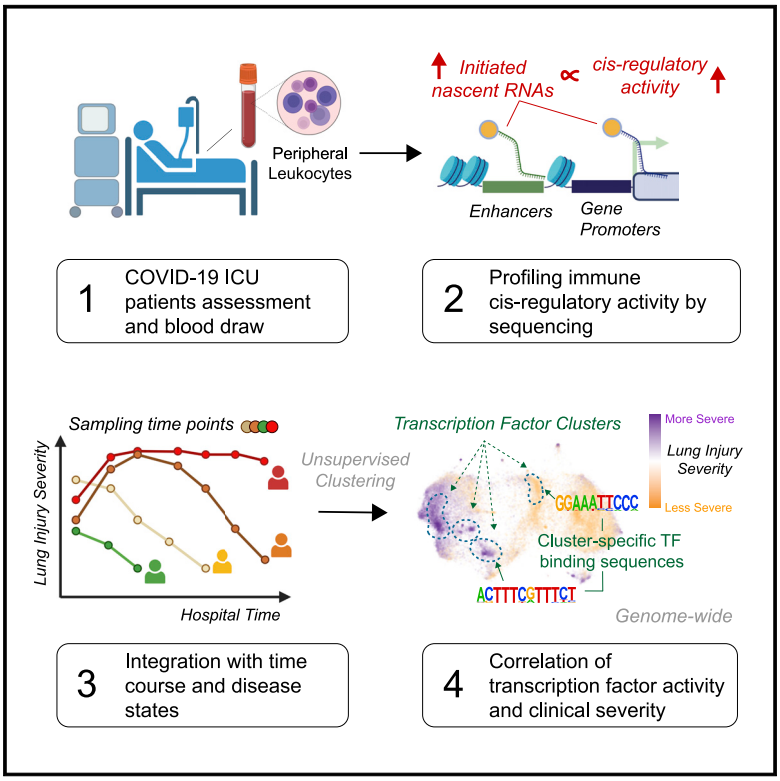
Cell Reports
Medicine

Article

Dynamic activity in *cis*-regulatory elements of leukocytes identifies transcription factor activation and stratifies COVID-19 severity in ICU patients

Graphical abstract



Highlights

- Immune *cis*-regulatory activities are dynamic in COVID-19 ICU patients' courses
- The activity of specific *cis*-regulatory clusters tracks clinical lung injury
- Distinct transcription factor motifs are enriched in *cis*-regulatory clusters
- STAT/BCL6, T1ISRE/STAT, and E2F/MYB activities are associated with poor outcomes

Authors

Michael Tun Yin Lam, Sascha H. Duttke, Mazen F. Odish, ..., Satchidananda Panda, Christopher W. Benner, Nicole G. Coufal

Correspondence

mtlam@health.ucsd.edu

In brief

Understanding how non-coding regulatory DNA functions in critical illnesses is understudied. Lam et al. integrate the dynamic activities of regulatory DNA with the clinical progression of COVID-19 ICU patients. Analysis of DNA binding sequences in these regulatory elements reveals the roles of transcription factor activities as indicators of disease severity.

Lam et al., 2023, Cell Reports Medicine *4*, 100935
February 21, 2023
https://doi.org/10.1016/j.xcrm.2023.100935

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Dynamic activity in *cis*-regulatory elements of leukocytes identifies transcription factor activation and stratifies COVID-19 severity in ICU patients

Michael Tun Yin Lam,[1,2,3,15,*] Sascha H. Duttke,[4,13] Mazen F. Odish,[1,13] Hiep D. Le,[2,13] Emily A. Hansen,[5,6,13] Celina T. Nguyen,[5,13] Samantha Trescott,[5,6,13] Roy Kim,[5,6,13] Shaunak Deota,[2] Max W. Chang,[7] Arjun Patel,[1] Mark Hepokoski,[1] Mona Alotaibi,[1] Mark Rolfsen,[8] Katherine Perofsky,[6,9] Anna S. Warden,[5] Jennifer Foley,[9] Sydney I. Ramirez,[10,11] Jennifer M. Dan,[10,11] Robert K. Abbott,[12] Shane Crotty,[10,11] Laura E. Crotty Alexander,[1,3] Atul Malhotra,[1] Satchidananda Panda,[2] Christopher W. Benner,[7,14] and Nicole G. Coufal[5,6,9,14]

[1]Division of Pulmonary, Critical Care, and Sleep Medicine, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA
[2]Laboratory of Regulatory Biology, Salk Institute for Biological Studies, La Jolla, CA 92037, USA
[3]Pulmonary and Critical Care Section, VA San Diego Healthcare System, La Jolla, CA 92161, USA
[4]School of Molecular Biosciences, College of Veterinary Medicine, Washington State University, Pullman, WA 99163, USA
[5]Sanford Consortium for Regenerative Medicine, La Jolla, CA 92037, USA
[6]Department of Pediatrics, University of California, San Diego, La Jolla, CA, USA
[7]Division of Endocrinology and Metabolism, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA
[8]Internal Medicine Residency Program, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA
[9]Rady Children's Hospital, San Diego, CA 92123, USA
[10]Division of Infectious Diseases, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA
[11]Center for Infectious Diseases and Vaccine Research, La Jolla Institute for Immunology (LJI), La Jolla, CA 92037, USA
[12]Department of Pathology, University of Texas Medical Branch, Galveston, TX 77555, USA
[13]These authors contributed equally
[14]Senior author
[15]Lead contact
*Correspondence: mtlam@health.ucsd.edu
https://doi.org/10.1016/j.xcrm.2023.100935

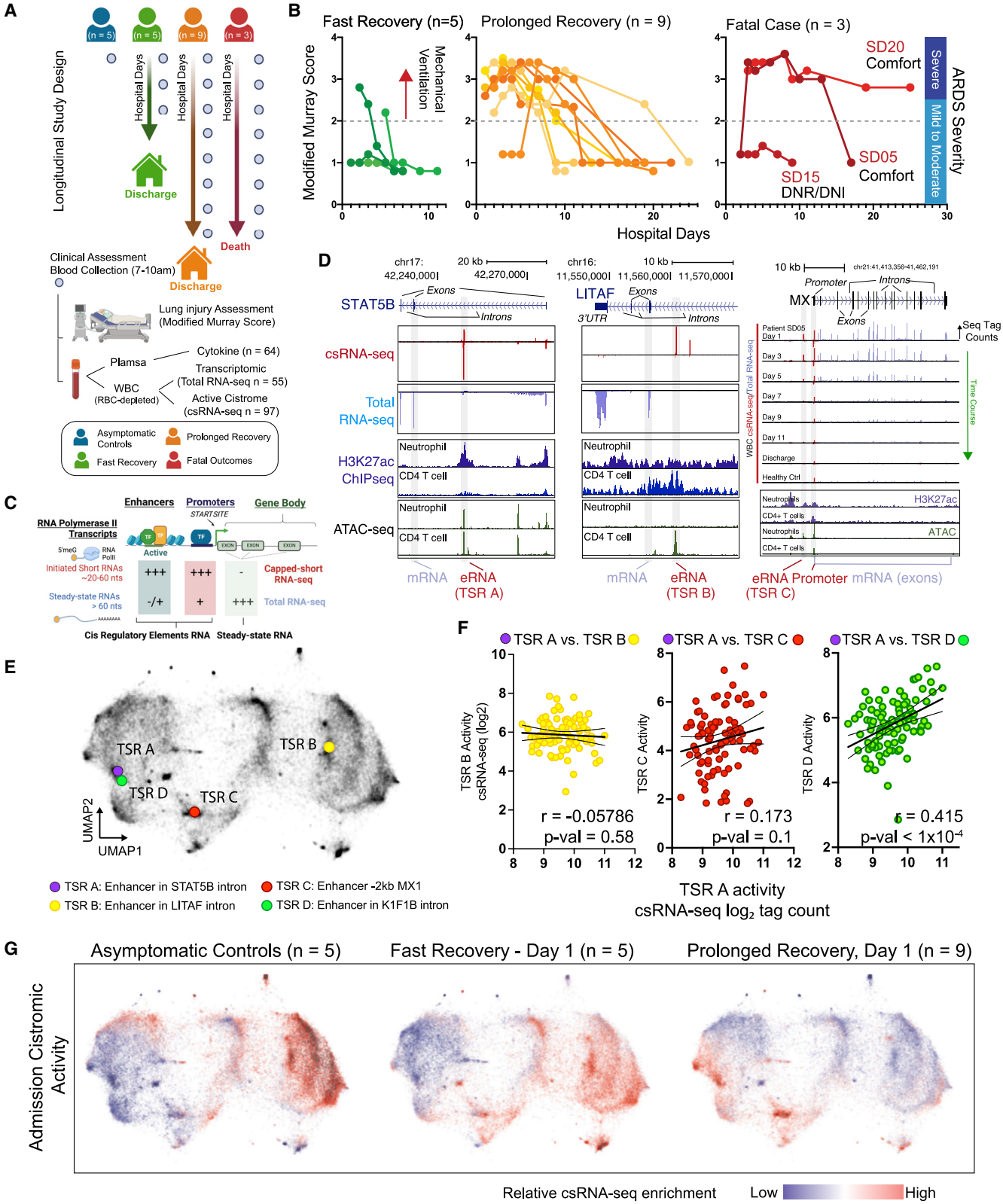
SUMMARY

Transcription factor programs mediating the immune response to coronavirus disease 2019 (COVID-19) are not fully understood. Capturing active transcription initiation from *cis*-regulatory elements such as enhancers and promoters by capped small RNA sequencing (csRNA-seq), in contrast to capturing steady-state transcripts by conventional RNA-seq, allows unbiased identification of the underlying transcription factor activity and regulatory pathways. Here, we profile transcription initiation in critically ill COVID-19 patients, identifying transcription factor motifs that correlate with clinical lung injury and disease severity. Unbiased clustering reveals distinct subsets of *cis*-regulatory elements that delineate the cell type, pathway-specific, and combinatorial transcription factor activity. We find evidence of critical roles of regulatory networks, showing that STAT/BCL6 and E2F/MYB regulatory programs from myeloid cell populations are activated in patients with poor disease outcomes and associated with COVID-19 susceptibility genetic variants. More broadly, we demonstrate how capturing acute, disease-mediated changes in transcription initiation can provide insight into the underlying molecular mechanisms and stratify patient disease severity.

INTRODUCTION

Acute respiratory distress syndrome (ARDS) is the cardinal clinical feature and the primary contributor to mortality in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection.[1–6] Cytokine and single-cell analyses support a model for a prolonged hyperinflammatory response driving diffuse alveolar damage.[7–14] Tremendous effort has been invested in repurposing therapeutics targeting the hyperinflammatory response to improve clinical outcomes in severe coronavirus disease 2019 (COVID-19).[15–18] The clinical benefits of these repurposed therapeutics, however, were only observed in patient subgroups, while therapies provided to inappropriate patient subpopulations may even cause harm.[15] This reflects the complexity and heterogeneity of ARDS and highlights the long-standing goal of subdividing ARDS into phenotypes based on distinct biological processes to identify the right patient for the right therapy.[19]

Understanding the dynamic relationship between transcription factor activity and disease severity may offer insights for precision therapy. Transcriptional responses are an important



(legend on next page)

component of the host response to infection. Not only do SARS-CoV-2-infected cells up-regulate antiviral gene expression programs to halt viral spread, but they also signal to activate regulatory networks in other cells and tissues to mount a coordinated immune response to the pathogen. Transcription factors (TFs) integrate upstream inflammatory and immunological signals to direct changes in the transcriptional programs mediating cellular adaptation and function. TFs bind DNA in *cis* at regulatory regions through specific recognition DNA sequences, also known as TF motifs.[20] Unbiased, genome-wide profiling of *cis*-regulatory elements, or the cistrome, coupled with computational analysis for motif enrichment, is a powerful tool to discover transcriptional regulatory mechanisms.[21] The analysis and interpretation lead to estimates of TF activity and thus provide unique insights into how transcription is regulated. Understanding what these TFs are, how their activities vary across individuals and time, and how they are dysregulated in severe disease is critical for understanding COVID-19 and the host response to sepsis and severe lung injury.

This study used unbiased cistrome profiling of peripheral leukocytes from COVID-19 patients to decipher regulatory networks activated or repressed during severe SARS-CoV-2 infection. Given that transcription is a hallmark of regulatory activity from promoters and enhancers,[22–25] we used capped small RNA sequencing (csRNA-seq) to measure the transcriptional activity of *cis*-regulatory elements on a genome-wide scale. csRNA-seq captures short, 5′-capped RNAs (20–60 nt) associated with engaged RNA polymerase II to define transcriptional initiation of stable and unstable transcripts such as enhancer RNA (eRNA). csRNA-seq is thus highly sensitive to changes in transcription, making it a superb assay for evaluating TF activity by linking DNA motifs to changes in cistromic activity.[26] Moreover, csRNA-seq focuses motif analysis on the active cistrome in contrast to assessing chromatin accessibility (i.e., assay for transposase-assessible chromatin with high throughput sequencing, ATAC-seq), where regulatory elements may be accessible but transcriptionally inactive (e.g., open-poised, insulators, etc.).[27] By revealing the genomic locations of individual regulatory elements, identifying TFs through their genomic binding sequences, and quantifying their activities, csRNA-seq provides unique interpretations for clinical specimens.

We designed a longitudinal trial measuring the cistromic activity of peripheral leukocytes from hospitalized COVID-19 patients with severe disease. This experimental design utilized individual patients' trajectories to construct a TF-centric dataset encompassing a wide range of disease states from peak disease to recovery. We profiled the active cistrome from 22 individuals with a median of seven time points encompassing early, mid, and late hospitalization. Our analysis revealed multiple regulatory programs associated with specific TFs correlated with clinical phenotypes. These TF programs include well-known inflammatory pathways as well as previously underappreciated TFs in the context of COVID-19. Using the expression of specific target genes to measure the TF regulatory network activity, we independently stratified COVID-19 patients with poor outcomes in a large published cohort with early admission transcriptomics. These findings showcase the utility of profiling transcription initiation to reveal regulatory programs from blood samples and provide insight into the key activated TFs during the host response to COVID-19.

