories are columns. The number of trajectories (T > 150) is user-defined and very robust across a wide dynamic range. Finally, we visualize the samples and their relationships in trajectory space using PCA or UMAP.

We used the following parameters for the dSpace analysis: G = 5, K = 65, L = 49, T = 500, WP = 20. We used Pearson correlation as the metric for computing distance between two samples. We fitted a principal line through data visualized in the first two components of tSpace (tPC1, tPC2) using the princurve R package. Princurve calculates lambda, an arc length distance for each data point, which we used to align subjects along the isolated trajectory. Furthermore, the covariance matrix of the transposed trajectory matrix (covariance mapping) coupled with the hierarchical clustering identified clusters of patients with shared trajectory space. The covariance matrix of the transposed trajectory matrix allows identification of patients that belong to diverging trajectories, and hierarchical clustering of covariance matrix allowed us to group patients that are in severe and non-severe branches, thus enabling isolation of both branches. Each of the determined clusters is a reflection of the position of patients in the trajectory space. Hierarchical clustering was calculated using hclust and Dist R functions with ''euclidean'' and ''complete'' parameters.

Severe and non-severe trajectories shared 1020 samples from the healthy and no symptoms categories. Therefore, we aligned them using dynamic time warping (dtw R package) and split them into 4 stages. All 251 genes and the fitted trajectory (lambda value) were used for alignment. We applied a permutation test (Efron and Tibshirani, 2002) for each of the 4 stages and identified 96 genes that were differentially expressed within the same stage between the two severity branches.

Statistical analysis of trajectories identified with dSpace

We applied a permutation test (Efron and Tibshirani, 2002) for each of the 4 stages and identified 96 genes that were differentially expressed within the same stage between the two severity branches. In our testing we used 1000 permutations, and for significance FDR < 0.001 and |effects size| > 0.3. We performed data analysis using R.

Calculation of the SoM score

The Severe or Mild (SoM) score is a 42-gene model that utilizes the expression of genes from the 4 gene modules to distinguish between severe and mild viral infection. For each sample, we compute the geometric mean of the expression of genes from each module. Then, we calculate a score by taking the sum of the geometric means of modules 1 and 2 and dividing that by the sum of the geometric means of modules 3 and 4, as shown in the following equation:

$$SoM\ score = \frac{\left(\prod_{gene\ \varepsilon\ Module1} x_i(gene)\right)^{\frac{1}{||Module1||}} + \left(\prod_{gene\ \varepsilon\ Module2} x_i(gene)\right)^{\frac{1}{||Module2||}}}{\left(\prod_{gene\ \varepsilon\ Module3} x_i(gene)\right)^{\frac{1}{||Module3||}} + \left(\prod_{gene\ \varepsilon\ Module4} x_i(gene)\right)^{\frac{1}{||Module4||}}} \quad \text{(Equation1)}$$

Correlation of cell proportions, MVS score, and SoM score with severity

We measured the correlation of the cell proportions or the score with viral infection severity using Spearman's rank correlation coefficient. We used the Mann–Whitney U test (Wilcoxon rank-sum test) to compare the cell proportions or the score between two groups. We tested the trend of the cell proportions or the score along viral infection severity categories using the Jonckheere-Terpstra trend test. We measured the correlation between the 251 and 396 gene versions of the MVS using Pearson's correlation coefficient. Data analysis was performed using R.

Module score association with plasma proteomics

The association between modules scores and plasma proteomics data from the Seattle cohort were calculated using multivariate generalized estimating equations to account for repeated samples per patient and adjust for covariates (age, sex, BMI, race/ethnicity). The p values were adjusted using Bonferroni correction. Data analysis was performed using R.

Immunity, Volume *54*

Supplemental information

Multi-cohort analysis of host immune response identifies conserved protective and detrimental modules associated with severity across viruses

Hong Zheng, Aditya M. Rao, Denis Dermadi, Jiaying Toh, Lara Murphy Jones, Michele Donato, Yiran Liu, Yapeng Su, Cheng L. Dai, Sergey A. Kornilov, Minas Karagiannis, Theodoros Marantos, Yehudit Hasin-Brumshtein, Yudong D. He, Evangelos J. Giamarellos-Bourboulis, James R. Heath, and Purvesh Khatri

Supplementary figure legends

Figure S1, refers to Figure 1: MVS score is associated with severity of viral infection in each dataset in discovery and validation cohorts. a) ROC curves for the MVS score in each dataset in the discovery cohorts. AUROC values varied from 0.859 (95% CI 0.69-1) to 1 (95% CI 1-1). **b)** ROC curves for the MVS score in 4 independent cohorts profiled using RNAseq. AUROC values varied from 0.84 (95% CI 0.76-0.92) to 0.972 (95% CI 0.932-1). **c**) Violin plots of the MVS score for all samples across 19 datasets.

Figure S2, refers to Figure 2: MVS score is predominantly from the myeloid cells and Neutrophil proportion is higher in patients with severe viral infection. a-c) UMAP visualization of cell types, severity and MVS score in each of the three scRNA-seq cohorts. **d-f)** Forest plots for the changes in proportions of neutrophils estimated from bulk transcriptomic profiles using deconvolution between d) patients with non-severe viral infection and healthy controls, e) patients with severe viral infection and healthy controls, and f) patients with severe and non-severe viral infection.

Figure S3, refers to Figure 4: Samples from viral challenge studies almost exclusively cluster with samples from non-severe viral infection within dSpace. a) dSpace trajectory analysis of 3,183 samples from 25 cohorts, including 1,509 samples from 4 viral challenge studies (2 influenza, 1 HRV, 1 RSV). Each point represents a sample; viral challenge study samples are demarcated as triangles. **b)** Proportion of samples for each severity category or challenge study group within each cluster.

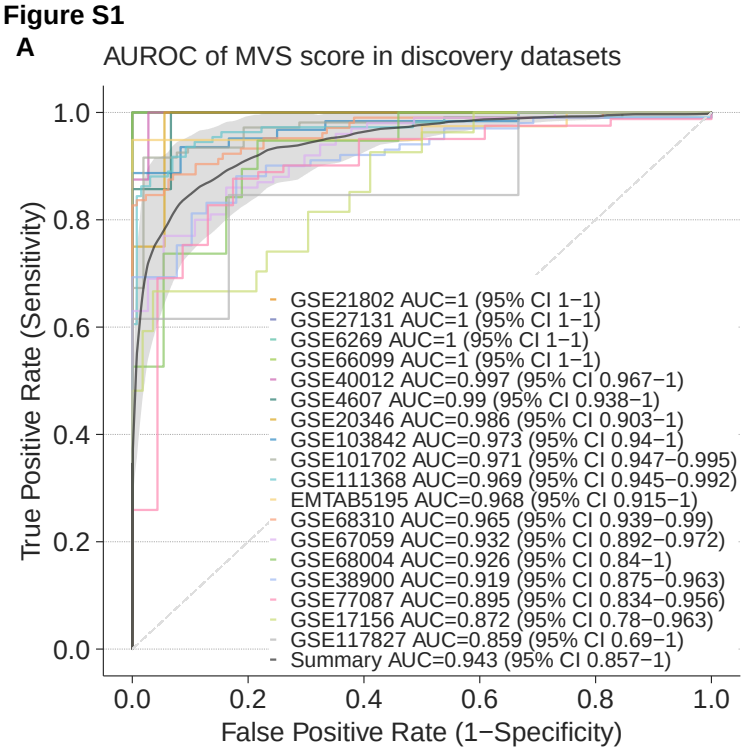
Figure S4, refers to Figure 5: Expression of genes associated with NK cells, myeloid cell-derived suppression, HSPCs, and an overall protective host response correlate with severity of viral infection. **a-b)** *KLRD1* and *PIK3R1* expression in **a)** the discovery cohort and **b)** SARS-CoV-2 infection. **c-d)** Expression of myeloid cell-associated genes, including MDSC markers and *ORM1* in **c)** the discovery cohort and **d)** SARS-CoV-2 infection. **e-f)** Expression of genes over-expressed in patients with severe viral infection but not in those with non-severe viral infections compared to healthy controls, and preferentially expressed in circulating HSPCs in **e)** the discovery cohort and **f)** SARS-CoV-2 infection. **g-h)** Expression of genes identified to be significantly higher in patients with mild viral infection compared to those with serious, critical, or fatal viral infection, or healthy controls, in **g)** the discovery cohort and **h)** SARS-CoV-2 infection.

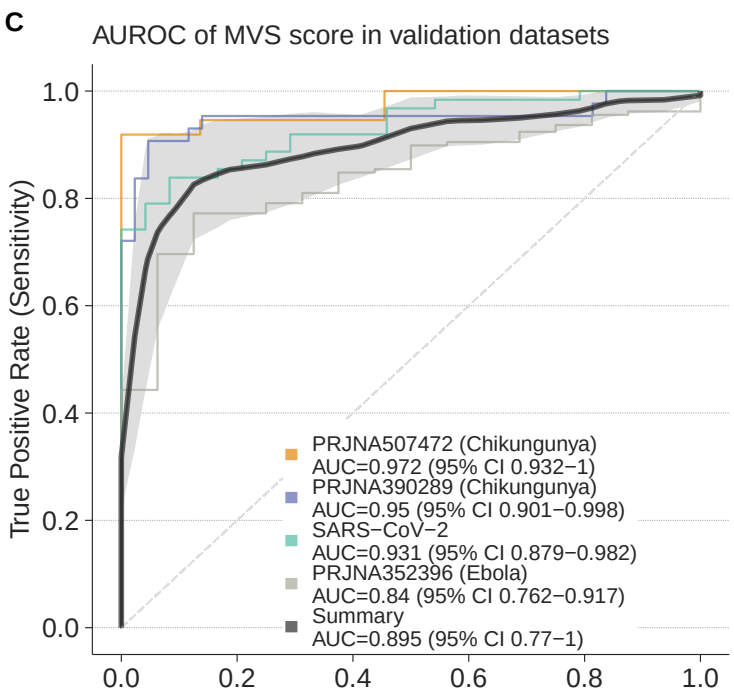
Figure S5, refers to Figure 5: Expression of *HLA-DPB1* in bulk transcriptome and sc-RNAseq. a) Expression of *HLA-DPB1* in 3,183 bulk transcriptome samples in the discovery cohort. **b**) Expression of *HLA-DPB1* in SARS-CoV-2 cohort. **c**) UMAP visualization of the expression of *HLA-DPB1* in three sc-RNAseq cohorts. UMAP of cell type and severity are shown in Figure 2b and 2d. d) Expression of *HLA-DPB1* in CD14 monocyte, cDCs, and B cells in the three sc-RNAseq cohorts.