RESULTS

Transcription initiation analysis reveals the active cistrome from peripheral leukocytes during severe SARS-CoV2 infection

We generated csRNA-seq libraries profiling transcriptional initiation activity at *cis*-regulatory elements in peripheral leukocytes of COVID-19 patients over hospitalization (Figures 1A and 1C). We collected blood from 17 patients and five healthy controls. The median number of samples per patient was seven (n = 92 disease samples). Of 17 patients, 16 required admission to the intensive care unit (ICU), and 13 needed support with mechanical ventilation for respiratory failure because of ARDS (Table S1). 48% of the samples were collected from patients with severe lung injury (modified Murray score > 2.5 [28]). Five patients were discharged home within 10 days (fast recovery, median hospitalization = 7 days), nine survivors had prolonged hospitalization (prolonged recovery, median hospitalization = 25 days), and three were non-survivors (Figure 1B).

From the patients' peripheral leukocytes, we cumulatively identified 93,465 genomic regions with evidence of transcription initiation, termed transcription start regions (TSRs) (Figures 1C and S1). More than 95% of TSRs overlapped with open chromatin regions defined by ATAC-seq in one or more leukocyte cell populations (Figure S1).[29,30] 58% of TSRs mapped to genomic

Figure 1. Differential cistrome activation in peripheral leukocytes of hospitalized COVID-19 patients

(A and B) Longitudinal study design for sampled plasma and peripheral leukocytes of patients across different stages of lung injury. Ninety-seven samples were included for active cistrome analysis using capped short RNA-seq (csRNA-seq). For fatal cases, the collection ended after patients transitioned to comfort care. One patient declined resuscitation or intubation (DNR/DNI).

(C) csRNA-seq captures short 5′-capped RNA from active *cis*-regulatory elements (gene promoters, enhancers), collectively termed transcription start regions (TSRs).

(D) csRNA-seq identifies transcriptional activity (red) in enhancers (eRNAs) at the *STAT5B* (left), *LITAF* (center), and *MX1* (right) loci. The *STAT5B* enhancer resides in a neutrophil-specific active chromatin region marked by acetylated histone H3K27 (H3K27ac+) and open chromatin (ATAC-seq+). The *LITAF* enhancer resides in a CD4 T cell-active chromatin region. *MX1* eRNA correlates with *MX1* gene expression over time.

(E and F) Unbiased clustering of 93,465 TSRs by activity similarity across 97 samples using uniform mani-fold approximation projection (UMAP). Positively associated TSRs (Pearson's correlation; e.g., TSR A and TSR D) clustered closer together (F). The lines represent simple linear regression with 95% confidence intervals.

(G) First-day enrollment leukocytes show globally distinct *cis*-regulatory activity between COVID-19 patients with fast (n = 5) and prolonged recovery (n = 9). Asymptomatic individuals (n = 5) serve as controls.

See also Figures S1 and S2.

regions carrying epigenetic marks associated with active enhancers (Figures S1A–S1C).[31] For example, the intronic regions of the *STAT5B* and *LITAF* loci have open chromatin regions with transcription initiation activity consistent with enhancer elements (Figure 1D, left and center panel). The interferon-induced *MX1* gene has two TSRs representing the gene promoter and a −2-kb enhancer (Figure 1D, right panel). Notably, the TSR activity of the *MX1* enhancer correlates (TSR C) with the *MX1* mRNA transcript level. We further evaluated the csRNA-seq sample characteristics using principal-component analysis (PCA). We did not observe sample separation based on individual patients or sequencing batches (Figure S1E). Rather, csRNA-seq samples were distributed largely by disease severity based on clinical status on the collection day (Figure S1F). Overall, we generated a dataset from peripheral leukocytes that identifies promoter and enhancer activities in leukocytes from patients with severe SARS-CoV-2 infection.

Agnostic TSR clustering contextualizes the activity of the immune cistrome with lung injury severity

We surmised that the change in cistromic activity across stages of disease severity reflects the underlying transcriptional regulatory mechanism(s). We hypothesize that analyzing genomic regulatory elements grouped by similar activity patterns provides mechanistic insights between transcriptional regulation and disease states. To this end, we performed unsupervised clustering to identify TSRs with similar activity patterns across all samples (Figure S2A) and visualized using uniform manifold approximation and projection (UMAP) (Figure 1E). Unsupervised clustering identified distinct congregations of TSRs with correlated activity patterns (e.g., TSR A and TSR D; Figure 1F), which we examined in relation to disease states and genomic sequence features.

We first interrogated the relationship of the cistrome activity with patient groups. We analyzed the first-day enrollment samples between patients with fast and prolonged recovery, contrasting their cistromic activity with asymptomatic controls (Figure 1G). The average signals for each TSR were overlaid onto the UMAP, showing a distinct pattern based on disease severity (Figure 1G).

To fully utilize our entire dataset, we next examined the relationship of cistromic activity to disease states, such as the severity of lung injury or multiorgan dysfunction. We evaluated the patient's clinical status at each sampling time point. Because the principal clinical diagnosis for our cohort is acute respiratory failure, we utilized the Murray lung injury score to quantify disease severity dynamics.[28] The lung injury score is a composite parameter of the radiographic disease burden on chest X-ray, hypoxemia (ratio of arterial oxygen partial pressure to fractional inspired oxygen, PaO_2/FiO_2), degree of respiratory support from the ventilator, and lung compliance; the latter three are daily measurements reflecting the patient's day-to-day clinical status. We then correlated each TSR's activity with the lung injury scores. The activity of the *STAT5B* intronic enhancer (TSR A), for example, has a positive correlation with lung injury (Figure 2A; rho = 0.3559, $p = 5 \times 10^{-4}$), indicating that its activity levels were generally higher in samples collected from patients experiencing severe lung injury. In contrast, the activity of the *LITAF* intronic enhancer (TSR B) was anticorrelated with lung injury (Figure 2B; rho = −0.4487, $p < 1 \times 10^{-4}$). We computed a lung injury correlation coefficient for all TSRs (Figure 2C; n = 93,465) and color-graded each TSR on the UMAP (Figure 2D). TSRs clustered in high concordance with lung injury severity. We corroborated the cistromic activity-lung injury relationship with sequential organ failure assessment (SOFA) and acute physiology and chronic health evaluation (APACHE-II) scores, two other clinical scores that estimate ICU mortality by encapsulating multiple organ dysfunction.[32,33] TSRs also clustered in high concordance with SOFA and APACHE-II scores (Figure S3). These analyses support the hypothesis that the cistromic activity pattern reflects disease severity.

Because *cis*-regulatory elements are often cell type specific,[34] we analyzed the cell type specificity of each TSR and related it to the spectrum of lung injury. We utilized a hematopoietic cell-type-specific reference ATAC-seq dataset from healthy donors to identify discrete cistrome clusters associated with open chromatin in neutrophils, plasmablasts, monocytes, and T lymphocytes (Figures 2Eand S2B).[29,30] Consistent with the elevated neutrophil-to-lymphocyte ratio observed in severe COVID-19,[35] TSR clusters associated with severe lung injury preferentially localized to neutrophil open chromatin, although TSRs in plasmablasts and lymphocytes are also associated with more severe lung injury (Figures 2D and 2E, violet). Notably, not all TSR activity patterns related to neutrophil open chromatin were positively correlated with lung injury (Figures 2D and 2E, asterisks), suggesting activation of discrete transcriptional programs in neutrophils that varied by disease state.

Lung injury severity-associated transcription factors in peripheral leukocytes

To gain further insight into the transcription regulatory mechanisms in severe COVID-19, we investigated the association of TF motifs in the active *cis*-regulatory elements in relation to disease severity. We ranked the TSRs based on the correlation of their activity with lung injury score, searched for TF motifs within −150 to +50 bp of the transcription start sites, and applied a logistic regression model (Figure 2F). Motif enrichment in ranked lists of peaks (MEIRLOP)[36] identifies TF motif enrichment in TSRs associated with the lung injury severity spectrum (Figure 2G).

Motifs enriched in lung injury-associated TSRs pointed to TFs associated with five different functional categories: response to virus and inflammation (signal transducer and activators of transcription [STAT], nuclear factor κB [NF-κB]), redox homeostasis (antioxidant-responsive element [ARE]), myeloid lineage determination (CEBP: AP-1 composite), lymphoid lineage determination (BCL-6), and proliferation (E2F, MYB). Interestingly, ARE is one of the top five motifs most strongly associated with severe lung injury. This motif is recognized by TFs involved in reduction-oxidation homeostasis, including members of the NRF, small MAF (sMAF), and BTB and CNC homology (BACH) families.[36] Genetic variants in the NRF member *NFE2L2* are associated with higher susceptibility to ARDS,[37] providing biological plausibility for this pathway in COVID-19 ARDS.

TSRs associated with low lung injury indices exhibited enrichment for the YY1 promoter element, glucocorticoid response element (GRE), X-box, and an ETS/RUNX composite motif recognized by lineage-determining TFs (LDTF) expressed in lymphocytes (Figure 3A). Dexamethasone is a pharmacological intervention that has shown significant benefit in severe

CellPress
OPEN ACCESS

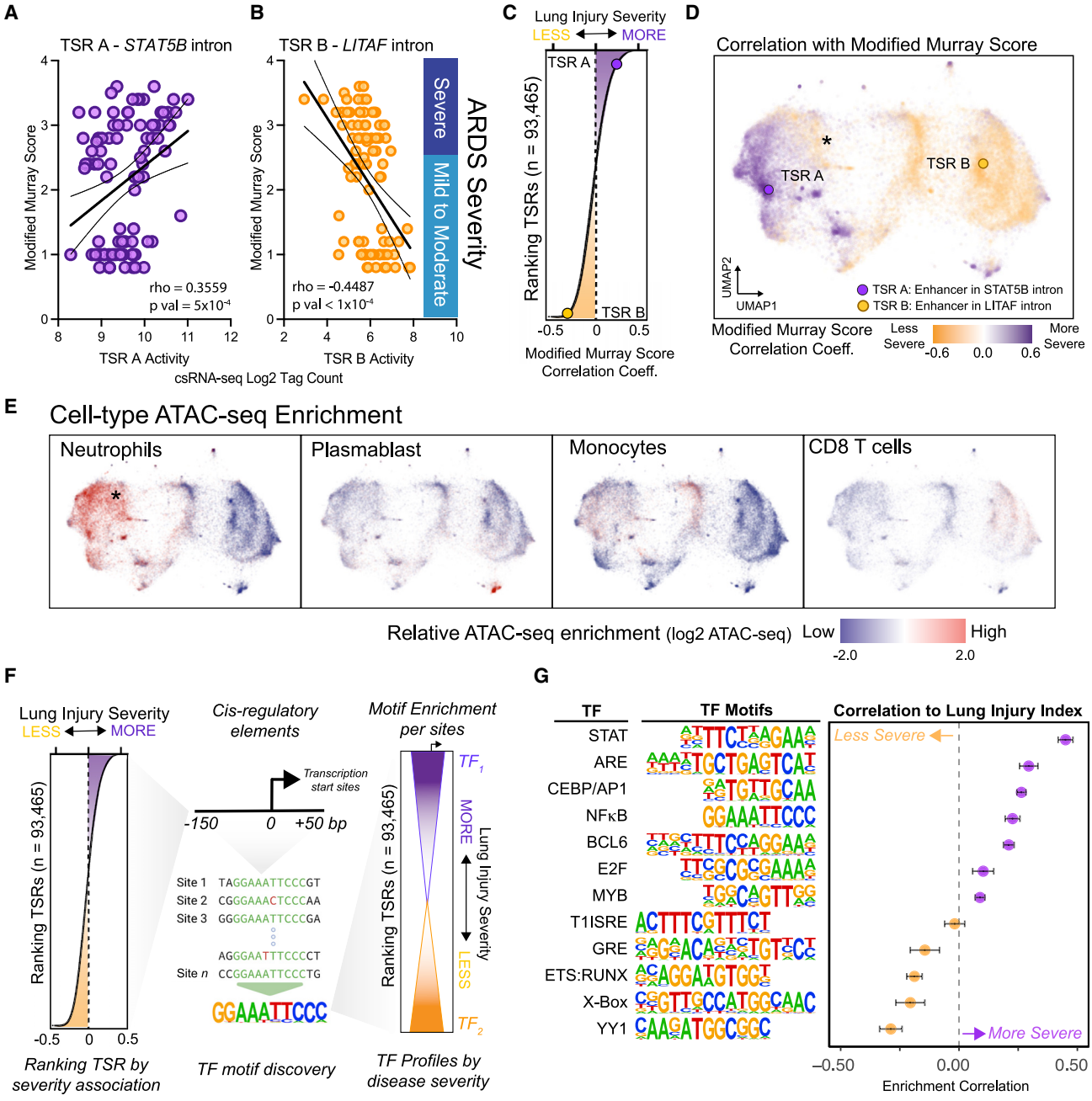


Figure 2. Lung injury-associated immune cistrome activation

(A and B) The enhancer activities in the *STAT5B* intron (TSR A) and *LITAF* intron (TSR B) are correlated with the modified Murray score for lung injury (Spearmen correlation, n = 92 samples with clinical scores). The lines represent simple linear regression and the 95% confidence intervals.

(C) A correlation coefficient was computed for each TSR (n = 93,465) based on its activity with lung injury scores.

(D) Each TSR in the UMAP is shaded based on the lung injury correlation coefficient (purple, positively correlated; gold, negatively correlated).