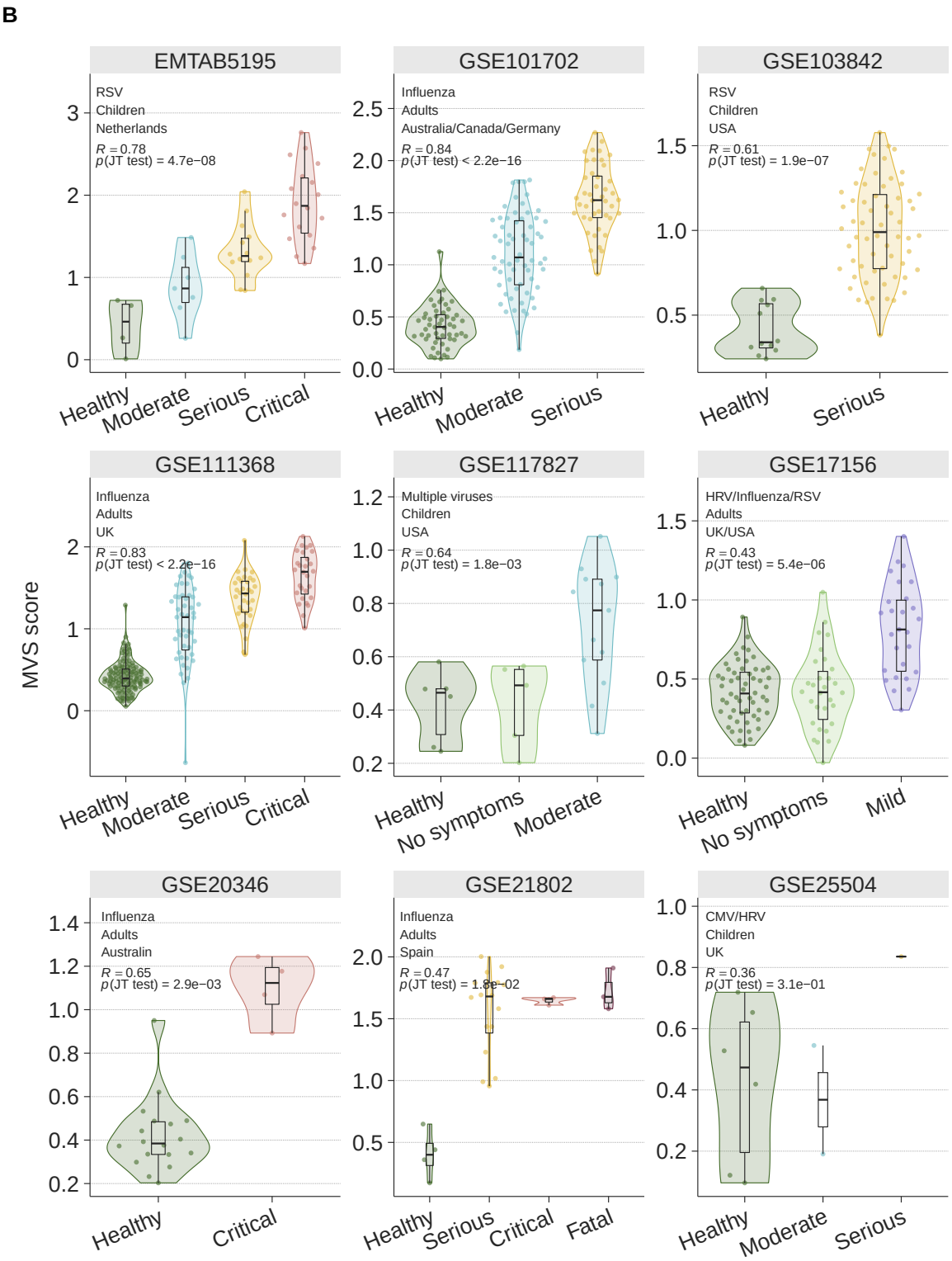
Figure S6, refers to Figure 6: The interferon-induced genes (*IFITM1*, *IFITM2*, *IFITM3)* and type I and II interferon receptors are highly correlated with protective response genes in mild but not severe viral infection. **a)** Expression of *IFITM1*, *IFITM2*, *IFITM3,* and type I and II interferon receptors across severity categories in patients with different viral infections including SARS-CoV-2. Each point in the violin plots represents a sample. **b)** Boxplots representing the correlation between IFITMs and type I and II interferon receptors, and the genes belonging to the protective response module. Each point represents a correlation between a gene pair, and lines connect the same pair across severity categories. P-values for the comparison between severity categories were computed using Wilcoxon signed-rank test.

Figure S7, refers to Figure 7: SoM score distinguishes mild and severe viral infection in discovery and validation cohorts with higher accuracy than the MVS score. a-b) The ROC curves of the MVS score for differentiating between severity categories of viral infection in **a)** discovery and **b)** validation cohorts. **c-d**) The ROC curves of the SoM score for differentiating between severity categories of viral infection in **c)** discovery and **d)** validation cohorts.

Figure S1


A
AUROC of MVS score in discovery datasets
True Positive Rate (Sensitivity)
False Positive Rate (1−Specificity)
GSE21802 AUC=1 (95% CI 1−1)
GSE27131 AUC=1 (95% CI 1−1)
GSE6269 AUC=1 (95% CI 1−1)
GSE66099 AUC=1 (95% CI 1−1)
GSE40012 AUC=0.997 (95% CI 0.967−1)
GSE4607 AUC=0.99 (95% CI 0.938−1)
GSE20346 AUC=0.986 (95% CI 0.903−1)
GSE103842 AUC=0.973 (95% CI 0.94−1)
GSE101702 AUC=0.971 (95% CI 0.947−0.995)
GSE111368 AUC=0.969 (95% CI 0.945−0.992)
EMTAB5195 AUC=0.968 (95% CI 0.915−1)
GSE68310 AUC=0.965 (95% CI 0.939−0.99)
GSE67059 AUC=0.932 (95% CI 0.892−0.972)
GSE68004 AUC=0.926 (95% CI 0.84−1)
GSE38900 AUC=0.919 (95% CI 0.875−0.963)
GSE77087 AUC=0.895 (95% CI 0.834−0.956)
GSE17156 AUC=0.872 (95% CI 0.78−0.963)
GSE117827 AUC=0.859 (95% CI 0.69−1)
Summary AUC=0.943 (95% CI 0.857−1)


C
AUROC of MVS score in validation datasets
True Positive Rate (Sensitivity)
False Positive Rate (1−Specificity)
PRJNA507472 (Chikungunya)
AUC=0.972 (95% CI 0.932−1)
PRJNA390289 (Chikungunya)
AUC=0.95 (95% CI 0.901−0.998)
SARS−CoV−2
AUC=0.931 (95% CI 0.879−0.982)
PRJNA352396 (Ebola)
AUC=0.84 (95% CI 0.762−0.917)
Summary
AUC=0.895 (95% CI 0.77−1)


B
MVS score
EMTAB5195
RSV
Children
Netherlands
R = 0.78
p(JT test) = 4.7e−08
Healthy
Moderate
Serious
Critical
GSE101702
Influenza
Adults
Australia/Canada/Germany
R = 0.84
p(JT test) < 2.2e−16
Healthy
Moderate
Serious
GSE103842
RSV
Children
USA
R = 0.61
p(JT test) = 1.9e−07
Healthy
Serious
GSE111368
Influenza
Adults
UK
R = 0.83
p(JT test) < 2.2e−16
Healthy
Moderate
Serious
Critical
GSE117827
Multiple viruses
Children
USA
R = 0.64
p(JT test) = 1.8e−03
Healthy
No symptoms
Moderate
GSE17156
HRV/Influenza/RSV
Adults
UK/USA
R = 0.43
p(JT test) = 5.4e−06
Healthy
No symptoms
Mild
GSE20346
Influenza
Adults
Australin
R = 0.65
p(JT test) = 2.9e−03
Healthy
Critical
GSE21802
Influenza
Adults
Spain
R = 0.47
p(JT test) = 1.8e−02
Healthy
Serious
Critical
Fatal
GSE25504
CMV/HRV
Children
UK
R = 0.36
p(JT test) = 3.1e−01
Healthy
Moderate
Serious

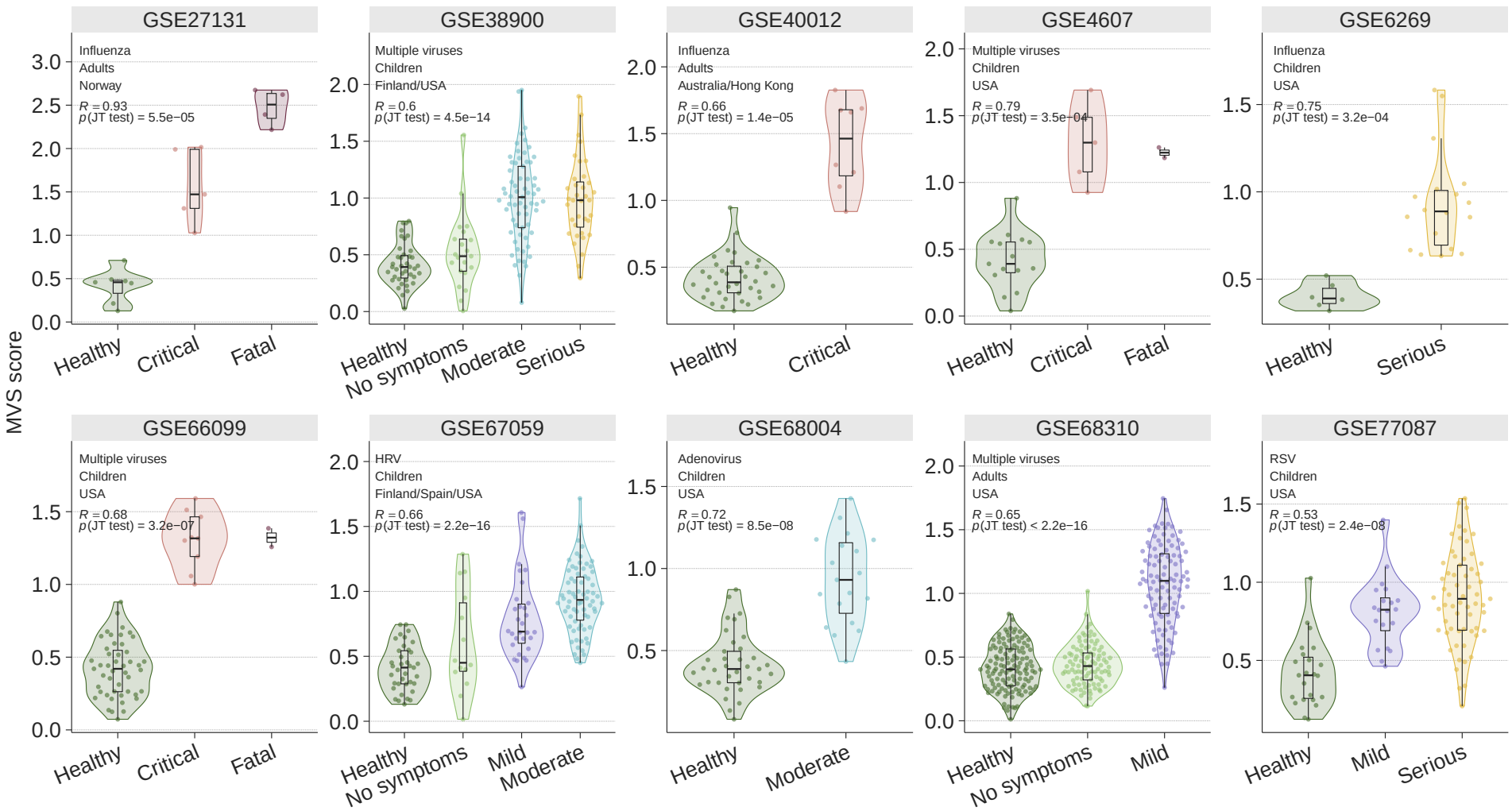
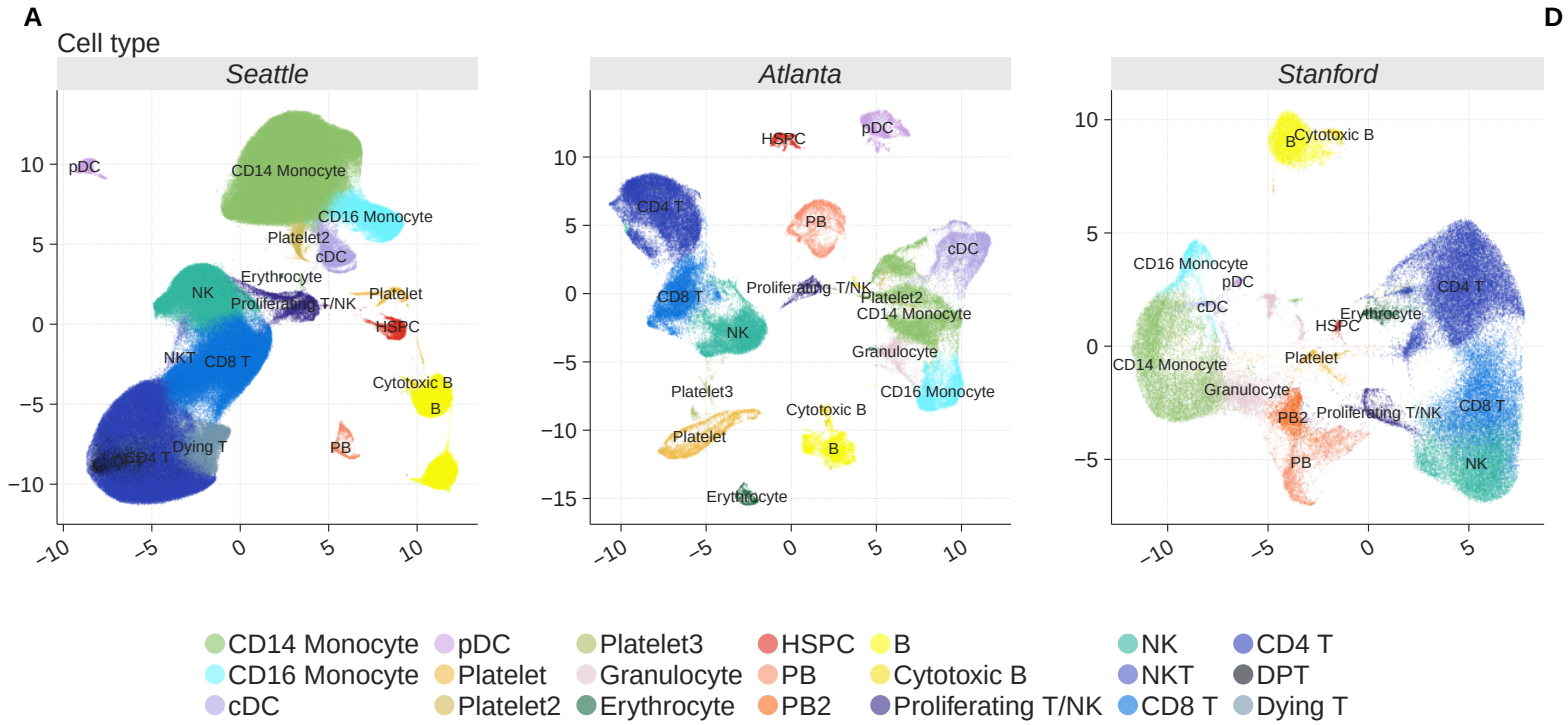

MVS score
GSE27131
Influenza
Adults
Norway
R = 0.93
p(JT test) = 5.5e−05
Healthy
Critical
Fatal
GSE38900
Multiple viruses
Children
Finland/USA
R = 0.6
p(JT test) = 4.5e−14
Healthy
No symptoms
Moderate
Serious
GSE40012
Influenza
Adults
Australia/Hong Kong
R = 0.66
p(JT test) = 1.4e−05
Healthy
Critical
GSE4607
Multiple viruses
Children
USA
R = 0.79
p(JT test) = 3.5e−04
Healthy
Critical
Fatal
GSE6269
Influenza
Children
USA
R = 0.75
p(JT test) = 3.2e−04
Healthy
Serious
GSE66099
Multiple viruses
Children
USA
R = 0.68
p(JT test) = 3.2e−07
Healthy
Critical
Fatal
GSE67059
HRV
Children
Finland/Spain/USA
R = 0.66
p(JT test) = 2.2e−16
Healthy
No symptoms
Mild
Moderate
GSE68004
Adenovirus
Children
USA
R = 0.72
p(JT test) = 8.5e−08
Healthy
Moderate
GSE68310
Multiple viruses
Adults
USA
R = 0.65
p(JT test) < 2.2e−16
Healthy
No symptoms
Mild
GSE77087
RSV
Children
USA
R = 0.53
p(JT test) = 2.4e−08
Healthy
Mild
Serious

Figure S2


A
Cell type
Seattle
Atlanta
Stanford
pDC
CD14 Monocyte
CD16 Monocyte
Platelet2
cDC
Erythrocyte
NK
Proliferating T/NK
Platelet
HSPC
NKT
CD8 T
Cytotoxic B
B
Dying T
CD4 T
PB
Platelet3
Granulocyte
PB2
DPT
B
Severity
Healthy
No symptoms
Mild
Moderate
Serious
Critical
Fatal
C
MVS score
low
high

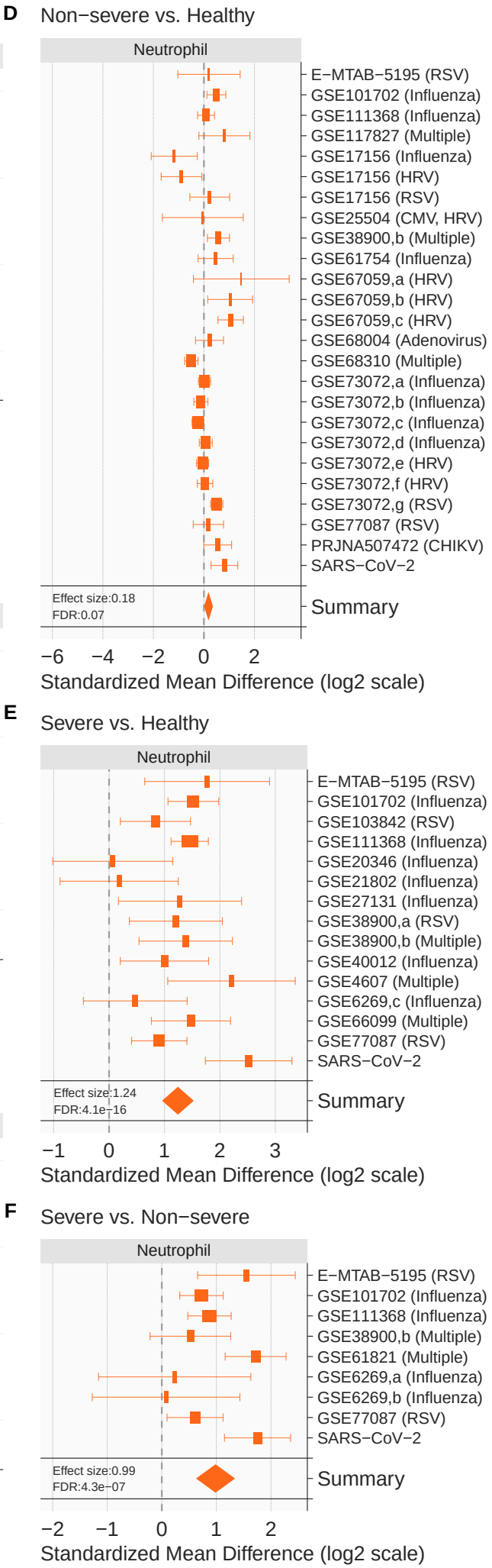

D
Non−severe vs. Healthy
Neutrophil
E−MTAB−5195 (RSV)
GSE101702 (Influenza)
GSE111368 (Influenza)
GSE117827 (Multiple)
GSE17156 (Influenza)
GSE17156 (HRV)
GSE17156 (RSV)
GSE25504 (CMV, HRV)
GSE38900,b (Multiple)
GSE61754 (Influenza)
GSE67059,a (HRV)
GSE67059,b (HRV)
GSE67059,c (HRV)
GSE68004 (Adenovirus)
GSE68310 (Multiple)
GSE73072,a (Influenza)
GSE73072,b (Influenza)
GSE73072,c (Influenza)
GSE73072,d (Influenza)
GSE73072,e (HRV)
GSE73072,f (HRV)
GSE73072,g (RSV)
GSE77087 (RSV)
PRJNA507472 (CHIKV)
SARS−CoV−2
Effect size:0.18
FDR:0.07
Summary
Standardized Mean Difference (log2 scale)
E
Severe vs. Healthy
Neutrophil
E−MTAB−5195 (RSV)
GSE101702 (Influenza)
GSE103842 (RSV)
GSE111368 (Influenza)
GSE20346 (Influenza)
GSE21802 (Influenza)
GSE27131 (Influenza)
GSE38900,a (RSV)
GSE38900,b (Multiple)
GSE40012 (Influenza)
GSE4607 (Multiple)
GSE6269,c (Influenza)
GSE66099 (Multiple)
GSE77087 (RSV)
SARS−CoV−2
Effect size:1.24
FDR:4.1e−16
Summary
Standardized Mean Difference (log2 scale)
F
Severe vs. Non−severe
Neutrophil
E−MTAB−5195 (RSV)
GSE101702 (Influenza)
GSE111368 (Influenza)
GSE38900,b (Multiple)
GSE61821 (Multiple)
GSE6269,a (Influenza)
GSE6269,b (Influenza)
GSE77087 (RSV)
SARS−CoV−2
Effect size:0.99
FDR:4.3e−07
Summary
Standardized Mean Difference (log2 scale)