(E) Open-chromatin ATAC-seq enrichment in neutrophils, plasmablasts, monocytes, and CD 8 T cells for each TSR visualized on the immune cistrome UMAP. Part of a neutrophil's active cistrome is associated with lower lung injury scores (asterisks, D and E). Red delineates high ATAC-seq enrichment relative to other hematopoietic cell types based on Calderon et al.[29] and Perez et al.[30]

(F) Conceptual schematic using logistic regression analysis to identify transcription factor (TF) motifs associated with high (purple) or low (gold) lung injury scores (MEIRLOP).

(G) Each dot represents the enrichment coefficient of TF motifs in TSRs with activity correlated or anticorrelated with the lung injury score. Error bars represent the 95% confidence intervals. The enrichments of all motifs, except for T1ISRE, are statistically significant (adj. $p < 1 \times 10^{-4}$).

See also Figures S2–S4.

COVID-19.[15] Patients who received glucocorticoid therapy in this cohort had elevated GRE activity, reflecting a transcriptional response to treatment (Figure S4). Of note, glucocorticoid therapy was not a widespread treatment for COVID-19 during the time of this study, enabling us to observe changes in cistromic activity because of treatment.

Distinct *cis*-regulatory element clusters are enriched for specific TFs

The TF motif enrichment analysis based on lung injury severity ranking provided a broad perspective on the relationship between TF activity and disease. Because the TSRs are congregated into distinct groups based on the UMAP clustering analysis, we reasoned that focusing TF motif analysis for each TSR cluster would refine TF activity in relation to cell type specificity and disease severity. To this end, we performed TF motif analysis in sequences −150 to +50 bp from the transcription start site (TSS) for each of the 26 TSR clusters. Overall, we found specific combinations of TF motif enrichment for each TSR cluster (Figure S5), supporting the notion that TSR activity patterns are associated with specific TF activity.

Of interest, we found two distinct TSR clusters enriched for NF-κB motifs (Figure 3A, clusters A and B, green). The NF-κB motifs are enriched within −200 bp of the TSS of these clusters, as opposed to cluster C, which has specific enrichment for interferon-regulatory factors (IRF) rather than NF-κB (Figure 3A, cluster C, yellow). The first NF-κB cluster is myeloid centric (Figure 3A, cluster A) with 904 TSRs and coenriched for motifs bound by CEBP (red) and PU.1 (blue) proteins, the LDTFs for the myeloid compartment. The myeloid NF-κB cluster strongly correlates with severe lung injury (NF-κB myeloid; Figure 3C, top panel) and overlaps with the open chromatin of neutrophils (Figures 2E, S2B, and S2C). In contrast, the second NF-κB TSR cluster (Figure 3A, cluster B) is lymphocyte centric and anticorrelated to the lung injury score (NF-κB lymphoid; Figure 3C, top panel). The NF-κB lymphoid TSR cluster is co-enriched for the HEB motif, an E-protein critical for T cell development[38] (NF-κB lymphoid; Figure 3C, top panel). The enrichment for CEBP and PU.1 motifs is lower in the lymphoid NF-κB cluster than the myeloid (Figure 3A, clusters A and B, red and blue).

To verify that our analysis can distinguish NF-κB regulatory programs originating from diverse cell types, we examined public data on genome-wide NF-κB p65/RELA binding (chromatin immunoprecipitation sequencing [ChIP-seq]) in activated myeloid[39] and lymphoid cells[40] (Figure S6). TSRs in the myeloid NF-κB cluster had significant overlap with the NF-κB ChIP-seq signal from activated monocyte-derived macrophages. In contrast, the lymphoid NF-κB cluster had significant overlap with the NF-κB ChIP-seq signal from activated $CD4^+$ T cells (Figure S6B). Together, these findings demonstrate that motif analysis of co-regulated TSRs provides insights into activated pathways, the TFs that regulate them, and the cell types in which they are active.

Combinatorial activity of signal-determining TFs in severe COVID-19 acute lung injury

Using the NF-κB example as a model, we further dissected the TF enrichments in each TSR cluster with respect to disease severity. We identified co-enrichments of multiple signal-determining TF (SDTF) motifs within several clusters, suggesting a co-regulatory mechanism in the context of acute lung injury. This SDTF-SDTF co-regulatory network is exemplified by a TSR cluster co-enriched for motifs of type 1 interferon-sensitive responsive element (T1ISRE) and STAT, consistent with the known role of STAT1/2 in interferon signaling during viral infection (T1ISRE/STAT; Figures 3B and 3C, top panels).[41] Similarly, the myeloid NF-κB cluster exhibits enrichment for the RBPJ motif, the transcriptional regulator of the Notch signaling pathway that interacts with NF-κB in various inflammatory contexts[42–44] (NF-κB myeloid; Figures 3B and 3C, top panels). Last, in a TSR cluster enriched for STAT, the second most enriched motif is BCL6 (STAT/BCL6; Figures 3B and 3C, top panels), which has been shown to antagonize STAT activity.[45–47] Knowledge of these relationships has therapeutic potential because targeting single TFs may affect multiple pathways, while targeting cooperative TFs may achieve pathway specificity while reducing off-target effects.[48]

We also observed other TSR clusters with motif enrichment for multiple SDTFs, representing previously unrecognized SDTF-SDTF interactions, especially in the disease context of COVID-19. First, an enhancer-centric 1,296-TSRs cluster exhibited co-enrichment for motifs recognized by the autophagy-related factor MiT/TFE family (MITF, TFE3, TFEB, and TFEC),[49] the redox homeostasis-related motif ARE, and the inflammation-related NF-κB. Genes in this TSR cluster were functionally enriched for autophagy, lysosomes, and membrane trafficking (MiT/ARE/NF-κB; Figures 3B and 3C, center panels). Furthermore, regions containing ARE and MiT motifs are 3.4-fold higher in this TSR cluster relative to other active TSRs (chi-square $p < 1.0 \times 10^{-5}$, two-tailed), suggesting that the MiT/TFE and NRF/sMAF/BACH families are jointly activating these regulatory regions (Figure S7). The activity of this TSR cluster is positively correlated with lung injury (MiT/ARE/NF-κB; Figure 3C, center panel).

A second enhancer-centric, monocytic cluster with 1,235 TSRs is co-enriched for ARE, activator protein 1 (AP1), and SMAD factor motifs, the signal transducer for the transforming growth factor β (TGF-β) pathway that regulates genes essential for wound healing and pulmonary fibrosis in monocytes.[50,51] The activity of this TSR cluster is anticorrelated with lung injury severity (ARE/SMAD/AP1; Figures 3B and 3C, center panels; Figure S5).

We also identified co-enrichment for cell cycle and proliferation E2 factor (E2F) and MYB TF families in a cluster of 2,364 TSRs (E2F/MYB; Figures 3B and 3C, bottom panels). The activity of this cluster is highly correlated with lung injury severity (correlation = 0.36). Interestingly, the E2F motif was also enriched in five other clusters, each with different associations with the lung injury severity score (Figure S5). Thus, the unique association of the E2F/MYB cluster with severe lung injury suggests a synergistic role of the E2F and MYB pathways in severe ARDS.

We pursued two orthogonal approaches to validate the TF motif prediction for the TSR clusters. First, we cross-referenced active TSR clusters with publicly available TF ChIP-seq datasets. TSR clusters with enrichment of TF motifs overlap the ChIP-seq signals associated with the individual TFs compared with TSR

CellPress
OPEN ACCESS

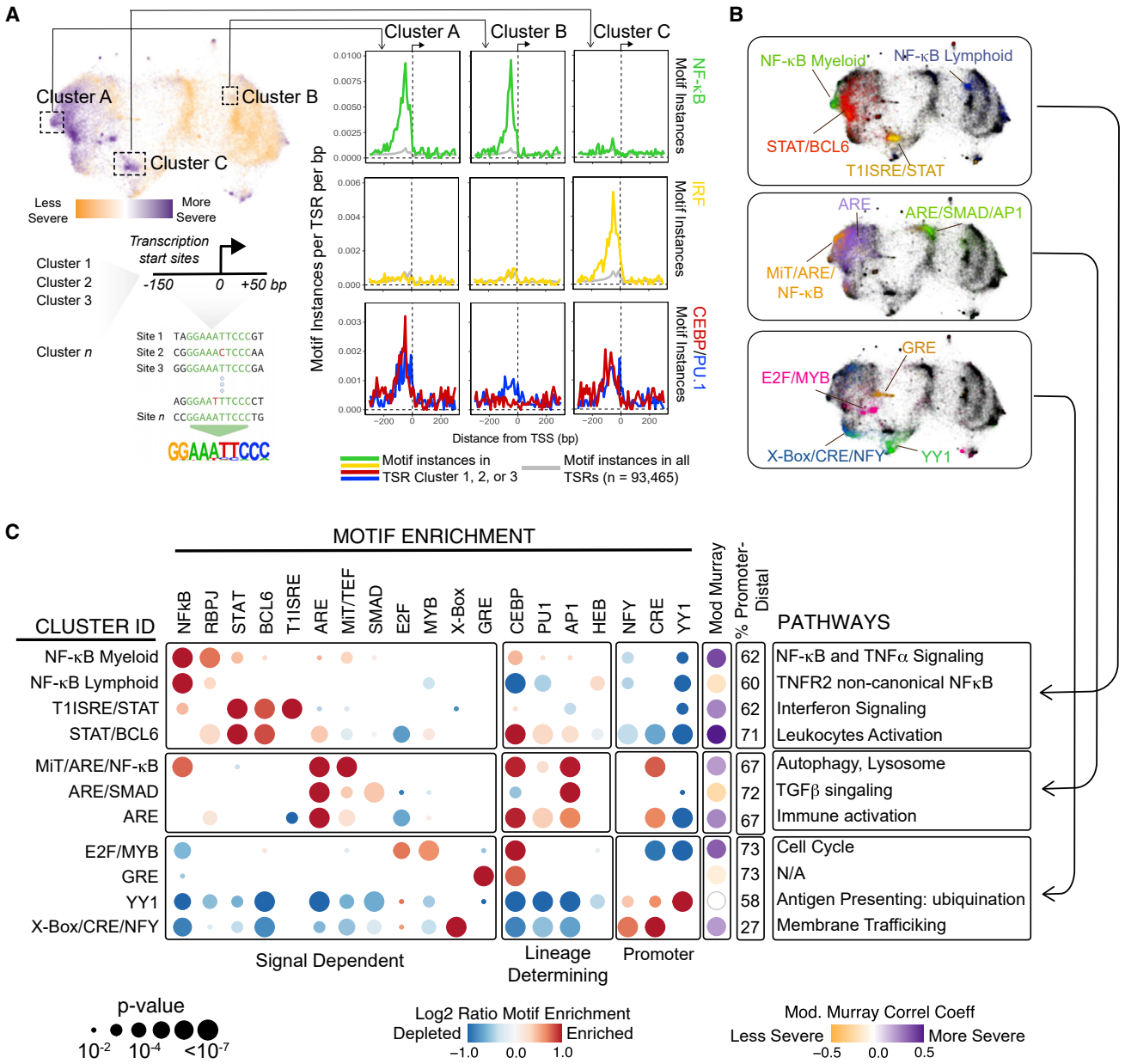


Figure 3. Distinct cistrome clusters identify co-enrichment of TF motifs suggestive of co-regulatory mechanisms

(A) Cluster-specific motif analysis shows enrichment of the NF-κB binding sequence upstream of transcription start sites (TSSs) in clusters A and B but not C (green histograms). The IRF motif is enriched in cluster C (yellow). The incidence of NF-κB or the IRF motif in the entire active cistrome was used as a control (n = 93,465, gray). Clusters A and C have increased enrichment for CEBP (red) and PU.1 motifs (blue), lineage-determining TFs (LDTFs) for myeloid cells.

(B) Discrete TSR clusters enriched for the representative TF motifs.

(C) Motif analysis depicting co-enrichment of signal-dependent, LDTF, and promoter TF motifs. Red indicates motif enrichment in the TSR cluster relative to all TSRs (log2 ratio) and blue depletion. The dot size represents the Fisher's exact test p value. Functional enrichment/gene ontology (GO) analysis identifies top pathways from genes associated with each TSR cluster.

See also Figures S5–S7.

clusters that lack motif enrichment (Figure S6C). Second, we tested whether genes in the TSR cluster overlap with the perturbation signature expected for the TF using Connectivity Map (Cmap). Cmap identifies gene sets regulated by small-molecule modulators of known signaling pathways for drug repurposing and discovery.[52] Cmap analysis revealed multiple cell cycle inhibitors for the genes in the E2F/MYB cluster, BCL and JAK/STAT inhibitors for the STAT/BCL6 cluster, TGF-β receptor inhibitors for the ARE/SMAD/AP1 TSR cluster, and cortisone for the GRE cluster (Table S2).

Overall, the combination of unbiased discovery of TSR clusters based on transcription dynamics with motif analysis revealed TF-regulated biological processes during severe SARS-CoV-2 infection.