Figure S3

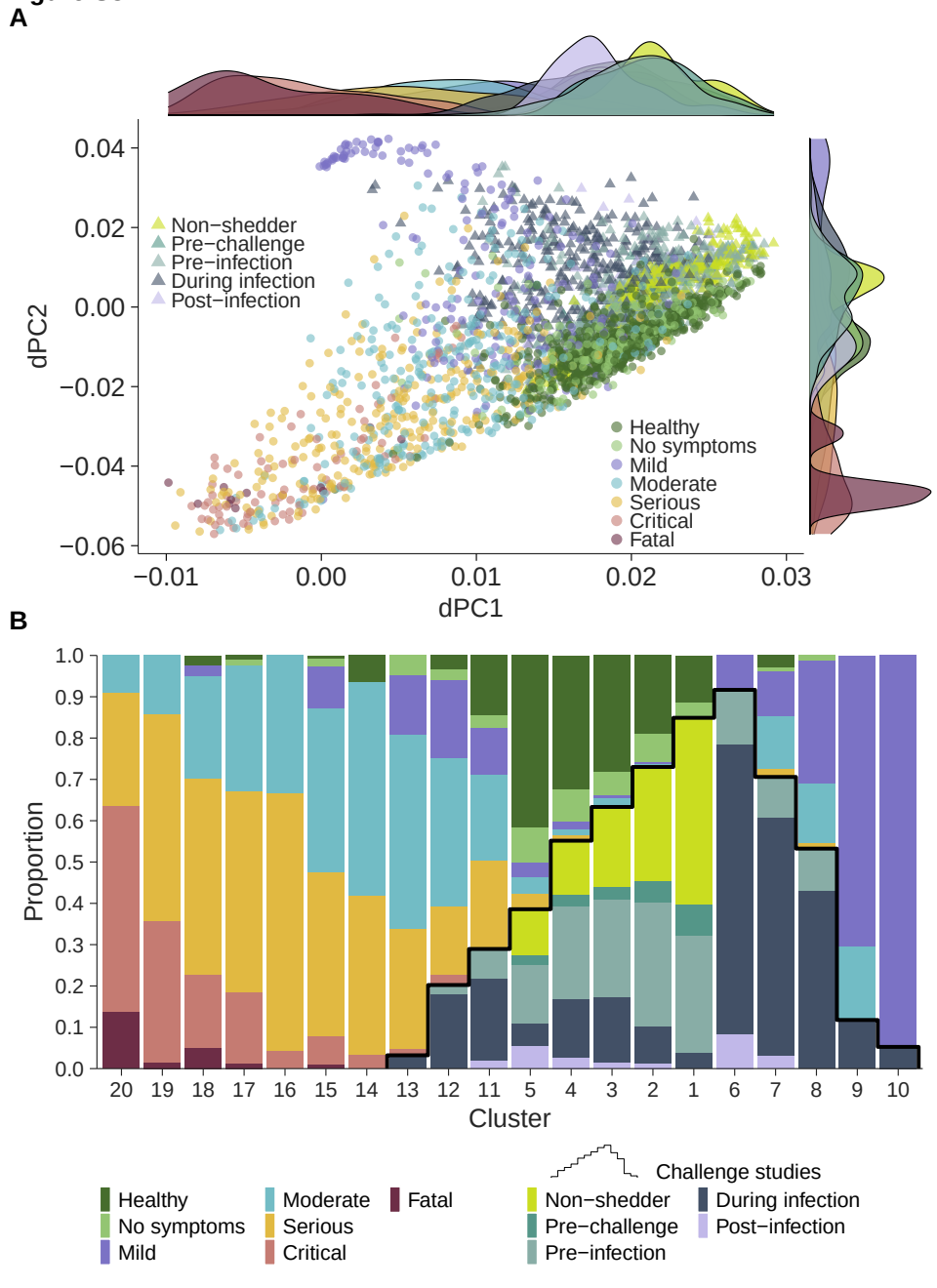

A
Non−shedder
Pre−challenge
Pre−infection
During infection
Post−infection
Healthy
No symptoms
Mild
Moderate
Serious
Critical
Fatal
dPC1
dPC2
B
Proportion
Cluster
Challenge studies

Figure S4


A
Discovery
KLRD1
PIK3R1
Expression
Healthy
No symptoms
Mild
Moderate
Serious
Critical
Fatal
B
SARS−CoV−2
C
Discovery
CAMP
BCAT1
LCN2
TXN
GRN
BCL6
OLR1
ARG1
ITGAM
ORM1
D
SARS−CoV−2
E
Discovery
DEFA4
TCEAL9
PRC1
CEP55
AZU1
KIF15
F
SARS−CoV−2
G
Discovery
OASL
VRK2
UBE2L6
CCL2
VAMP5
NAPA
H
SARS−CoV−2

Figure S5

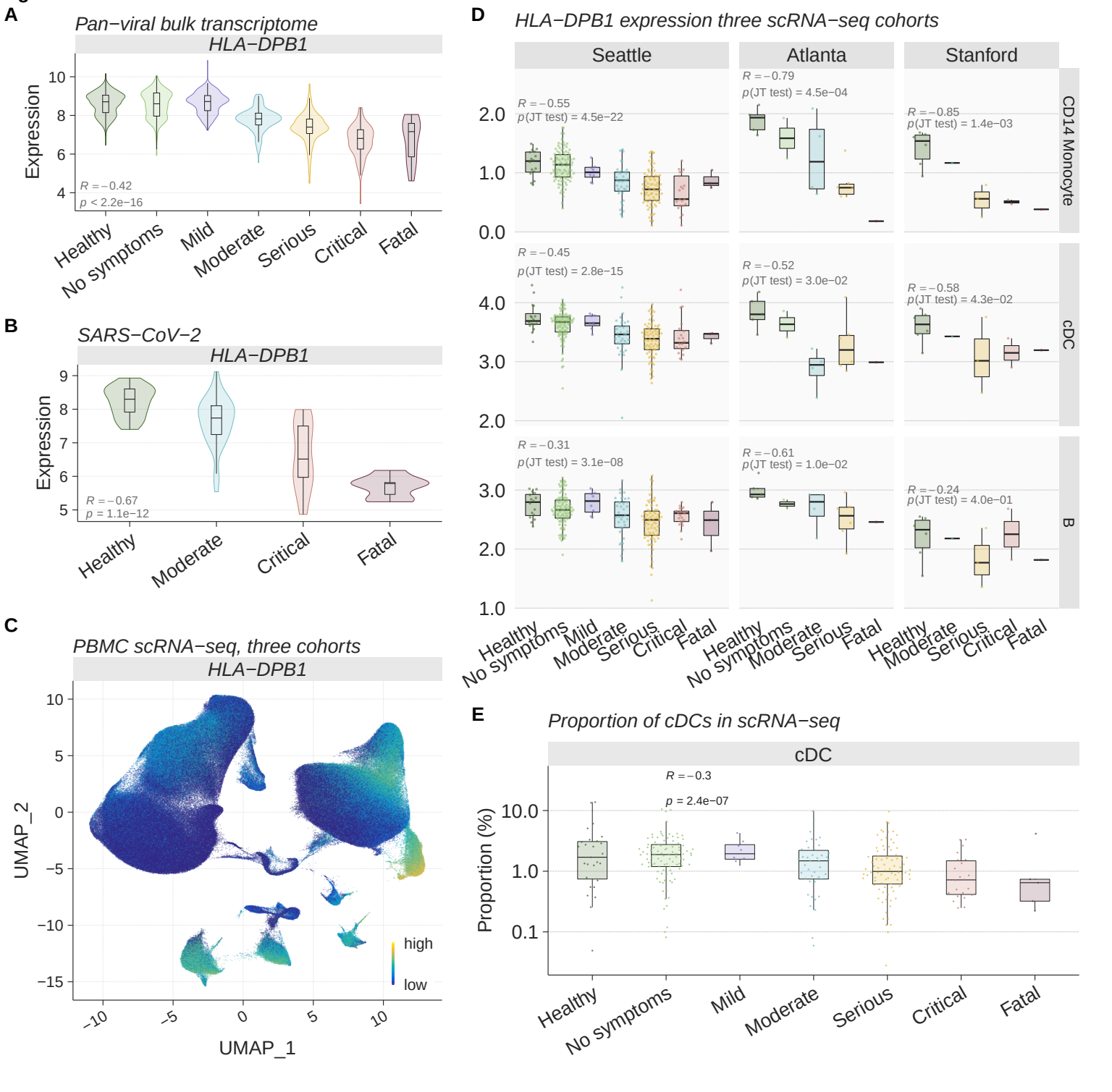

A
Pan−viral bulk transcriptome
HLA−DPB1
Expression
R = −0.42
p < 2.2e−16
Healthy
No symptoms
Mild
Moderate
Serious
Critical
Fatal
B
SARS−CoV−2
HLA−DPB1
Expression
R = −0.67
p = 1.1e−12
Healthy
Moderate
Critical
Fatal
C
PBMC scRNA−seq, three cohorts
HLA−DPB1
UMAP_2
UMAP_1
high
low
D
HLA−DPB1 expression three scRNA−seq cohorts
Seattle
Atlanta
Stanford
CD14 Monocyte
cDC
B
R = −0.55
p(JT test) = 4.5e−22
R = −0.79
p(JT test) = 4.5e−04
R = −0.85
p(JT test) = 1.4e−03
R = −0.45
p(JT test) = 2.8e−15
R = −0.52
p(JT test) = 3.0e−02
R = −0.58
p(JT test) = 4.3e−02
R = −0.31
p(JT test) = 3.1e−08
R = −0.61
p(JT test) = 1.0e−02
R = −0.24
p(JT test) = 4.0e−01
Healthy
No symptoms
Mild
Moderate
Serious
Critical
Fatal
E
Proportion of cDCs in scRNA−seq
cDC
R = −0.3
p = 2.4e−07
Proportion (%)
Healthy
No symptoms
Mild
Moderate
Serious
Critical
Fatal

Figure S6

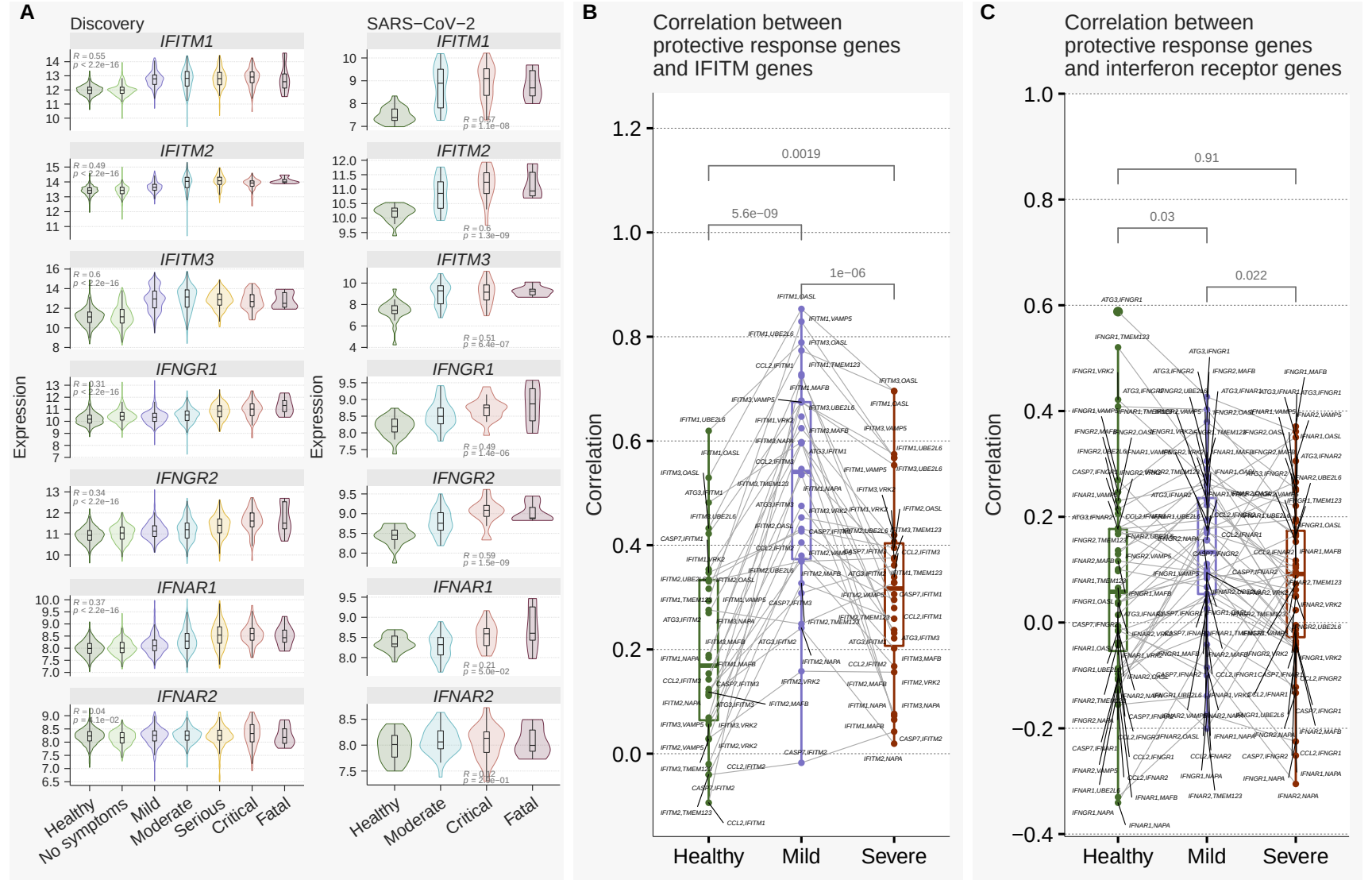
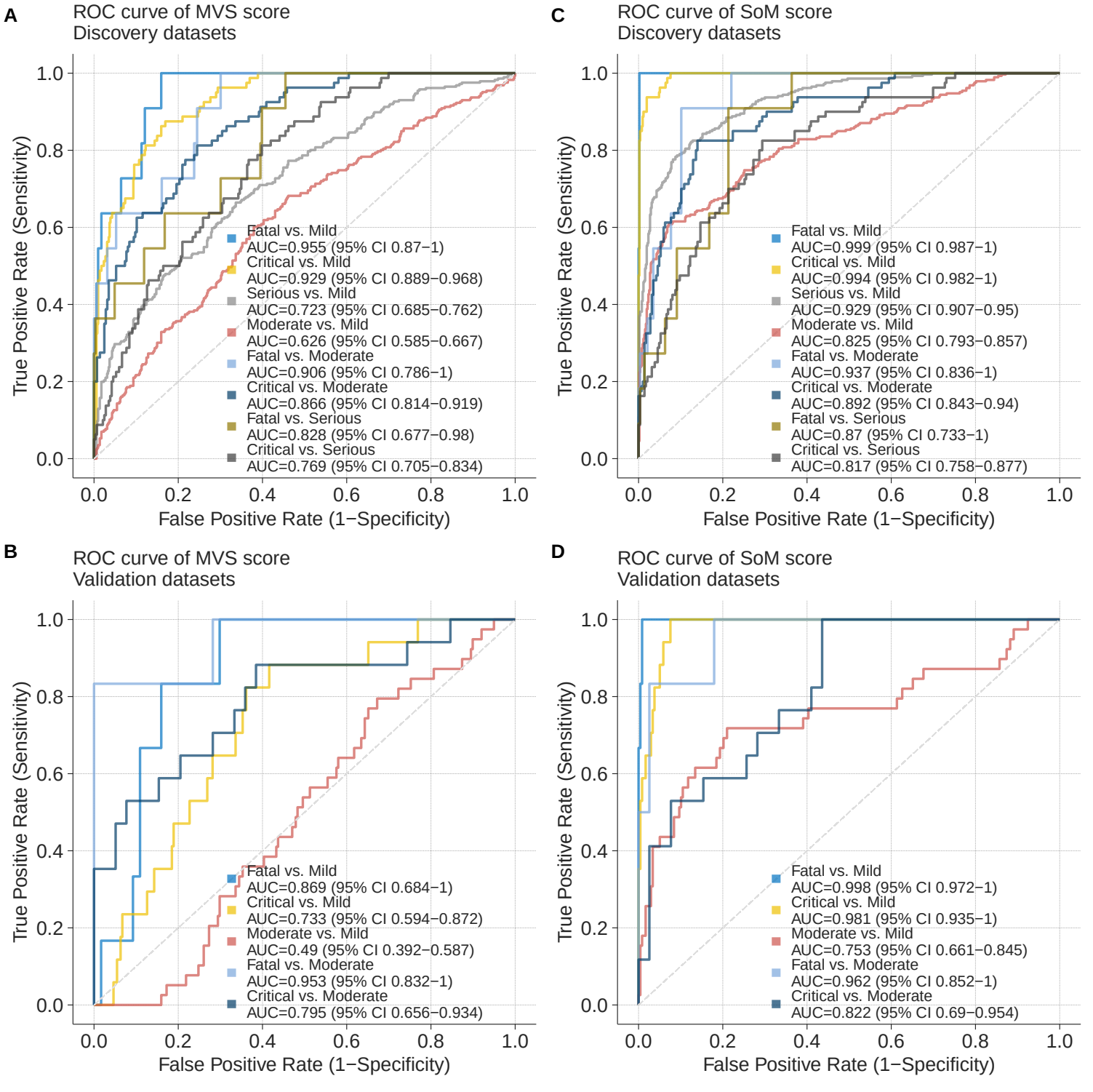

A
Discovery
SARS−CoV−2
IFITM1
IFITM2
IFITM3
IFNGR1
IFNGR2
IFNAR1
IFNAR2
Expression
R = 0.55
p < 2.2e−16
R = 0.49
p < 2.2e−16
R = 0.6
p < 2.2e−16
R = 0.31
p < 2.2e−16
R = 0.34
p < 2.2e−16
R = 0.37
p < 2.2e−16
R = 0.04
p = 4.1e−02
R = 0.67
p = 1.1e−08
R = 0.6
p = 1.3e−09
R = 0.51
p = 6.4e−07
R = 0.49
p = 1.4e−06
R = 0.59
p = 1.5e−09
R = 0.21
p = 5.0e−02
R = 0.02
p = 2.6e−01
Healthy
No symptoms
Mild
Moderate
Serious
Critical
Fatal
Healthy
Moderate
Critical
Fatal
B
Correlation between protective response genes and IFITM genes
Correlation
0.0019
5.6e−09
1e−06
Healthy
Mild
Severe
C
Correlation between protective response genes and interferon receptor genes
Correlation
0.91
0.03
0.022
Healthy
Mild
Severe

Figure S7


A
ROC curve of MVS score
Discovery datasets
True Positive Rate (Sensitivity)
False Positive Rate (1−Specificity)
Fatal vs. Mild
AUC=0.955 (95% CI 0.87−1)
Critical vs. Mild
AUC=0.929 (95% CI 0.889−0.968)
Serious vs. Mild
AUC=0.723 (95% CI 0.685−0.762)
Moderate vs. Mild
AUC=0.626 (95% CI 0.585−0.667)
Fatal vs. Moderate
AUC=0.906 (95% CI 0.786−1)
Critical vs. Moderate
AUC=0.866 (95% CI 0.814−0.919)
Fatal vs. Serious
AUC=0.828 (95% CI 0.677−0.98)
Critical vs. Serious
AUC=0.769 (95% CI 0.705−0.834)
C
ROC curve of SoM score
Discovery datasets
True Positive Rate (Sensitivity)
False Positive Rate (1−Specificity)
Fatal vs. Mild
AUC=0.999 (95% CI 0.987−1)
Critical vs. Mild
AUC=0.994 (95% CI 0.982−1)
Serious vs. Mild
AUC=0.929 (95% CI 0.907−0.95)
Moderate vs. Mild
AUC=0.825 (95% CI 0.793−0.857)
Fatal vs. Moderate
AUC=0.937 (95% CI 0.836−1)
Critical vs. Moderate
AUC=0.892 (95% CI 0.843−0.94)
Fatal vs. Serious
AUC=0.87 (95% CI 0.733−1)
Critical vs. Serious
AUC=0.817 (95% CI 0.758−0.877)
B
ROC curve of MVS score
Validation datasets
True Positive Rate (Sensitivity)
False Positive Rate (1−Specificity)
Fatal vs. Mild
AUC=0.869 (95% CI 0.684−1)
Critical vs. Mild
AUC=0.733 (95% CI 0.594−0.872)
Moderate vs. Mild
AUC=0.49 (95% CI 0.392−0.587)
Fatal vs. Moderate
AUC=0.953 (95% CI 0.832−1)
Critical vs. Moderate
AUC=0.795 (95% CI 0.656−0.934)
D
ROC curve of SoM score
Validation datasets
True Positive Rate (Sensitivity)
False Positive Rate (1−Specificity)
Fatal vs. Mild
AUC=0.998 (95% CI 0.972−1)
Critical vs. Mild
AUC=0.981 (95% CI 0.935−1)
Moderate vs. Mild
AUC=0.753 (95% CI 0.661−0.845)
Fatal vs. Moderate
AUC=0.962 (95% CI 0.852−1)
Critical vs. Moderate
AUC=0.822 (95% CI 0.69−0.954)

Data S1: Study Summaries (related to Figure 1)

This section describes each of the datasets used in the analyses. In order to provide the most accurate information, their descriptions have been used verbatim from the original corresponding manuscripts as much as possible.