The kinetics of transcriptional factor programs in survivors of severe COVID-19

The significance of day-to-day changes in *cis*-regulatory activity in relation to disease progression is understudied. With the longitudinal study design, we investigated the temporal profile of cistromic activity, reflecting global changes in peripheral immune TF activities through the course of severe COVID-19 infection. We observed heterogeneity in cistromic activity that parallels heterogeneity in clinical progression; patients with similar initial lung injury severity exhibited different dynamics in cistromic activities depending on the treatment, response, and outcome. For example, the cistromic activity for three patients who presented with severe COVID-19 ARDS requiring life support on mechanical ventilation (modified Murray score > 2.5), showed similarly elevated activity in STAT/BCL6 and T1ISRE/STAT on the first day of enrollment (Figure 4A). Patient SD07 had the shortest recovery course. SD07 was treated with prednisone, extubated from mechanical ventilation on day 6, and liberated from any respiratory support on day 9. TSR activities for STAT/BCL6 and T1ISRE/STAT were low on day 9 (Figure 4A, top panel). In the same time frame, patients SD09 and SD20 remained critically ill, requiring ongoing mechanical ventilation support (Figure 4A, center and bottom panels). However, for SD09, the TSR activity for STAT/BCL6 and T1ISRE/STAT decreased on day 9, and this patient slowly improved and was extubated on day 17. For SD20, the immune cistromic activity was essentially unchanged by day 9. SD20 did not show significant clinical improvement and was transitioned to comfort care on day 25 (Figure 4A, bottom panel). The cistromic activity and disease progress relationship is further emphasized by the complete TSR trajectories of these patients (Figure 4B). Using the median TSR levels as a surrogate for the cluster activity, patient SD09 had increasing STAT/BCL6 activity from days 1–5 before seeing a downtrend on day 9 (Figure 4B, blue). The STAT/BCL6 activities for SD20 were consistently elevated through the course (Figure 4B, red). The day-to-day changes in immune cistromic activity provide information pertinent to clinical progression.

The persistent elevation in STAT/BCL6 activity in a fatal case (SD20) highlights the importance of understanding the cistromic activity trajectory for survivors of severe disease. Deviations from the survivor trajectory may have significant clinical prognostic or therapeutic implications. To this end, we focused on the dynamic change in the active cistrome in survivors with severe COVID-19 ARDS with prolonged hospitalization. The nine survivors had similar outcomes with multiple repeated sampling, consisting of 63 csRNA-seq libraries representing a median of 22 hospitalization days with seven time points per patient. We used the median TSR activity to represent the overall activity of each TSR cluster. We observed a consistent reduction in T1ISRE/STAT activity per individual (Figure 4C) or as an aggregate cohort (Figure 4D), reflecting resolution of viral infection and interferon activity. The activity for STAT/BCL6 is more heterogeneous between individuals (Figure 4C) but globally decreased over hospitalization (Figure 4D). These patterns reflect the diverse clinical progression, with some individuals experiencing complications such as ventilator-associated bacterial superinfection (e.g., SD08, SD09). For independent corroboration, we profiled key serum cytokines associated with these TF pathways. The csRNA activity from the T1ISRE/STAT cluster paralleled the circulatory levels of IP-10, an interferon-induced cytokine associated with severe COVID-19[11,13] (Figure S8; Pearson's correlation = 0.860, $p < 1 \times 10^{-4}$). Similarly, STAT/BCL6 TF activity correlates with interleukin-6 (IL-6) plasma cytokine levels (Figure 4D; Pearson's correlation = 0.544, $p < 5 \times 10^{-4}$).

We next expanded our longitudinal analysis for the entire immune cistrome. We calculated a correlation coefficient for each TSR based on its activity to hospital admission day (Figure 4E) and overlaid this information on the UMAP (Figure 4F). Similar to the T1ISRE/STAT and STAT/BCL6 clusters, the activity levels are high for myeloid NF-κB, lymphoid NF-κB, cell proliferation E2F/MYB, and plasmablast OCT/MEF2C in early hospitalization (Figure 4E, blue). The monocyte ARE/AP1/SMAD and neutrophil cluster CEBP/PU.1 are associated with later time points (Figure 4E, yellow).

The inverse temporal profile of the neutrophil TSRs clusters (i.e., STAT/BCL6 versus CEBP/PU.1) indicates a transition in cistromic activity during recovery from COVID-19 ARDS. Indeed, genes associated with early-active neutrophil-specific TSR clusters were enriched for proinflammatory Gene Ontology terms associated with neutrophil activation (Figure 4F). Conversely, genes for the late neutrophil TF programs were enriched in homeostatic pathway terms (Figure 4F). When we compared the kinetics of neutrophil TF clusters with the patients' overall daily neutrophil counts, the decreases in STAT/BCL6 and T1ISRE/STAT activity correlate with a decrease in neutrophil counts (Figure S9). However, the temporal profile of CEBP/PU.1 activity differs from the neutrophil counts. Furthermore, the decreasing temporal activity of myeloid-NF-κB has a distinct pattern that cannot be explained by neutrophil cell count alone (Figure S9). These findings show that longitudinal profiling can identify distinct TF programs acting at different phases of disease resolution independent of the cell population composition in the blood.

Distinct TF regulatory networks overlap genetic variants associated with COVID-19 clinical outcomes

We next evaluated the clinical significance of TF activity with COVID-19 outcomes. Because csRNA-seq maps the functional non-coding regions of the genome, we assessed the overlap of risk-associated genetic variants to the vicinity of active *cis*-regulatory elements, hypothesizing that clinically relevant TF regulatory networks have higher enrichment of genetic variants associated with COVID-19 outcomes. We identified COVID-19-associated single-nucleotide polymorphisms (SNPs) from the COVID-19 Host Genetic Initiative Consortium[53] that overlap within −300 to +100 bp of the TSS of the active cistrome. The SNP (rs34460587) in the *LZTFL1* locus in chromosome 3p21.31, for example, overlaps with a TSR that has an activity pattern grouped in the STAT/BCL6 TSR clusters (Figures 5A and 5B). We assessed the significance of the overlap between genetic variants and the

CellPress
OPEN ACCESS

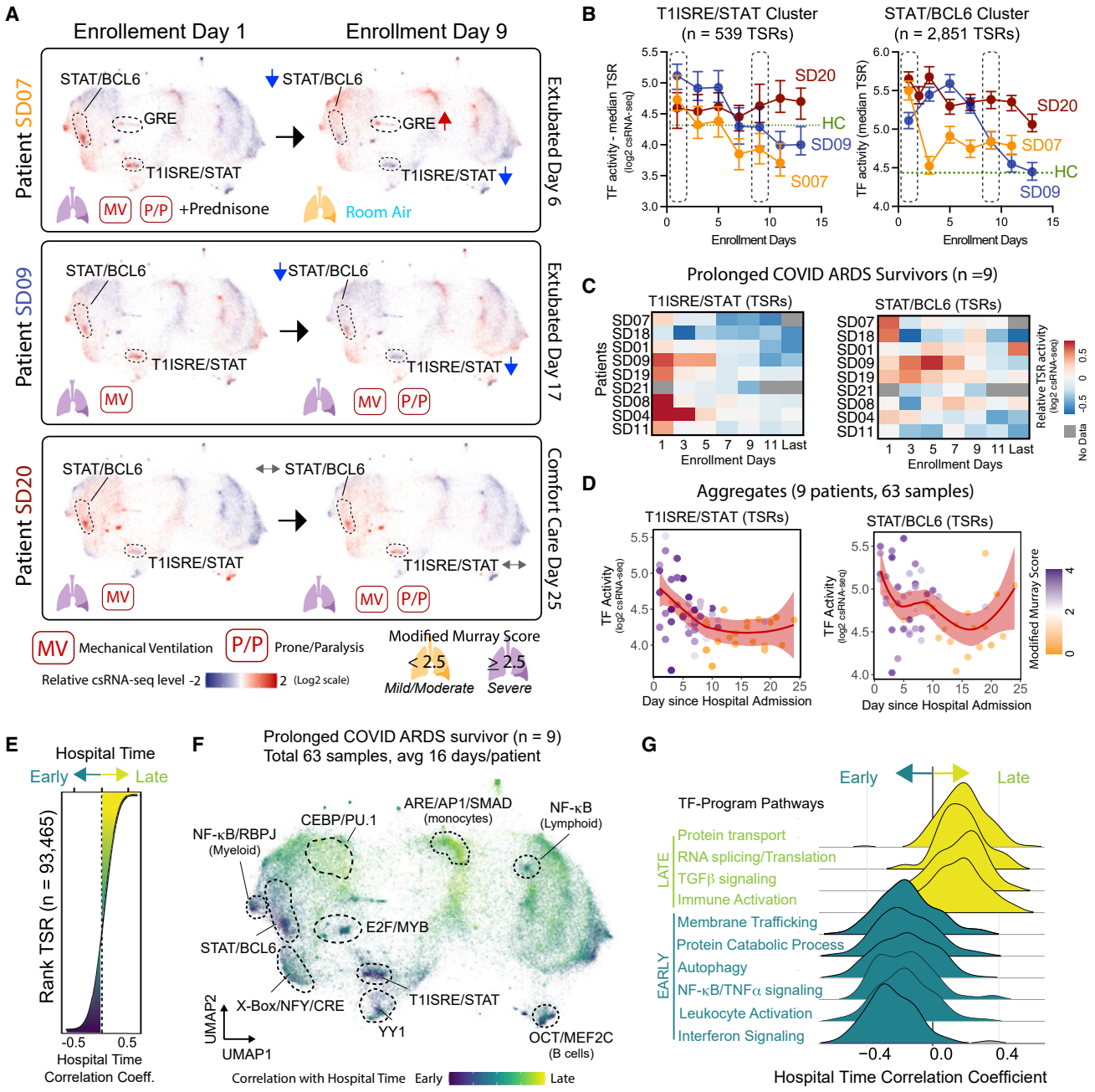


Figure 4. The kinetics of transcriptional factor activity during the hospitalization course of COVID-19

(A) *cis*-regulatory activity from three COVID-19 ARDS patients on enrollment days 1 and 9.

(B) Full-time-course activities for T1ISRE/STAT and STAT/BCL6 are represented by the median log2 TSR signals and the 95% confidence intervals (CIs). Dotted boxes represent day 1 and 9. The medians for healthy controls (HCs) are shown as green dashes.

(C) Heatmap of individual survivors' time courses for T1ISRE/STAT and STAT/BCL6 activity. Red, high; blue, low; gray, no data.

(D) Aggregate activity for T1ISRE/STAT and STAT/BCL6 TSRs (n = 9 patients, total 63 samples). Each point represents the median log2 of TSR clusters. The color of each point indicates the lung injury score (purple, severe lung injury; gold, less severe). The line and shaded region correspond to the smooth conditional mean and 95% CIs, respectively.

(E) A correlation coefficient was computed for each TSR (n = 93,465) based on its activity with the hospital time of the prolonged survivors.

(F) Each TSR in the UMAP is shaded based on the hospital time correlation coefficient (blue, early hospital course; yellow, late).

(G) Gene pathways enriched in the early and late TF programs. A ridge plot shows the time-TSR activity correlation coefficient of genes in the respective pathways.

See also Figures S8–S10.

TSR clusters while considering the underlying genetic structure of the data using regulatory element locus intersection (RELI).[54] We focused on SNPs associated with hospitalization, with non-hospitalized COVID-19 cases as controls, mirroring the focus of our COVID-19 ARDS cistromic data. TSRs positively correlated with lung injury indices (modified Murray score coefficient > 0.15) have significant enrichment for disease-associated SNPs (Adjusted p value, adj p, 2.57×10^{-13}) (Figure 5D). The E2F/MYB, STAT/BCL6, and T1ISRE/STAT clusters are also significantly enriched for COVID-19-associated SNPs (adj p 7.62×10^{-7}, 1.67×10^{-9}, and 2.16×10^{-13}, respectively). The late-onset monocytic TSR cluster with ARE/SMAD/AP1 motifs is also enriched for SNPs associated with COVID-19 hospitalization (adj p 5.12×10^{-11}). In sum, enrichments of disease-associated SNPs within *cis*-regulatory clusters provide evidence that immune regulation through E2F/MYB, STAT/BCL6, T1ISRE/STAT, and ARE/SMAD/AP1 may influence COVID-19 disease outcomes.

Cistrome-disease relationship reveals dysregulated E2F/MYB, STAT/BCL6, and T1ISRE/STAT activity

Reflecting on the genome-wide association study (GWAS) analysis, we hypothesized that the activity of the TF network enriched for disease-associated genetic variants is likely correlated with disease outcome. Consistent with this notion, the deceased patients in our cohort had persistently elevated STAT/BCL6 and E2F/MYB activities compared with survivors with prolonged hospital courses (Figure S10). To extend these observations and extrapolate the generalizability of our findings, we expanded our cistrome-based TF network analysis to gene expression, assuming that specific genes associated with TSR clusters can be used to quantify TF activity (Figure 6A). We focused on the E2F/MYB, STAT/BCL6, and T1ISRE/STAT regulatory networks because of the statistical significance of risk variant overlaps and activity early in hospitalization. We performed total RNA-seq from samples matching the csRNA-seq (n = 55) and identified genes specific for E2F/MYB, STAT/BCL6, and T1ISRE/STAT (n = 106, 176, and 58 respectively; STAR Methods). The median gene expressions specific for E2F/MYB, STAT/BCL6, or T1ISRE/STAT clusters were highly correlated with the median TSR activities for those clusters (r = 0.92, 0.89, and 0.93, respectively) (Figures 6B and S11), supporting a strong association between leukocyte RNA and the cistrome at the level of TF programs.

Using this approach, we evaluated the correlation of E2F/MYB, STAT/BCL6, and T1ISRE/STAT activities with disease severity in a large independent COVID-19 cohort (n = 100) with blood leukocyte transcriptomics isolated soon after hospital admission.[55] This cohort used the number of hospital-free days on the 45th day after admission (HFD45) to delineate clinical severity – severe cases with prolonged hospitalization have fewer HFDs. E2F/MYB and STAT/BCL6 activity independently correlated with disease severity, with higher TF-network activity corresponding to lower numbers of HFDs (Figure S12). T1ISRE/STAT activity did not show a linear relationship with disease severity; rather, patients at each extreme of T1ISRE/STAT activity trended toward poor disease prognosis (Figure S12).