PRJNA352396: Liu *et al.* (Liu et al., 2017) collected peripheral blood from infected and convalescent, recovering Ebola virus (EBOV) patients located in Guinea during the 2013-2016 outbreak. They first used deep sequencing to define the transcriptomic profile of blood taken from acute patients who either went on to survive or die from EBOV. These transcriptional signatures from acutely ill patients were compared to profiles obtained from former patients who had recovered from EBOV and were EBOV-negative by qRT-PCR and also to data obtained from healthy volunteers mined from historical datasets.

PRJNA507472: Soares *et al.* (Soares-Schanoski et al., 2019) collected blood samples from adult subjects reporting arbovirus-like symptoms and controls in Sergipe and São Paulo Brazil. A subset of the patients returned for clinical follow-up, and these Chikungunya (CHIKV) infected patients were under treatment and their joint manifestations were monitored quarterly. Patients who presented joint pain associated or not with joint or periarticular edema at the onset or significant worsening (in cases of previous joint disease) of the acute febrile syndrome after three months were considered chronic. Real-time RT-PCR was performed to test for CHIKV, Zika virus, and Dengue (DENV).

PRJNA390289: Michlmayr *et al.* (Michlmayr et al., 2018) selected children (aged 6 months to 14 years) with Chikungunya (CHIKV) infection who were enrolled in the ongoing study at the National Pediatric Reference Hospital (Hospital Infantil Manuel de Jesús Rivera; HIMJR) in Nicaragua. CHIKV cases were laboratory-confirmed by detection of CHIKV using real-time RT–PCR in some cases followed by virus isolation, in acute-phase samples. In addition, seroconversion by IgM capture ELISA and/or a >4-fold increase in antibody titer by Inhibition ELISA in paired acute and convalescent sera were evaluated. Participants were also screened for dengue virus (DENV) infection, and CHIKV/DENV co-infections were excluded. Children with severe symptoms were excluded. All selected cases had an acute-phase sample collected on days 1–2 of illness and a convalescent sample collected on days 15–17 post-illness for serum, whole blood in PAXgene solution, and PBMCs; these were subject to transcriptomic analysis via RNAseq, CHIKV viral titer assays, multiplex ELISA for cytokines, and mass cytometry for cellular phenotyping. Follow-up samples were collected at 3, 6, and 12 months post-infection to study long-term outcomes of CHIKV infection. Sampling times closely adhered to the targeted acute (standard deviation [SD] = 0.5 days) and convalescent (SD = 0.6 days) timepoints. Clinical information was collected every 12 h, and after systematic monitoring by a clinical supervisor was digitized by double-data entry with quality control checks performed daily and weekly.

SARS-CoV-2: Giamarellos-Bourboulis *et al.* (Giamarellos-Bourboulis et al., 2020) collected samples from adult patients with community-acquired pneumonia (CAP) and sepsis and CAP by SARS-CoV-2 of either gender within the first 24 h of hospital admission. CAP was defined as the presence of new infiltrate in chest X-ray in a patient without any contact with any health- care facility the last 90 days; sepsis was defined by the Sepsis-3 criteria. COVID was diagnosed for patients with CAP confirmed by chest X-ray or chest computed tomography and positive molecular testing of respiratory secretions. For patients who required mechanical ventilation (MV), blood sampling was performed within the first 24 h from MV and results were used for this analysis. Exclusion criteria were infection by the human immunodeficiency virus; neutropenia; and any previous intake of immunosuppressive medication (corticosteroids, anti-cytokine biologicals, and biological response modifiers). Severe respiratory failure (SRF) was defined as a severe decrease of the respiratory ratio necessitating intubation and MV. All data of patients with bacterial CAP screened for participation in the randomized clinical trial with the acronym PROVIDE (ClinicalTrials.gov NCT03332225) were used. All patients admitted for CAP by SARS-CoV-2 from March 3 until March 30, 2020 participated in the study. RNAseq was performed on PBMCs.

GSE77087: De Steenhuijsen Piters *et al.* (de Steenhuijsen Piters et al., 2016) prospectively enrolled previously healthy children (<2 years) with a first episode of RSV infection during four consecutive RSV seasons at Nationwide Children's Hospital, Columbus, Ohio either at the outpatient clinics or within a median of 24 hours (interquartile range 17–39 h) of admission in the pediatric ward or the pediatric intensive care unit (PICU). Asymptomatic healthy control subjects were enrolled during routine primary care visits or elective surgery not involving the respiratory tract. In addition to the need for hospitalization, RSV disease severity was assessed using a clinical disease severity score and by the need for supplemental oxygen, PICU admission, and length of stay. At enrollment, they obtained a blood sample from patients and controls for white blood cell count and transcriptome analysis, a nasopharyngeal bacterial swab for bacterial quantitative polymerase chain reaction (PCR) and microbiome analysis, and a nasal wash for RSV quantitation. RNA was extracted from whole-blood samples and hybridized onto Illumina HT12-V4 beadchips.

GSE73072: *Liu et al.* (Liu et al., 2016) quantified the diagnostic advantage of pairing a person's baseline reference with his or her target sample using gene microarray data collected in a large-scale serially sampled respiratory virus challenge study. The data was collected over 7 challenge studies performed by Duke University under a contract awarded by the DARPA Predicting Health and Disease program. These 7 challenge studies are referred to as RSV DEE1, H3N2 DEE2, H1N1 DEE3, H1N1 DEE4, H3N2 DEE5, HRV UVA, and HRV DUKE. These studies consist of 2886 microarray chips assaying 12,023 genes of 148 human volunteer subjects under 4 different inoculation regimes (HRV, RSV, H1N1, H3N2). During each challenge study, subjects had peripheral blood taken prior to inoculation with virus (baseline), immediately prior to inoculation (pre-challenge) and at set intervals following challenge. RNA was extracted at Expression Analysis (Durham, NC) from whole blood using the PAXgene™ 96 Blood RNA Kit (PreAnalytiX, Valencia, CA) employing the manufacturer's protocol. Hybridization and microarray data collection were performed using the Human Genome U133A 2.0 Array (Affymetrix, Santa Clara, CA).

GSE68310: Zhai *et al.* (Zhai et al.) enrolled healthy adults aged 18 to 49 to be followed for acute respiratory illness (ARI) through two consecutive influenza seasons 2009-2010 and 2010-2011. Baseline blood specimens were obtained during enrollment and subjects were given thermometers and instructions to call and report for evaluation within 48 hours of ARI onset. Those persons who had a new ARI were seen within 48 hours of onset (day 0) and 2, 4, and 6 days later for repeat evaluation, blood specimen collections, and medical care (including the antiviral zanamivir if indicated) and 21 days later for collection of convalescent specimens. Nasal wash samples were collected for virus detection by RT-PCR on day 0 and day 2. Surveillance for influenza was terminated after 5.5 months; all subjects were asked to return for specimen collection and to provide a medical and ARI history in spring of next year. Serum specimens obtained at baseline, day 0 and day 21 visits for illnesses, and the terminal visit were tested simultaneously using hemagglutination-inhibition (HAI) antibody assay for Influenza H1N1, H3N2, and Influenza B strains. Peripheral blood RNA (PaxGene) obtained from blood specimens at each visit were analyzed using Illumina Human HT-12 v4. The study was repeated 2010-2011.

GSE68004: Jaggi *et al.* (Jaggi et al., 2018) prospectively enrolled and obtained blood samples in a total of 162 children < 18 years of age at three hospitals (Nationwide Children's Hospital in Columbus, OH, Children's Medical Center in Dallas, TX and Rady Children's Hospital in San Diego, CA) hospitalized with Kawasaki Disease or other febrile illnesses and 37 healthy, asymptomatic children as controls who were age-, gender- and race-matched. Of the 125 hospitalized children, 19 patients had confirmed adenovirus infection and only these samples and the controls were used for our analysis. Healthy controls were enrolled when undergoing elective surgery for non-infectious related causes, or at routine outpatient visits.

GSE67059: Heinonen *et al.* (Heinonen et al., 2016) prospectively enrolled previously healthy children <2 years old recruited over six respiratory seasons (2007–2011 and 2013–2015) from four study sites (Columbus, OH; Dallas, TX; Turku, Finland; and Malaga, Spain). Children with respiratory symptoms were enrolled at the emergency department (outpatients) or within 48 hours of hospitalization (inpatients) and were classified into:

HRV- healthy controls (HC; n=37); HRV+ asymptomatic (n=14); HRV+ outpatients (n=30); and HRV+ inpatients (n=70). Exclusion criteria: documented bacterial or viral coinfections, systemic corticosteroid treatment in the preceding 2 weeks, prematurity (<36 wk of gestation), immunodeficiency, or chronic medical conditions. Whole blood RNA samples were hybridized into Illumina arrays. Healthy control subjects were enrolled during well-child visits or minor elective surgical procedures not involving the respiratory tract who did not have a history of a respiratory tract infection within 2 weeks of enrollment. Blood and nasopharyngeal (NP) samples were collected at enrollment from all subjects and analyzed for transcriptional profiles and respiratory viruses. Disease severity was assessed according to the need for hospitalization (inpatients [moderate or severe disease] vs. outpatients [mild disease]) and using a standardized clinical disease severity score (CDSS).

GSE66099: Sweeney *et al.* (Sweeney et al., 2015) created this dataset which is composed of unique patients present in the six other GEO datasets published by the Genomics of Pediatric SIRS and Septic Shock Investigators, which includes all unique patients from GSE4607 (described below), GSE8121 (Shanley et al., 2007), GSE9692 (Cvijanovich et al., 2008), GSE13904 (Wong et al., 2009a), GSE26378 (Wynn et al., 2011), and GSE26440 (Wong et al., 2009b). Only unique patients from the Day 1 timepoint are included. The original studies examined pediatric patients admitted to the ICU, who were later classified as either SIRS (non-infectious) or Sepsis or Septic Shock (infectious). There is also a group of healthy controls. Although the original studies examine patients at both ICU day 1 and ICU day 3, patients here are only aggregated ICU day 1 patients. All samples were downloaded as CEL files and re-normalized with gcRMA using R package 'affy'.

GSE4607: Wong *et al.* (Wong et al., 2012; 2008; 2007) enrolled children < 10 years of age admitted to the pediatric intensive care unit (PICU) who met criteria for pediatric-specific septic shock as defined by the 2005 International pediatric sepsis consensus conference. Control patients were recruited from the participating institutions using the following exclusion criteria: any acute illness, a recent febrile illness (within 2 weeks), recent use of anti-inflammatory medications (within 2 weeks), or any history of chronic or acute disease associated with inflammation. Whole blood samples were obtained within 24 hours of admission to the PICU, referred to as "Day 1" of septic shock, for microarray analysis with Affymetrix Human Genome U133 Plus 2.0 GeneChip. All study patients were followed for 28 days to determine survival. A subset of patients with influenza, rotavirus, and varicella, and controls were used in our analysis, and we excluded samples that overlapped with those in GSE66099.