Because of the non-linear relationship of T1ISRE/STAT activity and disease outcome or the variability when using STAT/BCL6 or E2F/MYB activity alone, we queried whether interactions between TF networks further stratified patients with different disease severity. The cohort was divided into "high" and "low" groups based on the median of the TF network activity and compared for each TF network pair (Figures 6C and S13D). The STAT/BCL6^{hi} E2F/MYB^{hi} group included significantly more fatal (HFD45 = 0, chi-square, p = 0.004, two-tailed) and low-HFD45 cases (Figures 6D and S13). When delineated by T1ISRE/STAT and E2F/MYB activity, the four groups have no statistically significant differences in fatality (chi-square, p = 0.47, two-tailed) or HFD45 (ANOVA, adj p = 0.34) (Figure 6F). T1ISRE/STAT and STAT/BCL6 exhibited significant interaction. The T1ISRE/STAT^{lo} STAT/BCL6^{hi} group (n = 14) included the highest fatalities and lowest average HDF45 score (Figures 6E and S13). The T1ISRE/STAT^{hi} STAT/BCL6^{hi} group also had low HFD45, explaining that the extremes of T1ISRE/STAT activity correlate with more severe disease (Figure S12). Furthermore, by using the same approach in a COVID-19 single-cell RNA-seq dataset, we found that high STAT/BCL6, high E2F/MYB, and low T1ISRE/STAT activity were specifically observed in immature neutrophils and a distinct mature neutrophil population, and both were more abundant in severe COVID-19 patients (Figure S14).[14] Prior studies have demonstrated an association of severe COVID-19 with emergence of immature neutrophils.[12,14,56] Our analysis showed that gene targets specific for the STAT/BCL6 and E2F/MYB programs accounted for approximately 50% and 10% of the transcripts in immature neutrophils, respectively (Figure S12H), highlighting the TF programs driving emergence of developing granulocytes in critical illness. Our findings are also consistent with the current literature on interferon dysregulation in COVID-19,[9,14,57] but delineate patients with combined high STAT/BCL6 and low T1ISRE/STAT activity as the most vulnerable to poor clinical outcome.

DISCUSSION

We report here a longitudinal study characterizing the change in the immune *cis*-regulatory activity of ARDS patients associated with COVID-19. Our analysis of the regulatory landscape catalogs the dynamic regulation of eRNAs and provides a TF-centric analysis and interpretation. Because the identity of the TF is revealed through the genomic DNA sequence, csRNA-seq coupled with motif analysis is, in essence, an unbiased functional assay for TF activity[58] and provides a dataset that is substantially different and complementary to traditional transcriptomics or other types of epigenetics profiling (e.g., ATAC-seq). Active cistromic analysis provides unique insights into the underlying TF networks and mechanisms in complex diseases when integrated with GWAS and bulk and single-cell transcriptomics.

With this approach, we identified pathways and TFs implicated in severe COVID-19, including known therapeutic targets such as the glucocorticoid receptor,[15] the interferon pathway,[59] and the JAK/STAT pathway, including IL-6 antagonist[16,60] and JAK inhibitors.[17,18] Our analysis identified TFs in the immune response to severe COVID-19 and implicated their activity in neutrophils, including motifs for antioxidant response elements involved in oxidative homeostasis with NRF, sMAF, and BACH. The genetic association of ARDS with the NRF family *NFE2L2*

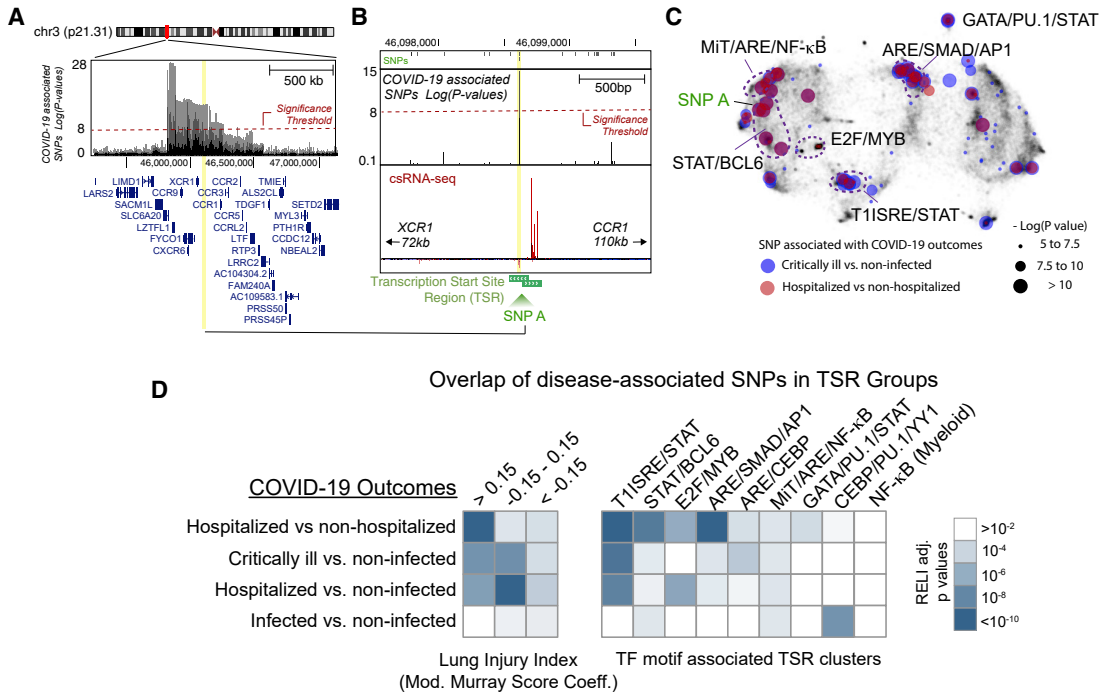


Figure 5. Distinct TSR clusters exhibit significant enrichment of single-nucleotide polymorphisms (SNPs) associated with COVID-19 clinical outcomes

(A) The *LZTFL1* locus in chromosome 3p21.31 exemplifies the overlap of COVID-19-associated SNPs ($p < 5 \times 10^{-8}$) with *cis*-regulatory elements detected by csRNA-seq.

(B) SNP A (rs34460587, $-\log p > 15$ for hospitalization versus non-hospitalized COVID-19 cases) lies within −300 to +100 bp of the TSS located in the intergenic region between the *CCR1* and *XCR1* genes.

(C) UMAP showing the global distribution of TSRs overlapping COVID-19-associated SNPs.

(D) TSRs associated with higher lung injury or TF clusters (T1ISRE/STAT, STAT/BCL6, E2F/MYB, and ARE/SMAD/AP1) have significant overlaps with disease-associated SNPs.[53] The color represents the adjusted p values, determined by regulatory element locus intersection (RELI), that account for the underlying genetic structure.[54]

in patients[37] and higher mortality because of bacterial pneumonia[61,62] or, more significant, acute lung injury because of high tidal volume[63] in mice lacking *Nfe2l2* supports the biological plausibility of our findings.

One key implication of this study is identification of co-TF regulatory networks in the context of systemic disease. With csRNA-seq and unsupervised clustering, we identified *cis*-regulatory elements and genes clustering into distinct groups because of similarities in transcriptional initiation activity. Co-enrichment of TF motifs within a single cluster suggests cooperative regulation.[64] Our dataset revealed networks with co-enrichment of (1) E2F and MYB; (2) MiT/TFE, ARE, and NF-κB; (3) NF-κB and Notch; (4) STAT and BCL6; and (5) ARE, SMAD, and AP-1. The patient's active immune cistrome thus provides evidence of unrecognized interactions between otherwise well-described pathways. The convergence of ARE, NF-κB, and MiT/TEF motifs in a single TF network suggests biological significance in the interaction of redox, inflammation, and autophagy during COVID-19 ARDS.[49,65,66] Similarly, the co-enrichment of ARE, SMAD, and AP-1 in monocytes in the late disease course of prolonged survivors suggests co-regulation of TFs in the TGF-β and redox pathways. Consistent with this, COVID-19 patients with pulmonary fibrosis harbor monocyte-derived alveolar macrophages enriched for profibrotic programs, including the TGF-β pathway.[51] Our finding suggests that peripheral monocytes are primed for a wound-healing program through co-regulation of TGF-β and redox pathways, shown previously in macrophages during wound healing.[50] This finding of co-TF enrichment in TSR clusters provides a collaborative model where TFs of different families converge in response to combinatorial biological signals in the cellular milieu.

Our longitudinal study offers insights into TF activity trajectories of severe COVID-19 survivors. We already observed deviation of E2F/MYB and STAT/BCL6 activity from two fatal patients (Figure S10). More longitudinal studies are needed to confirm and capture the heterogeneity in the trajectories of survivors and fatal cases. While the research fields for ARDS and sepsis are finding success in applying molecular assays for stratifying and endotyping patients early during their hospital admission,[67,68] the field is still in its infancy in designing, analyzing, and interpreting longitudinal trials with molecular data. The lack of molecular guidance in mid- to late disease is a crucial unmet need in critical care medicine. We believe our study contributes to this field, complementing some key studies[69,70] while presenting the active cistrome as a completely novel approach.

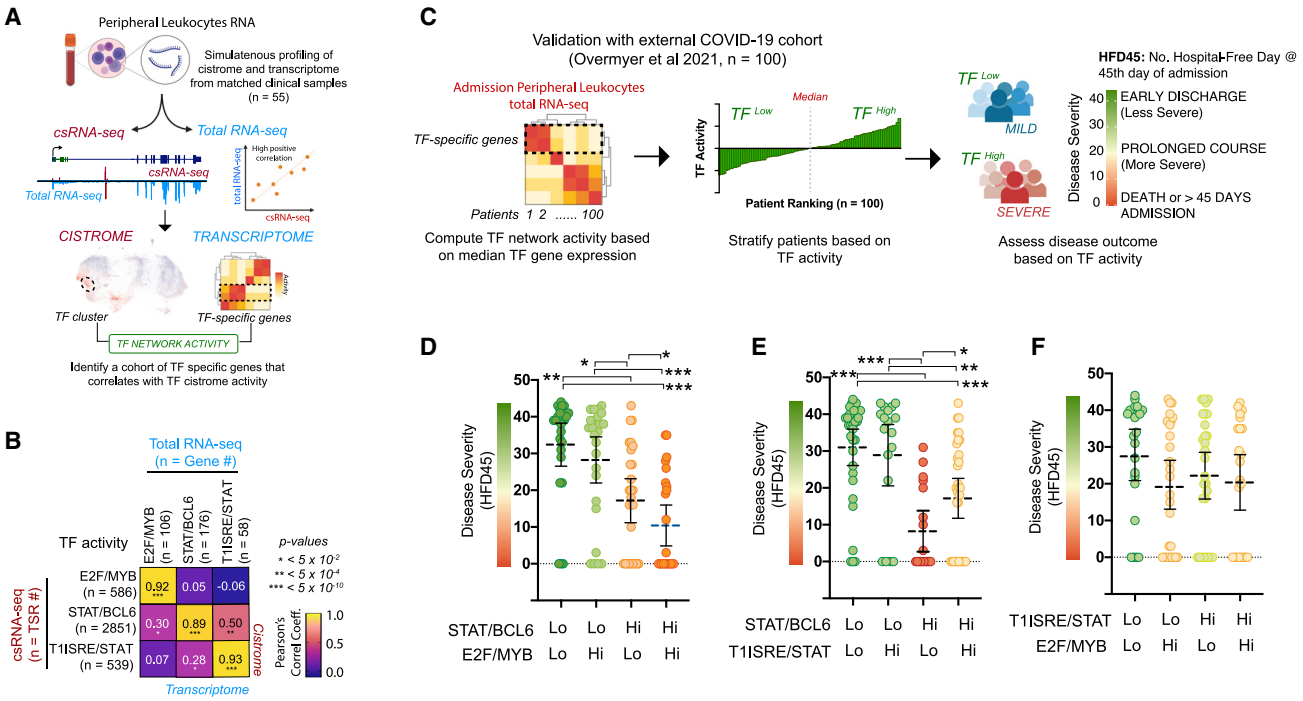


Figure 6. Dysregulated E2F/MYB, STAT/BCL6, and T1ISRE/STAT programs are associated with severe COVID-19

(A) csRNA-seq and total RNA-seq performed on matched RNA from peripheral leukocytes of COVID-19 patients (n = 55). To represent TF activity, genes specific to the TF network with expression highly correlated with the TSR cluster activity (csRNA-seq) were selected.

(B) The median TF-specific gene expression positively correlated with the median cistromic activity for the E2F/MYB, STAT/BCL6, and T1ISRE/STAT clusters, respectively. Pairwise Pearson's correlation analysis also shows specificity in TF activity.

(C–F) Combinatorial TF activity signatures are associated with COVID-19 outcomes using a published validation cohort (n = 100).[55] TF activities for STAT/BCL6, E2F/MYB, and T1ISRE/STAT were computed and ranked with bulk leukocyte RNA-seq (C). Hospital-free day 45 (HFD45) delineates clinical severity. TF activity signatures of (D) high STAT/BCL6 and high E2F/MYB or (E) high STAT/BCL6 and low T1ISRE/STAT correlate with the highest clinical severity. Combination of T1ISRE/STAT with E2F/MYB does not correlate with severity outcome (F). Error bars represent 95% CIs from the mean (horizontal lines). One-way ANOVA with adjusted p values; *p <0.05, **p < 0.005, ***p < 0.0005.

See also Figures S11–S14.

While a disease-associated genetic variant localizes the disease risk to a genomic locus, enrichment of disease SNPs in multiple regulatory elements within one TF network indicates the biological significance of the TF pathway.[54] We identified over-representation of COVID-19-associated SNPs at TSRs regulated by the T1ISRE/STAT, STAT/BCL6, and E2F/MYB networks. SNP enrichment in the T1ISRE/STAT cluster emphasizes functional variants in the regulatory regions, complementary to the ongoing effort of identifying coding sequence variants in type I interferon pathway genes.[71] We further demonstrated that the activity of these TF regulatory networks parallels disease outcomes using two independent COVID-19 cohorts.[14,55] While our dataset cannot provide evidence of causal genetic variants, integrating the cistrome, GWAS, and transcriptomic analyses supports functional roles of co-regulatory TF networks.

In summary, we methodically examined active genomic regulatory elements by profiling levels of initiating transcripts directly from peripheral leukocytes of critically ill COVID-19 patients (Figure 7). We identified and provided evidence of TF pathways and co-regulatory mechanisms implicated in severe COVID-19 ARDS. Furthermore, we demonstrated the feasibility of using a cistromic approach to assess TF activity, stratifying disease severity and patient outcomes. We also observed treatment response (Figure S4). Knowing the patient's TF activity profile may guide patient selection and provide clinical insights into treatment variations, especially for targeted molecular therapies. Our work thus provides a TF-centric assessment that complements the ongoing efforts of endotyping patients with critical illnesses.[67,68,72] Our approach may also apply to non-COVID-19 ARDS, a heterogeneous clinical syndrome with high mortality (35%–40%) that depends on supportive care without targeted pharmacological therapy.[73] Our study demonstrates that unbiased active cistrome profiling offers an unprecedented TF-centric resolution for understanding human disease.

Limitations of the study

Our work has several limitations. First, the power of this longitudinal study lies in the trajectories of the individual patients, which limits patient numbers. A dedicated study with a larger cohort for csRNA-seq would be ideal for confirmation.

Second, we profiled the cistrome of all peripheral leukocytes, a heterogeneous population of cells. The imbalance in cellular proportion influences clustering resolution, which is more

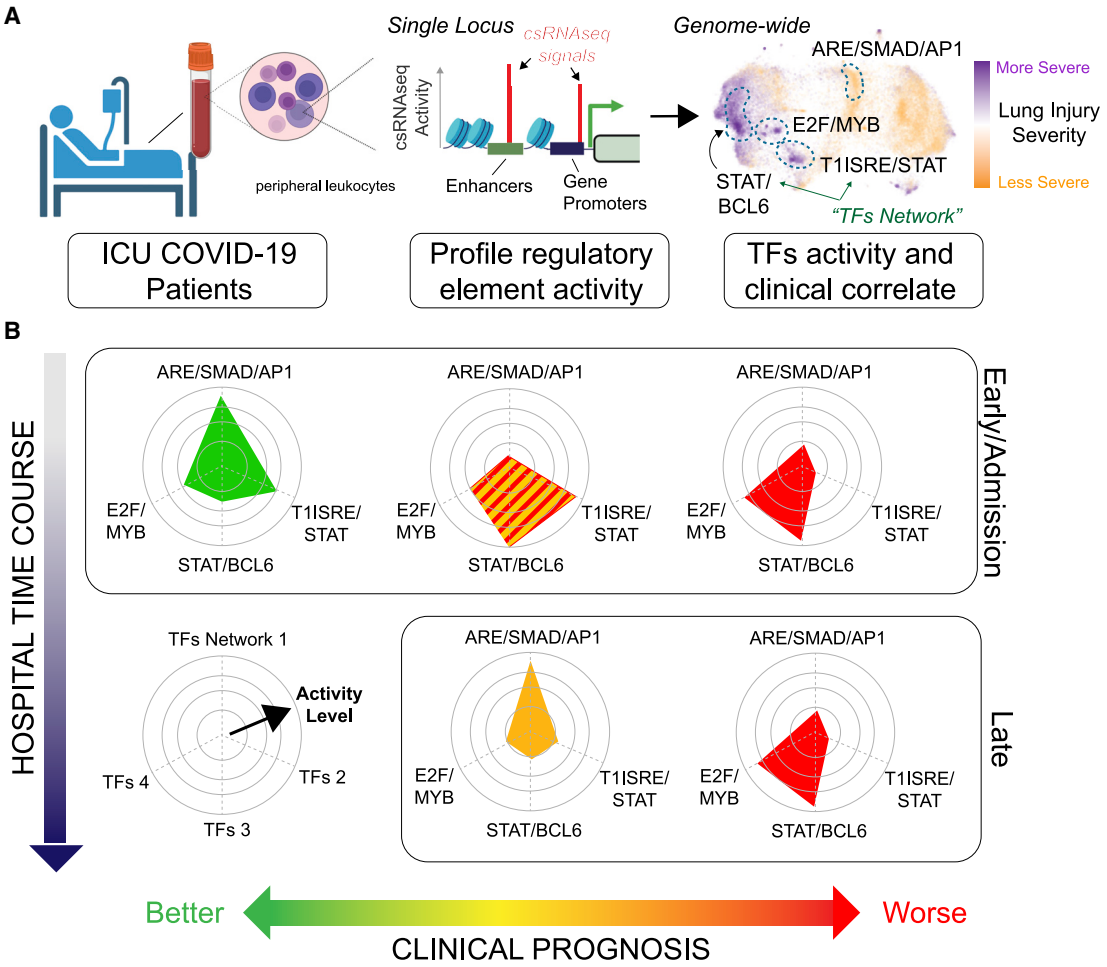


Figure 7. Summary

(A) The active cistromes from patients' leukocytes were profiled using csRNA-seq. Distinct clusters of *cis*-regulatory elements with co-enrichment of TF motifs, or "TF networks," track clinical disease severity.

(B) The activities of the TF networks in early or late hospital progression prognosticated outcomes.

sensitive to cells making up most of the population. Cell sorting before cistrome analysis would address this issue but presents technical, feasibility, and ethical challenges requiring larger blood volumes from clinically unstable patients.[74] We cross-examined publicly available cistrome databases and successfully identified major inflammatory pathways in smaller subsets of circulatory immune cells, including the lymphocytic NF-κB and the monocytic ARE/SMAD/AP-1 programs. We also identified the combined STAT/BCL6 and E2F/MYB signature from immature neutrophils, which usually represent less than 10%–20% of total leukocytes even in critical illnesses. We also demonstrated that the changes in TF signal are not merely based on cell abundance (Figure S9). Nonetheless, dedicated single-cell RNA[8,12,14,75] and chromatin studies[56,76] are unmatched in resolution at the cell type level. Our study design focuses on detecting active regulatory elements, revealing insights into active TF programs in peripheral leukocytes of patients transitioning between different stages during critical illness.

While we found the dynamic clinical scores, such as the Murray lung injury score, SOFA, and APACHE-II, useful to interpret longitudinal data, these scores have shortcomings. Palliative and end-of-life decisions may influence these scores. For example, the severity score increases when patients require mechanical ventilation. For patients with prior "do not resuscitate/do not intubate" wishes (Figure 1B, patient SD15 with "DNR/DNI") or extubated compassionately for comfort care (Figure 1B, patient SD05 with "Comfort"), the severity score lowers by criterion. It may not reflect disease severity at the molecular level.

Moreover, our study is observational and based on a patient cohort with COVID-19 pneumonia as the principal diagnosis. Our observed cistromic activity can represent COVID-19-specific responses or a more general response to viral infection, sepsis, and organ injury. Future cistrome-based studies that include lung injury or sepsis from other acute viral infections or inflammatory etiologies are warranted to delineate specific TF responses.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Patient cohorts
- METHOD DETAILS
  - Characterizing lung injury with modified Murray Score
  - Multiorgan dysfunction assessment with SOFA and Apache-II scores
  - Blood sample processing
  - Peripheral leukocytes isolation
  - Plasma isolation
  - RNA extraction and purification
  - Capped small RNA-sequencing
  - Total RNA sequencing
  - Cytokine measurements
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Principal component analysis
  - Unsupervised machine learning for TSR cluster identification
  - Motif enrichment analysis
  - Integration of csRNA-seq data with hematopoietic NGS epigenome profiling
  - Overlap of disease-associated SNPs from COVID-19 GWAS with TSRs
  - Target genes selection and pathway analysis
  - Calculating TF-network activity from csRNA-seq
  - Target gene selection to calculate TF-network activity from transcriptomes
  - Integrated analysis with single-cell RNA-seq data
  - Drug repurposing & connectivity mapping

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2023.100935.

ACKNOWLEDGMENTS

We are grateful to patients and their families for participation. We also thank the nurses in the UC San Diego-affiliated Medical Center ICU and the step-down unit for assistance. We thank Dr. Christopher K. Glass and Dr. Sven Heinz for advice. This work was supported by the NGS Core Facility of the Salk Institute with funding from NIH-NCI CCSG: P30 014195, the Chapman Foundation, and the Helmsley Charitable Trust. M.T.Y.L. is supported by the Academic Sleep Pulmonary Integrated Research/Clinical Fellowship through the American Thoracic Society, NIH (5T32HL134632-04), and Veterans Affair (1IK2BX005908). M.F.O. is supported by the NIH/National Institute of General Medical Science (T32GM121318). R.K.A. is funded by NIH K99/R00 AI145762. S.C. is funded by NIH/Cooperative Centers on Human Immunology (U19AI142742). S.H.D. is supported by NIH (R00 GM135515). C.W.B. is funded by the NIH (U19A135972, R01GM134366, U01DA051972, and U01AI150748). S.P. is supported by the Wu Tsai Human Performance Alliance and the Joe and Clara Tsai Foundation and an unrestricted gift from Barbara Friedman. E.A.H. received funding from EDUC2-08388 Cal Poly San Luis Obispo. N.G.C. is funded by the Doris Duke Foundation (2020081) and by NIH K08 NS109200. Selected figures were generated using BioRender. This study was approved by the UC San Diego Institutional Review Board (IRB-190699).

AUTHOR CONTRIBUTIONS

C.W.B., M.T.Y.L., and N.G.C. conceived the study. J.F. and N.G.C. oversaw the trial's logistics. M.F.O., M.T.Y.L., and N.G.C. identified, enrolled, and obtained consent from patients. M.F.O., M.T.Y.L., M.R., M.H., M.A., and K.P. collected patient samples. E.A.H., C.T.N., S.T., R.K., M.T.Y.L., S.D., and N.G.C. processed samples. M.F.O., M.T.Y.L., A.P., L.E.C.A., K.P., and N.G.C. gathered and interpreted the clinical data. S.H.D. generated csRNA-seq libraries. H.D.L. generated total RNA-seq libraries. M.W.C. processed the sequencing data. C.W.B., M.W.C., M.T.Y.L., and S.H.D. analyzed and interpreted the csRNA-seq and total RNA-seq dataset. S.I.R., J.M.D., R.K.A., and S.C. performed the cytokine analysis. A.S.W. and N.G.C. performed the CMap analysis. A.M. and S.P. provided critical feedback on the study design and the manuscript. N.G.C., M.T.Y.L., and C.W.B. interpreted data and wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

A.M. is funded by the NIH. He reports income from Livanova, Equillium, and Corvus related to medical education. ResMed provided a philanthropic donation to UC San Diego. S.C. has consulted for Avalia, Roche, and GlaxoSmithKline.

INCLUSION AND DIVERSITY

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure ethnic or other types of diversity in the recruitment of human subjects. We support inclusive, diverse, and equitable conduct of research.

Received: March 9, 2022
Revised: September 8, 2022
Accepted: January 17, 2023
Published: January 25, 2023

REFERENCES

1. Wu, C., Chen, X., Cai, Y., Xia, J., Zhou, X., Xu, S., Huang, H., Zhang, L., Zhou, X., Du, C., et al. (2020). Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in wuhan, China. JAMA Intern. Med. *180*, 934–943. https://doi.org/10.1001/jamainternmed.2020.0994.
2. Yang, X., Yu, Y., Xu, J., Shu, H., Xia, J.a., Liu, H., Wu, Y., Zhang, L., Yu, Z., Fang, M., et al. (2020). Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir. Med. *8*, 475–481. https://doi.org/10.1016/s2213-2600(20)30079-5.
3. Grasselli, G., Greco, M., Zanella, A., Albano, G., Antonelli, M., Bellani, G., Bonanomi, E., Cabrini, L., Carlesso, E., Castelli, G., et al. (2020). Risk factors associated with mortality among patients with COVID-19 in intensive care units in lombardy, Italy. JAMA Intern. Med. *180*, 1345–1355. https://doi.org/10.1001/jamainternmed.2020.3539.
4. Richardson, S., Hirsch, J.S., Narasimhan, M., Crawford, J.M., McGinn, T., Davidson, K.W., Consortium, a.t.N.C.-R., Barnaby, D.P., Becker, L.B., Chelico, J.D., et al. (2020). Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York city area. JAMA, J. Am. Med. Assoc. *323*, 2052–2058. https://doi.org/10.1001/jama.2020.6775.
5. Tzotzos, S.J., Fischer, B., Fischer, H., and Zeitlinger, M. (2020). Incidence of ARDS and outcomes in hospitalized patients with COVID-19: a global literature survey. Crit. Care *24*, 516. https://doi.org/10.1186/s13054-020-03240-7.

6. Network, C.-I.G.o.b.o.t.R., and the, C.-I.C.U.I. (2021). Clinical characteristics and day-90 outcomes of 4244 critically ill adults with COVID-19: a prospective cohort study. Intensive Care Med. *47*, 60–73. https://doi.org/10.1007/s00134-020-06294-x.

7. Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., Cheng, L., Li, J., Wang, X., Wang, F., et al. (2020). Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nat. Medicine *26*, 842–844. https://doi.org/10.1038/s41591-020-0901-9.

8. Wilk, A.J., Rustagi, A., Zhao, N.Q., Roque, J., Martinez-Colon, 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. Medicine *26*, 1070–1076. https://doi.org/10.1038/s41591-020-0944-y.

9. Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Pere, 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 (New York, N.Y.) *31*, eabc6027–6015. https://doi.org/10.1126/science.abc6027.

10. Mathew, D., Giles, J.R., Baxter, A.E., Oldridge, D.A., Greenplate, A.R., Wu, J.E., Alanio, C., Kuri-Cervantes, L., Pampena, M.B., D'Andrea, K., et al. (2020). Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. Science (New York, N.Y.) *369*. https://doi.org/10.1126/science.abc8511.