GSE6269: Ramilo *et al.* (Ramilo et al., 2007) profiled PBMCs from patients (age ≤ 18 years) with acute infections caused by: (1) an RNA virus (influenza A); (2) two gram-positive bacteria (*Staphylococcus aureus* and *Streptococcus pneumoniae*); and (4) a Gram- negative bacterium (*Escherichia coli)*. Patients were treated according to standard hospital protocols, and antimicrobial therapy was promptly initiated in the emergency department. The patients were treated with up to three different drugs from an overall set of 13 drugs. The study profiled these samples using three types of microarrays from two manufacturers, Affymetrix (HG U133A and HG U133 plus 2.0) and Illumina, representing technical variability.

GSE61821: Hoang *et al.* (Hoang et al., 2014) recruited patients from two cohorts 1) Mild influenza and other febrile illness (OFI) samples, and 2) moderate and severe samples as described below.

Mild and OFI cohort: Patients were recruited between January 2008 and January 2010 from an undifferentiated fever inclusion study (EDEN) in Singapore, who were ≥18 years of age and presented ≤72 h from onset of fever ≥38°C. Patients were tested by PCR for influenza A and B, RSV, parainfluenza 1–3, coronavirus, metapneumovirus, enterovirus and adenovirus in nasal swabs and for dengue virus 1–4, human parvovirus B19, Cytomegalovirus, and Epstein Barr virus in EDTA blood. They performed whole-blood transcriptional profiling on all samples from both groups at 72 hours after fever onset, at 3–8 days and 3–4 weeks after self-reported fever onset.

Moderate and severe cohort: Samples of moderate and severe influenza patients were collected from a multi-center, double-blinded, randomized control trial of different doses of oseltamivir for the treatment of

hospitalized patients with influenza (Registered ClinicalTrials.gov: NCT00298233). The study took place across 13 hospitals in the Southeast Asia Infectious Disease Clinical Research Network (SEAICRN). Whole-blood samples collected before oseltamivir treatment (study day 0) and at follow up (day 28) were used. The inclusion criteria were: (i) age ≥1 year, (ii) duration of illness ≤10 (non-H5N1) or ≤14 (H5N1) days, (iii) positive result for influenza virus A or B using a rapid antigen test or qualitative RT-PCR in a respiratory specimen, (iv) presence of at least one respiratory symptom (cough, dyspnea or sore throat), (v) disease requiring hospital admission, and (vi) one of the following signs of severe influenza: (a) new infiltrate on a chest X-ray, (b) tachypnea (respiratory rate ≥30 for ages ≥12 years, rate ≥40 for ages 6 to 12 years, rate ≥45 for ages 3 to 6 years, rate ≥50 for ages 1– to 3 years), (c) dyspnea (unable to speak full sentences, or use of accessory respiratory muscles), or (d) hypoxia (arterial oxygen saturation ≤92% on room air by a transcutaneous method). Subjects infected with avian influenza virus A/H5N1 were enrolled with any degree of severity. The exclusion criteria were: (i) pregnancy or urine β-hCG positivity, (ii) breast feeding, (iii) prior oseltamivir therapy for >72 hours duration or double dose (any duration) within the past 14 days, (iv) allergy or severe intolerance of oseltamivir, (v) creatinine clearance (CrCl) <10 mL/min. Severe influenza was defined as: i) requiring mechanical ventilation or ii) presenting with severe tachypnea (respiratory rate ≥30 for ages ≥12 years, rate ≥40 for ages 6 to 12 years, rate ≥45 for ages 3 to 6 years, rate ≥50 for ages 1–to 3 years) and hypoxia (arterial oxygen saturation ≤92% on room air by a transcutaneous method. Gene expression microarray was performed using one-color array technology on the Illumina platform (Illumina Inc, San Diego, CA, USA).

GSE61754: Davenport *et al.* (Davenport et al., 2014) recruited 22 healthy volunteers aged 18–45 years, and 11 were vaccinated with 1.5×10^8 plaque forming units (pfu) of MVA-NP+M1. Thirty days after vaccination, all volunteers underwent intranasal challenge with H3N2 influenza (A/Wisconsin/67/2005) at a dose of 1 ml of $10^{5.25}$ TCID50/ml in a quarantine facility. Volunteers were all challenged within a 2-h period. All subjects had hemagglutination inhibition (HI) titers of <1:10 to the challenge virus on admission to the quarantine unit. After viral challenge, volunteers had a physical examination by a trial physician daily and self-reported their symptoms using a modified Jackson scoring system (Zaas et al., 2009) twice daily, from 12-h post-challenge, for 6 days. This system lists upper respiratory and systemic symptoms on a scale of 0–3: 'no symptoms', 'just noticeable', 'bothersome but can still do activities' and 'bothersome and cannot do daily activities.' Scores were summed over the duration of the challenge period to give a total symptom score. Those volunteers with scores of ≥4 and positive viral culture from nasal wash samples were categorized as having laboratory-confirmed influenza (LCI). Severity of infection was graded as mild if the summed symptom score was 4–28, with scores of ≥29 classified as moderate/severe. Volunteers were kept in quarantine until day 6 post-challenge, and all had a negative rapid antigen test on nasal wash sample prior to discharge. Whole blood was collected in PAXgene tubes prior to influenza challenge (day 30) and then at three further time points (12, 24 and 48 h post-challenge) for genome-wide gene expression analysis using the Illumina HumanHT-12 v4 Expression BeadChip. All nasal washing and blood collections for viral shedding and gene expression analysis were taken in the same order that the volunteers were challenged in.

GSE40012: Parnell *et al.* (Parnell et al., 2012) recruited patients with severe community-acquired pneumonia requiring ICU admission, patients with noninfective systemic inflammatory response syndrome (SIRS), and healthy controls. They defined SIRS as the presence of at least two of the following four clinical criteria: (a) fever or hypothermia (temperature > 100.4°F (38°C) or < 96.8°F (36°C)); (b) tachycardia (> 90 beats/min), (c) tachypnea (> 20 breaths/min or PaCO2 < 4.3 kPa (32 mm Hg)), or the need for mechanical ventilation; (d) an altered white blood cell count of > 12,000 cells/μl, < 4,000 cells/μl, or the presence of > 10% band forms. Pneumonia was defined as a microbiologically confirmed infection of the lungs in a patient fulfilling the SIRS criteria. Patients were between ages 22 to 75 years. Influenza A H1N1 2009 was confirmed by PCR, and bacterial pneumonia by microbiological cultures. The first blood sample from each patient was collected within 24 hours of ICU

admission. Patients were monitored for up to 5 days for microarray analysis of longitudinal gene-expression profiles.

GSE38900: Mejias *et al* (Mejias et al., 2013) prospectively enrolled children over six respiratory seasons including a cohort of hospitalized infants (age <2 years) with RSV, HRV, and influenza infections from three centers: (1) Nationwide Children's hospital, Columbus, Ohio, (2) Turku University (Turku, Finland), and (3) Children's Medical Center, Dallas, Texas and blood samples were collected for microarray analysis. Children with documented viral or bacterial co-infections (bacteremia, urinary tract infection, meningitis, acute gastroenteritis, or any bacterial pathogen isolated from a sterile site) were excluded. Children with congenital heart disease, chronic lung disease, immunodeficiency, prematurity (<36 weeks), and systemic steroid treatment within 2 weeks prior to presentation were also excluded. Control samples were obtained from healthy children without comorbidities, use of systemic steroids, or presence of any illness within 2 weeks prior to enrollment who were either undergoing elective surgery not involving the respiratory tract, or at outpatient visits for routine care. To exclude viral co-infections, respiratory samples were tested by viral culture or PCR in 94% of patients and controls.

GSE2729: Wang *et al.* (Wang et al., 2007) collected blood and fecal specimens from children < 3 years of age who were hospitalized for treatment of acute rotavirus gastroenteritis, who were otherwise generally in good health with no prior history of rotavirus diarrhea and had not received rotavirus vaccines. All patients had symptoms typical of rotavirus infection (fever, vomiting, diarrhea, or dehydration), and all were confirmed positive by the testing of fecal specimens by using enzyme immunoassay and PCR. Blood samples were drawn at the time of hospital admission, within 72 hours of enrollment, and 3 weeks later from some children. PBMCs were isolated for microarray analysis using Affymetrix Human U95Av2 high density oligonucleotide arrays. Controls were age-matched healthy children admitted for elective surgery with no symptoms of gastroenteritis and fecal specimens that tested negative for rotavirus by enzyme immunoassay.

GSE27131: Berdal *et al.* (Berdal et al., 2011) prospectively recruited ICU patients over a four week period who were > 18 years of age diagnosed with pandemic H1N1 influenza by upper airway PCR with bilateral chest X-rays infiltrates requiring mechanical ventilator support. Patients with significant comorbidity (malignancy and severe systemic inflammatory diseases) were excluded. Age and gender-matched healthy individuals served as controls. Peripheral venous blood was collected for microarray analysis at inclusion and after 3 and 6 days. Blood was centrifuged at 2500× *g* for 20 min to obtain platelet-poor plasma (EDTA) or allowed to clot before centrifugation, after which RNA from blood leukocytes was obtained by PAXgene Blood RNA System for microarray analysis.

GSE25504: Smith *et al.* (Smith et al., 2014) and Dickinson *et al.* (Dickinson et al., 2015) enrolled infants from the Neonatal Unit, Royal Infirmary of Edinburgh, and the Division of Pathway Medicine, University of Edinburgh having blood cultures taken for suspected infection. They obtained whole blood for expression profiling at the time of first clinical signs of suspected and microbiological blood culture. The infected group was defined as patients with suspected clinical infection that proved to have microbiological evidence of infection from a usually sterile body site, and samples were only included as positive if both clinicians agreed that infection was present. Exclusion criteria were: infants of mothers known to be positive for hepatitis B, HIV or hepatitis C viruses, mothers with known history of drug misuse, and mothers that had no antenatal screening for blood-borne viruses. Other exclusion criteria were infants who did not require clinical blood samples and infants for whom extra blood sampling might be of particular risk (such as anemia). Controls were defined as 'well' infants having blood taken for other clinical reasons, which included screening tests for maternal thyroid disease, routine neonatal screening, pigmented scrotum, electrolyte check for previously deranged sodium, bilirubin check due to jaundice, and neonatal encephalopathy, all of which were normal. Samples were divided into four cohorts which were each run on different microarray platforms, three of which were excluded from our meta-analysis (Affymetrix Human Genome U133 Plus 2.0, Illumina HumanHT-12 V3.0 expression beadchip, and

Codelink 55K Human Array) because they did not have at least 2 samples present in each category (viral infection vs control). Patient samples with bacterial or fungal infection were also excluded from our analysis. We therefore only analyzed samples from 9 infants with corrected gestational age ranges of 27^5 to 39^6 weeks and ran on Affymetrix Human Genome U219 Array from this dataset, which included 6 controls and 3 patients with viral infections (2 Rhinovirus, 1 CMV).

GSE21802: Bermejo-Martin *et al.* (Bermejo-Martin et al., 2010) recruited critically ill patients aged 18 to 65 years with influenza pneumonia during the acute phase of illness with acute respiratory distress and unequivocal alveolar opacification involving two or more lobes with negative respiratory and blood bacterial cultures on admission to the ICU. Only patients with confirmed H1N1 infection by RT-PCR were included in the study. Age-matched healthy volunteers were also recruited between workers of the University of Valladolid, Spain. Treatment decisions for all patients, including corticosteroid therapy, were not standardized and were used at the discretion of the attending physician.

GSE20346: Parnell *et al.* (Parnell et al., 2011) recruited hospitalized patients aged 21-75 years with severe infection, defined as the presence of at least one major organ failure requiring critical care intervention. These patients had either viral (seasonal H3N2 or pandemic H1N1/09 influenza virus) or bacterial infections and were followed for four days. Healthy volunteers were enrolled from a local influenza vaccination program.

GSE17156: Zaas *et al.* (Zaas et al., 2009) inoculated healthy volunteers (HV) with one of the following viruses (HRV, RSV, influenza) as described below.

> **HRV cohort:** The HRV challenge subjects were recruited through an active screening protocol at the University of Virginia (Charlottesville, VA). On the day of inoculation, 10^6TCID50 GMP HRV serotype 39 was administered intranasally. Subjects were admitted to a quarantine facility for 48 hours post-inoculation. Nasopharyngeal (NP) lavage samples were obtained from each subject daily for titers to accurately gauge the success and timing of inoculation. After 48 hours post-inoculation, subjects were released from quarantine and returned for 3 consecutive days for further sample acquisition.
>
> **RSV cohort:** The RSV challenge was performed by Retroscreen Virology, Ltd. (London). On the day of inoculation, a dose of 10^4 TCID50 RSV (serotype A) was administered intranasally. Blood and NP lavage samples were collected as in the HRV cohort, but continued throughout quarantine. Due to the incubation period of RSV A, subjects were not released from quarantine until they were 288 hours post-inoculation and negative for rapid RSV antigen detection.
>
> **Influenza cohort:** The influenza challenge A/Wisconsin/67/2005 (H3N2) was performed at Retroscreen Virology, Ltd. (Brentwood, UK). On the day of inoculation, a 10^6 dose of TCID50 influenza A was diluted and administered intranasally per standard methods at different doses: (1:10, 1:100, 1:1000, 1:10000) with 4-5 subjects receiving each dose. Due to the incubation period, subjects were not released from quarantine until 168th hours post-inoculation. Blood and NP lavage samples were collected throughout the duration of quarantine. All subjects received oral oseltamivir (Roche Pharmaceuticals) (75 mg) by mouth twice daily at day 6 post-inoculation and were negative by rapid antigen detection at the time of discharge.

GSE117827: Yu *et al.* (Yu et al., 2019) profiled children age 3 months to 18 years hospitalized for acute respiratory illness with positive results only for a single virus (Rhinovirus, RSV, Enterovirus, Coxsackievirus) on a multiplex nasopharyngeal swab obtained as part of the child's routine care (BioFire FilmArray Respiratory Panel, Salt Lake City, UT). Exclusion criteria included any underlying medical condition that required regular medical care, receipt of immunosuppressive medications, corticosteroids within the preceding 30 days, and receipt of antibiotics within 7 days. Controls were age-matched children having ambulatory surgery for non-acute conditions. Exclusion criteria were the same as for symptomatic subjects plus fever within the preceding 48 hours. Blood samples were profiled using microarray with Affymetrix.

GSE111368: Dunning *et al*. (Dunning et al., 2018) and the Mechanisms of Severe Acute Influenza Consortium (MOSAIC) profiled whole blood for microarray analysis from patients aged ≥ 16 years during two successive winters (2009–2010 and 2010–2011) with pandemic H1N1 and seasonal H3N2 (determined by RT-PCR) admitted to general wards or ICU. They included patients with prior or concurrent comorbidities: (most commonly asthma, pregnancy, immunocompromising conditions or co-infection with other respiratory pathogens), to reflect the populations known to be at greatest risk of severe influenza. Adult healthy age, sex and ethnicity-matched control subjects screened to exclude known illnesses or current use of medications were also included (Registered Clinical Trial NCT00965354). Samples were obtained at three time points: T1 (recruitment), T2 (approximately 48 h after T1) and T3 (at least 4 weeks after T1). Severity of illness was graded as: 1, no substantial respiratory compromise, with blood oxygen saturation of > 93% while the patient was breathing room air; 2, oxygen saturation of ≤ 93% while the patient was breathing room air, justifying or requiring supplemental oxygen by face mask or nasal cannulae (with or without continuous positive airway pressure support or non-invasive mechanical ventilation); 3, respiratory compromise requiring invasive mechanical ventilation with or without ECMO (extracorporeal membrane oxygenation). Nasopharyngeal aspirates and swabs collected at T1 underwent microscopy and culture for bacteria, and multiplex PCR was performed to detect S. aureus, C. pneumoniae, H. influenzae, S. pneumoniae, Pneumocystis pneumoniae, Legionella species, Klebsiella pneumoniae, Salmonella species, Moraxella catarrhalis, Mycoplasma pneumoniae and Bordetella pertussis. Throat swab samples obtained at T1 also underwent culture and microscopy. Where available, urine samples collected between T1 and T2 underwent pneumococcal antigen testing (BinaxNow, Alere). An independent microbiologist assessed the significance and validity of positive blood-culture results, in an attempt to exclude cases of pseudobacteremia.

GSE103842: Rodriguez-Fernandez *et al.* (Rodriguez-Fernandez et al., 2017) prospectively recruited previously healthy infants (age < 2 years) hospitalized to general wards or Pediatric ICU over 5 nonconsecutive respiratory seasons with RSV bronchiolitis. Patients were excluded if they were premature (gestational age ≤36 weeks), had a previous episode of documented RSV infection, chronic medical conditions, immunodeficiency, or had received systemic steroids or immunomodulatory drugs within 2 weeks of hospitalization. Healthy asymptomatic controls were enrolled during well-child visits or minor elective surgical procedures not involving the respiratory tract. Bronchiolitis was defined as the presence of rhinitis, tachypnea, wheezing, cough, crackles, use of accessory muscles, and/or nasal flaring with or without fever. Children who met the inclusion criteria and had a positive RSV test per standard of care, were enrolled within 24 hours of hospitalization, and a nasal wash sample was obtained for RSV typing and viral load quantitation. Nasal wash samples were also tested for other respiratory viruses in 75% of patients; 97% using a multiplex assay targeting 17 pathogens and 3% by viral culture. In a subset of infants, blood samples (1–3 mL) were collected and blood RNA was processed and hybridized into Illumina Human HT12 V4 beadchips and scanned on the Illumina Beadstation 500.

GSE101702: Tang *et al.* (Tang et al., 2019) profiled blood samples from hospitalized patients with varying severity of influenza infection. The eligibility criteria included (1) age > 18 years and (2) World Health Organization definition of influenza-like illness (fever of 38 C° or higher, cough and illness onset within the last ten days), and patients with a high likelihood of infection, based on history and clinical features. Airway samples (nasopharyngeal swab, throat sample or sputum) and a 2.5ml peripheral blood sample (into PAXgene tubes) were obtained in each participant within 24 hours of their presentation to the hospital. Airway samples were tested for bacterial pathogens and common respiratory viruses. Blood samples in PAXgene tubes were later processed for microarray analysis (Agilent 8× 60k Human V3) of participants and healthy control subjects without any medical illnesses. In patients admitted to the intensive care unit (ICU), additional respiratory samples were obtained from bronchoalveolar lavage or tracheal aspirates. Standard microbiological testing was performed in these samples, including sputum Gram stain and culture. Testing for atypical respiratory pathogens (Chlamydophila pneumoniae, Mycoplasma pneumoniae, and Legionella pneumophila) was performed in selected patients at the discretion of treating physicians. All patients were tested for respiratory

viruses using nucleic acid PCR. The PCR panel included primers for influenza A, influenza B, respiratory syncytial virus, rhinovirus, parainfluenza virus, and human metapneumovirus. Patients infected with other viruses than influenza were excluded.

E-MTAB-5195: Jong *et al.* (Jong et al., 2016) studied respiratory syncytial virus (RSV) infection in 39 hospitalized young children (age < 2 years) using early blood microarray transcriptome profiles followed until recovery and of which the level of disease severity was determined retrospectively. Patients were recruited from three hospitals, and nasopharyngeal wash and blood samples were prospectively obtained from patients with a viral bronchiolitis. Samples were taken within 24 hours after first contact with the hospital, and only patients with an RSV infection, as retrospectively determined by PCR were included in the study. Exclusion criteria were: immunodeficiency, systemic steroid treatment in the previous 2 weeks, blood transfusion, congenital heart and chronic lung disease. Patients were followed until recovery and were retrospectively classified as: mild for children without hypoxia, moderate for patients requiring supplemental oxygen (oxygen saturations < 90%, ≥ 10 minutes) and severe for children requiring mechanical ventilation due to apnea, exhaustion and/ or respiratory failure. Recovery samples were obtained after 4–6 weeks, during home visits. Blood samples were obtained from healthy controls without underlying diseases or medication subjected to elective surgery.

scRNAseq Datasets: The Seattle Cohort profiled 557,240 immune cells from 258 PBMC samples of HC and patients with SARS-CoV-2 infection (16 healthy, 94 asymptomatic, 8 mild, 37 moderate, 78 serious, 21 critical, 3 fatal) using CITE-seq. The patients with SARS-CoV-2 infection in the Seattle Cohort were profiled at two time points: (1) near the time of a positive clinical diagnosis and (2) a few days later. The Atlanta Cohort profiled 76,929 immune cells from 18 PBMC samples of HC and patients infected with one of 3 viruses (5 healthy, 1 moderate influenza, 1 serious influenza, 2 serious RSV, 2 convalescent SARS-CoV-2, 3 moderate SARS-CoV-2, 3 serious SARS-CoV-2, 1 fatal SARS-CoV-2) using CITE-seq. Finally, the Stanford Cohort profiled 68,801 immune cells from 13 PBMC samples of HC and patients with SARS-CoV-2 infection (6 healthy, 1 moderate, 3 serious, 2 critical, 1 fatal) using Seq-Well.