11. Moderbacher, C.R., Ramirez, S.I., Dan, J.M., Grifoni, A., Hastie, K.M., Weiskopf, D., Belanger, S., Abbott, R.K., Kim, C., Choi, J., et al. (2020). Antigen-specific adaptive immunity to SARS-CoV-2 in acute COVID-19 and associations with age and disease severity. Cell *183*, 996–1012.e1019. https://doi.org/10.1016/j.cell.2020.09.038.

12. Schulte-Schrepping, J., Reusch, N., Paclik, D., Baßler, K., Schlickeiser, S., Zhang, B., Krämer, B., Krammer, T., Brumhard, S., Bonaguro, L., et al. (2020). Severe COVID-19 is marked by a dysregulated myeloid cell compartment. Cell, 1–64. https://doi.org/10.1016/j.cell.2020.08.001.

13. Lucas, C., Wong, P., Klein, J., Castro, T.B.R., Silva, J., Sundaram, M., Ellingson, M.K., Mao, T., Oh, J.E., Israelow, B., et al. (2020). Longitudinal analyses reveal immunological misfiring in severe COVID-19. Nature *584*, 463–469. https://doi.org/10.1038/s41586-020-2588-y.

14. Combes, A.J., Courau, T., Kuhn, N.F., Hu, K.H., Ray, A., Chen, W.S., Chew, N.W., Cleary, S.J., Kushnoor, D., Reeder, G.C., et al. (2021). Global absence and targeting of protective immune states in severe COVID-19. Nature *1*, 10. https://doi.org/10.1038/s41586-021-03234-7.

15. Group, R.C., Horby, P., Lim, W.S., Emberson, J.R., Mafham, M., Bell, J.L., Linsell, L., Staplin, N., Brightling, C., Ustianowski, A., et al. (2021). Dexamethasone in hospitalized patients with covid-19. N. Engl. J. Med. *384*, 693–704. https://doi.org/10.1056/NEJMoa2021436.

16. Group, W.H.O.R.E.A.f.C.-T.W.; Shankar-Hari, M., Vale, C.L., Godolphin, P.J., Fisher, D., Higgins, J.P.T., Spiga, F., Savovic, J., Tierney, J., Baron, G., et al. (2021). Association between administration of IL-6 antagonists and mortality among patients hospitalized for COVID-19: a meta-analysis. JAMA. https://doi.org/10.1001/jama.2021.11330.

17. Kalil, A.C., Patterson, T.F., Mehta, A.K., Tomashek, K.M., Wolfe, C.R., Ghazaryan, V., Marconi, V.C., Ruiz-Palacios, G.M., Hsieh, L., Kline, S., et al. (2021). Baricitinib plus remdesivir for hospitalized adults with covid-19. N. Engl. J. Med. *384*, 795–807. https://doi.org/10.1056/NEJMoa2031994.

18. Marconi, V.C., Ramanan, A.V., de Bono, S., Kartman, C.E., Krishnan, V., Liao, R., Piruzeli, M.L.B., Goldman, J.D., Alatorre-Alexander, J., de Cassia Pellegrini, R., et al. (2021). Efficacy and safety of baricitinib for the treatment of hospitalised adults with COVID-19 (COV-BARRIER): a randomised, double-blind, parallel-group, placebo-controlled phase 3 trial. Lancet Respir. Med. https://doi.org/10.1016/s2213-2600(21)00331-3.

19. Prescott, H.C., Calfee, C.S., Thompson, B.T., Angus, D.C., and Liu, V.X. (2016). Toward smarter lumping and smarter splitting: rethinking strategies for sepsis and acute respiratory distress syndrome clinical trial design. Am. J. Respir. Crit. Care Med. *194*, 147–155. https://doi.org/10.1164/rccm.201512-2544CP.

20. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The human transcription factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.

21. Consortium, R.E., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330. https://doi.org/10.1038/nature14248.

22. Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187. https://doi.org/10.1038/nature09033.

23. Lam, M.T.Y., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M., et al. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature *498*, 511–515. https://doi.org/10.1038/nature12209.

24. Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C., and Glass, C.K. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol. Cell *51*, 310–325. https://doi.org/10.1016/j.molcel.2013.07.010.

25. Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of Nascent RNA Identifies a Unified Architecture of Initiation Regions at Mammalian Promoters and Enhancers (Nature Publishing Group), pp. 1–12. https://doi.org/10.1038/ng.3142.

26. Duttke, S.H., Chang, M.W., Heinz, S., and Benner, C. (2019). Identification and dynamic quantification of regulatory elements using total RNA. Genome Res. *29*, 1836–1846. https://doi.org/10.1101/gr.253492.119.

27. Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. *1*, 14. https://doi.org/10.1038/s41576-018-0089-8.

28. Murray, J.F., Matthay, M.A., Luce, J.M., and Flick, M.R. (1988). An expanded definition of the adult respiratory distress syndrome. The American Review of Respiratory Disease *138*, 720–723. https://doi.org/10.1164/ajrccm/138.3.720.

29. Calderon, D., Nguyen, M.L.T., Mezger, A., Kathiria, A., Müller, F., Nguyen, V., Lescano, N., Wu, B., Trombetta, J., Ribado, J.V., et al. (2019). Landscape of stimulation-responsive chromatin across diverse human immune cells. Nat. Genet. *51*, 1–17. https://doi.org/10.1038/s41588-019-0505-9.

30. Perez, C., Botta, C., Zabaleta, A., Puig, N., Cedena, M.-T., Goicoechea, I., Alameda, D., San José-Eneriz, E., Merino, J., Rodríguez-Otero, P., et al. (2020). Immunogenomic identification and characterization of granulocytic myeloid-derived suppressor cells in multiple myeloma. Blood *136*, 199–209. https://doi.org/10.1182/blood.2019004537.

31. Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Várnai, C., Thiecke, M.J., et al. (2016). Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell *167*, 1369–1384.e1319. https://doi.org/10.1016/j.cell.2016.09.037.

32. Vincent, J.L., Moreno, R., Takala, J., Willatts, S., De Mendonca, A., Bruining, H., Reinhart, C.K., Suter, P.M., and Thijs, L.G. (1996). The SOFA (Sepsis-related organ failure assessment) score to describe organ dysfunction/failure. On behalf of the working group on sepsis-related problems of the European society of intensive care medicine. Intensive Care Med. *22*, 707–710. https://doi.org/10.1007/BF01709751.

33. Knaus, W.A., Draper, E.A., Wagner, D.P., and Zimmerman, J.E. (1985). Apache II: a severity of disease classification system. Crit. Care Med. *13*, 818–829.

34. Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112. https://doi.org/10.1038/nature07829.

35. Liu, J., Liu, Y., Xiang, P., Pu, L., Xiong, H., Li, C., Zhang, M., Tan, J., Xu, Y., Song, R., et al. (2020). Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. J. Transl. Med. *18*, 206–212. https://doi.org/10.1186/s12967-020-02374-0.

36. Tonelli, C., Chio, I.I.C., and Tuveson, D.A. (2018). Transcriptional regulation by Nrf2. Antioxid. Redox Signal *29*, 1727–1745. https://doi.org/10.1089/ars.2017.7342.

37. Acosta-Herrera, M., Pino-Yanes, M., Blanco, J., Ballesteros, J.C., Ambrós, A., Corrales, A., Gandía, F., Subira, C., Domínguez, D., Baluja, A., et al. (2015). Common variants of NFE2L2 gene predisposes to acute respiratory distress syndrome in patients with severe sepsis. Crit. Care *19*, 256. https://doi.org/10.1186/s13054-015-0981-y.

38. D'Cruz, L.M., Knell, J., Fujimoto, J.K., and Goldrath, A.W. (2010). An essential role for the transcription factor HEB in thymocyte survival, Tcra rearrangement and the development of natural killer T cells. Nat. Immunol. *11*, 240–249. https://doi.org/10.1038/ni.1845.

39. Park, S.H., Kang, K., Giannopoulou, E., Qiao, Y., Kang, K., Kim, G., Park-Min, K.-H., and Ivashkiv, L.B. (2017). Type I interferons and the cytokine TNF cooperatively reprogram the macrophage epigenome to promote inflammatory activation. Nat. Immunol. *18*, 1104–1116. https://doi.org/10.1038/ni.3818.

40. Yukawa, M., Jagannathan, S., Vallabh, S., Kartashov, A.V., Chen, X., Weirauch, M.T., and Barski, A. (2020). AP-1 activity induced by co-stimulation is required for chromatin opening during T cell activation. J. Exp. Med. *217*. https://doi.org/10.1084/jem.20182009.

41. Park, A., and Iwasaki, A. (2020). Type I and type III interferons - induction, signaling, evasion, and application to combat COVID-19. Cell Host Microbe *27*, 870–878. https://doi.org/10.1016/j.chom.2020.05.008.

42. Shin, H.M., Minter, L.M., Cho, O.H., Gottipati, S., Fauq, A.H., Golde, T.E., Sonenshein, G.E., and Osborne, B.A. (2006). Notch1 augments NF-κB activity by facilitating its nuclear retention. EMBO J. *25*, 129–138. https://doi.org/10.1038/sj.emboj.7600902.

43. Wang, J., Shelly, L., Miele, L., Boykins, R., Norcross, M.A., and Guan, E. (2001). Human Notch-1 inhibits NF-kappa B activity in the nucleus through a direct interaction involving a novel domain. J. Immunol. *167*, 289–295. https://doi.org/10.4049/jimmunol.167.1.289.

44. Poulsen, L.I.C., Edelmann, R.J., Krüger, S., Diéguez-Hurtado, R., Shah, A., Stav-Noraas, T.E., Renzi, A., Szymanska, M., Wang, J., Ehling, M., et al. (2018). Inhibition of endothelial NOTCH1 signaling attenuates inflammation by reducing cytokine-mediated histone acetylation at inflammatory enhancers. Arterioscler. Thromb. Vasc. Biol. *38*, 854–869. https://doi.org/10.1161/ATVBAHA.117.310388.

45. Dent, A.L., Shaffer, A.L., Yu, X., Allman, D., and Staudt, L.M. (1997). Control of inflammation, cytokine expression, and germinal center formation by BCL-6. Science *276*, 589–592. https://doi.org/10.1126/science.276.5312.589.

46. Harris, M.B., Chang, C.C., Berton, M.T., Danial, N.N., Zhang, J., Kuehner, D., Ye, B.H., Kvatyuk, M., Pandolfi, P.P., Cattoretti, G., et al. (1999). Transcriptional repression of Stat6-dependent interleukin-4-induced genes by BCL-6: specific regulation of iepsilon transcription and immunoglobulin E switching. Mol. Cell Biol. *19*, 7264–7275. https://doi.org/10.1128/MCB.19.10.7264.

47. Liu, X., Lu, H., Chen, T., Nallaparaju, K.C., Yan, X., Tanaka, S., Ichiyama, K., Zhang, X., Zhang, L., Wen, X., et al. (2016). Genome-wide analysis identifies bcl6-controlled regulatory networks during T follicular helper cell differentiation. Cell Rep. *14*, 1735–1747. https://doi.org/10.1016/j.celrep.2016.01.038.

48. Ogawa, S., Lozach, J., Benner, C., Pascual, G., Tangirala, R.K., Westin, S., Hoffmann, A., Subramaniam, S., David, M., Rosenfeld, M.G., and Glass, C.K. (2005). Molecular determinants of crosstalk between nuclear receptors and toll-like receptors. Cell *122*, 707–721. https://doi.org/10.1016/j.cell.2005.06.029.

49. Perera, R.M., Stoykova, S., Nicolay, B.N., Ross, K.N., Fitamant, J., Boukhali, M., Lengrand, J., Deshpande, V., Selig, M.K., Ferrone, C.R., et al. (2015). Transcriptional control of autophagy-lysosome function drives pancreatic cancer metabolism. Nature *524*, 361–365. https://doi.org/10.1038/nature14587.

50. Eichenfield, D.Z., Troutman, T.D., Link, V.M., Lam, M.T., Cho, H., Gosselin, D., Spann, N.J., Lesch, H.P., Tao, J., Muto, J., et al. (2016). Tissue damage drives co-localization of NF-κB, Smad3, and Nrf2 to direct Rev-erb sensitive wound repair in mouse macrophages. Elife *5*, 554. https://doi.org/10.7554/eLife.13024.

51. Wendisch, D., Dietrich, O., Mari, T., von Stillfried, S., Ibarra, I.L., Mittermaier, M., Mache, C., Chua, R.L., Knoll, R., Timm, S., et al. (2021). SARS-CoV-2 infection triggers profibrotic macrophage responses and lung fibrosis. Cell. https://doi.org/10.1016/j.cell.2021.11.033.

52. Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell *171*, 1437–1452.e1417. https://doi.org/10.1016/j.cell.2017.10.049.

53. Initiative, C.-H.G. (2021). Mapping the human genetic architecture of COVID-19. Nature. https://doi.org/10.1038/s41586-021-03767-x.

54. Harley, J.B., Chen, X., Pujato, M., Miller, D., Maddox, A., Forney, C., Magnusen, A.F., Lynch, A., Chetal, K., Yukawa, M., et al. (2018). Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. Nat. Genet. *50*, 699–707. https://doi.org/10.1038/s41588-018-0102-3.

55. Overmyer, K.A., Shishkova, E., Miller, I.J., Balnis, J., Bernstein, M.N., Peters-Clarke, T.M., Meyer, J.G., Quan, Q., Muehlbauer, L.K., Trujillo, E.A., et al. (2021). Large-scale multi-omic analysis of COVID-19 severity. Cell Syst. *12*, 23–40.e27. https://doi.org/10.1016/j.cels.2020.10.003.

56. Wilk, A.J., Lee, M.J., Wei, B., Parks, B., Pi, R., Martinez-Colon, G.J., Ranganath, T., Zhao, N.Q., Taylor, S., Becker, W., et al. (2021). Multi-omic profiling reveals widespread dysregulation of innate immunity and hematopoiesis in COVID-19. J. Exp. Med. *218*. https://doi.org/10.1084/jem.20210582.

57. Wang, E.Y., Mao, T., Klein, J., Dai, Y., Huck, J.D., Jaycox, J.R., Liu, F., Zhou, T., Israelow, B., Wong, P., et al. (2021). Diverse functional autoantibodies in patients with COVID-19. Nature *1*, 30. https://doi.org/10.1038/s41586-021-03631-y.

58. Rubin, J.D., Stanley, J.T., Sigauke, R.F., Levandowski, C.B., Maas, Z.L., Westfall, J., Taatjes, D.J., and Dowell, R.D. (2021). Transcription factor enrichment analysis (TFEA) quantifies the activity of multiple transcription factors from a single experiment. Commun Biol. *4*, 661. https://doi.org/10.1038/s42003-021-02153-7.

59. Hung, I.F.-N., Lung, K.-C., Tso, E.Y.-K., Liu, R., Chung, T.W.-H., Chu, M.-Y., Ng, Y.-Y., Lo, J., Chan, J., Tam, A.R., et al. (2020). Triple combination of interferon beta-1b, lopinavir-ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial. Lancet (London, England) *395*, 1695–1704. https://doi.org/10.1016/S0140-6736(20)31042-4.

60. Salama, C., Han, J., Yau, L., Reiss, W.G., Kramer, B., Neidhart, J.D., Criner, G.J., Kaplan-Lewis, E., Baden, R., Pandit, L., et al. (2021). Tocilizumab in patients hospitalized with covid-19 pneumonia. N. Engl. J. Med. *384*, 20–30. https://doi.org/10.1056/NEJMoa2030340.

61. Thimmulappa, R.K., Lee, H., Rangasamy, T., Reddy, S.P., Yamamoto, M., Kensler, T.W., and Biswal, S. (2006). Nrf2 is a critical regulator of the innate immune response and survival during experimental sepsis. J. Clin. Invest. *116*, 984–995. https://doi.org/10.1172/JCI25790.

62. Gomez, J.C., Dang, H., Martin, J.R., and Doerschuk, C.M. (2016). Nrf2 modulates host defense during Streptococcus pneumoniae pneumonia in mice. J. of Immunol. *197*, 2864–2879. https://doi.org/10.4049/jimmunol.1600043.

63. Papaiahgari, S., Yerrapureddy, A., Reddy, S.R., Reddy, N.M., Dodd-O, J.M., Crow, M.T., Grigoryev, D.N., Barnes, K., Tuder, R.M., Yamamoto, M., et al. (2007). Genetic and pharmacologic evidence links oxidative stress to ventilator-induced lung injury in mice. Am. J. Respir. Crit. Care Med. *176*, 1222–1235. https://doi.org/10.1164/rccm.200701-060OC.

64. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

65. El-Houjeiri, L., Possik, E., Vijayaraghavan, T., Paquette, M., Martina, J.A., Kazan, J.M., Ma, E.H., Jones, R., Blanchette, P., Puertollano, R., and Pause, A. (2019). The transcription factors TFEB and TFE3 link the FLCN-AMPK signaling Axis to innate immune response and pathogen resistance. Cell Rep. *26*, 3613–3628.e3616. https://doi.org/10.1016/j.celrep.2019.02.102.

66. Pastore, N., Brady, O.A., Diab, H.I., Martina, J.A., Sun, L., Huynh, T., Lim, J.-A., Zare, H., Raben, N., Ballabio, A., and Puertollano, R. (2016). TFEB and TFE3 cooperate in the regulation of the innate immune response in activated macrophages. Autophagy *12*, 1240–1258. https://doi.org/10.1080/15548627.2016.1179405.

67. Calfee, C.S., Delucchi, K., Parsons, P.E., Thompson, B.T., Ware, L.B., and Matthay, M.A. (2014). Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. Lancet *2*, 611–620. https://doi.org/10.1016/S2213-2600(14)70097-9.

68. DPhil, E.E.D., MGen, K.L.B., DPhil, J.R., PhD, P.H., Cert, P.H.P., DPhil, T.C.M., PhD, A.R., MD, A.C.G., DPhil, C.G., DPhil, P.A.V.S.H., et al. (2016). Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. The Lancet Respiratory *4*, 259–271. https://doi.org/10.1016/S2213-2600(16)00046-1.

69. Burnham, K.L., Davenport, E.E., Radhakrishnan, J., Humburg, P., Gordon, A.C., Hutton, P., Svoren-Jabalera, E., Garrard, C., Hill, A.V.S., Hinds, C.J., and Knight, J.C. (2017). Shared and distinct aspects of the sepsis transcriptomic response to fecal peritonitis and pneumonia. Am. J. Respir. Crit. Care Med. *196*, 328–339. https://doi.org/10.1164/rccm.201608-1685OC.

70. Frishberg, A., Kooistra, E., Nuesch-Germano, M., Pecht, T., Milman, N., Reusch, N., Warnat-Herresthal, S., Bruse, N., Handler, K., Theis, H., et al. (2022). Mature neutrophils and a NF-kappaB-to-IFN transition determine the unifying disease recovery dynamics in COVID-19. Cell Rep. Med. *3*, 100652. https://doi.org/10.1016/j.xcrm.2022.100652.

71. Zhang, Q., Bastard, P., Liu, Z., Le Pen, J., Moncada-Velez, M., Chen, J., Ogishi, M., Sabli, I.K.D., Hodeib, S., Korol, C., et al. (2020). Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science *370*. https://doi.org/10.1126/science.abd4570.

72. Scicluna, B.P., van Vught, L.A., Zwinderman, A.H., Wiewel, M.A., Davenport, E.E., Burnham, K.L., Nurnberg, P., Schultz, M.J., Horn, J., Cremer, O.L., et al. (2017). Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. Lancet Respir. Med. *5*, 816–826. https://doi.org/10.1016/S2213-2600(17)30294-1.

73. Thompson, B.T., Chambers, R.C., and Liu, K.D. (2017). Acute respiratory distress syndrome. N. Engl. J. Med. *377*, 562–572. https://doi.org/10.1056/NEJMra1608077.

74. Vincent, J.L., Baron, J.-F., Reinhart, K., Gattinoni, L., Thijs, L., Webb, A., Meier-Hellmann, A., Nollet, G., Peres-Bota, D., and Investigators, A.B.C. (2002). Anemia and blood transfusion in critically ill patients. JAMA *288*, 1499–1507. https://doi.org/10.1001/jama.288.12.1499.

75. Tian, Y., Carpp, L.N., Miller, H.E.R., Zager, M., Newell, E.W., and Gottardo, R. (2022). Single-cell immunology of SARS-CoV-2 infection. Nat. Biotechnol. *40*, 30–41. https://doi.org/10.1038/s41587-021-01131-y.

76. You, M., Chen, L., Zhang, D., Zhao, P., Chen, Z., Qin, E.Q., Gao, Y., Davis, M.M., and Yang, P. (2021). Single-cell epigenomic landscape of peripheral immune cells reveals establishment of trained immunity in individuals convalescing from COVID-19. Nat. Cell Biol. *23*, 620–630. https://doi.org/10.1038/s41556-021-00690-1.

77. Kang, K., Bachu, M., Park, S.H., Kang, K., Bae, S., Park-Min, K.-H., and Ivashkiv, L.B. (2019). IFN-γ selectively suppresses a subset of TLR4-activated genes and enhancers to potentiate macrophage activation. Nat. Commun. *10*, 3320-3314. https://doi.org/10.1038/s41467-019-11147-3.

78. Wang, C., Nanni, L., Novakovic, B., Megchelenbrink, W., Kuznetsova, T., Stunnenberg, H.G., Ceri, S., and Logie, C. (2019). Extensive epigenomic integration of the glucocorticoid response in primary human monocytes and in vitro derived macrophages. Sci. Rep. *9*, 2772-2717. https://doi.org/10.1038/s41598-019-39395-9.

79. Qiao, Y., Giannopoulou, E.G., Chan, C.H., Park, S.H., Gong, S., Chen, J., Hu, X., Elemento, O., and Ivashkiv, L.B. (2013). Synergistic activation of inflammatory cytokine genes by interferon-γ-induced chromatin remodeling and toll-like receptor signaling. Immunity *39*, 454–469. https://doi.org/10.1016/j.immuni.2013.08.009.

80. 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. https://doi.org/10.1093/bioinformatics/bts635.

81. 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*, 31-21. https://doi.org/10.1186/s13059-014-0550-8.

82. Santos, N.P.D., Texari, L., and Benner, C. (2020). MEIRLOP: improving score-based motif enrichment by incorporating sequence bias covariates. BMC Bioinf. *21*, 1–22. https://doi.org/10.1186/s12859-020-03739-4.

83. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. *1*, 10. https://doi.org/10.1038/s41467-019-09234-6.

84. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233. https://doi.org/10.1038/s41598-019-41695-z.

85. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

86. Duttke, S.H., Beyhan, S., Singh, R., Neal, S., Viriyakosol, S., Fierer, J., Kirkland, T.N., Stajich, J.E., Benner, C., and Carlin, A.F. (2022). Decoding transcription regulatory mechanisms associated with coccidioides immitis phase transition using total RNA. mSystems, e0140421. https://doi.org/10.1128/msystems.01404-21.

87. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer International Publishing).

88. de Hoon, M.J.L., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics *20*, 1453–1454. https://doi.org/10.1093/bioinformatics/bth078.

89. Saldanha, A.J. (2004). Java Treeview–extensible visualization of microarray data. Bioinformatics *20*, 3246–3248. https://doi.org/10.1093/bioinformatics/bth349.

90. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K., and Kastenmuller, G. (2015). SNiPA: an interactive, genetic variant-centered annotation browser. Bioinformatics *31*, 1334–1336. https://doi.org/10.1093/bioinformatics/btu779.

91. Yoshida, H., Lareau, C.A., Ramirez, R.N., Rose, S.A., Maier, B., Wroblewska, A., Desland, F., Chudnovskiy, A., Mortha, A., Dominguez, C., et al. (2019). The cis-regulatory atlas of the mouse immune system. Cell *176*, 897–912.e820. https://doi.org/10.1016/j.cell.2018.12.036.

92. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

93. Duarte, R.R., and Copertino, D.C., Jr. (2020). Repurposing Fda-Approved Drugs for Covid-19 Using a Data-Driven Approach. https://doi.org/10.26434/chemrxiv.12148764.v1.

94. Le, B., Andreoletti, G., Oskotsky, T., Vallejo-Gracia, A., Ramirez, R.R., Yu, K., Kosti, I., Leon, K., Bunis, D., Li, C., et al. (2021). Transcriptomics-based Drug Repositioning Pipeline Identifies Therapeutic Candidates for COVID-19 (Research Square). https://doi.org/10.21203/rs.3.rs-333578/v1.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
RNA from peripheral leukocytes of COVID-19 ICU patients	This paper	UC San Diego, Department of Medicine
Plasma from peripheral blood of COVID-19 ICU patients	This paper	UC San Diego, Department of Medicine
Chemicals, peptides, and recombinant proteins		
Super-Script RT II	Thermo Fisher Scientific	Cat. No. 18064022
QuickCIP	New England Biolabs	Cat. No. M0525L
RNA 5′ Pyrophosphohydrolase (RppH)	New England Biolabs	M0356
SUPERase•In™ RNase Inhibitor (20 U/μL)	Thermo Fisher Scientific	Cat. No. AM2696
Terminator 5′-Phosphate-Dependent Exonuclease	Lucigen	Cat. No. TER51020
Trizol	Thermo Fischer Scientific	Cat. No. 15596018
Trizol LS	Thermo Fischer Scientific	Cat. No. 10296010
T4 RNA Ligase Reaction Buffer	New England Biolabs	Cat. No. B0216SVIAL
SpeedBead Magnetic Carboxylate Modified particles	Sigma Millipore	cat. No. 65152105050250
SYBR™ Gold Nucleic Acid Gel Stain	Thermo Fisher Scientific	Cat. No. S11494
GelGreen® Nucleic Acid Gel Stain	Biotium	Cat. No. 41005
commercial assays		
TruSeq® Stranded Total RNA Library Prep Gold	Illumina	Cat # 20020598
NEBNext®Multiplex Small RNA Library Prep Kit for Illumina®(Index Primers 1-48)	New England Biolabs	Cat. No. E7560S
ChIP-DNA clean and concentrator	Zymo Research	Cat. No. D5205
Human Anti-Virus Response Panel (13-plex)	BioLegend	Cat # 740349
Deposited data		
Raw data – csRNA-seq	NCBI Gene Expression Ominibus	GEO: GSE221065
Raw data – total RNA-seq	NCBI Gene Expression Ominibus	GEO: GSE221066
Leukocytes total RNA-seq	Overmyer et al. 2021[55]	GEO: GSE157103
Whole Blood single-cell RNAseq	Combes et al. 2021[14]	GEO: GSE163668
Blood cell type ATAC-seq	Calderon et al. 2019[29]	GEO: GSE118189
Neutrophil ATAC-seq	Perez et al. 2020[30]	GEO: GSE150018
Blood cells H3K27ac ChIP-seq	Javierre et al. 2016[31]	https://www.blueprint-epigenome.eu/
RELA ChIP-seq, monocytes-derived macrophages	Park et al. 2017[39]	GEO; GSE100381
NFKB1 ChIP-seq, CD4 T cells	Yukawa et al. 2020[40]	GEO: GSE116695
STAT3 ChIP-seq, monocytes-derived macrophages	Kang et al. 2019[77]	GEO: GSE120943
Glucocorticoid Receptor ChIP-seq, monocytes-derived macrophages	Wang et al. 2019[78]	GEO: GSE109438
IRF1 ChIP-seq, monocytes-derived macrophages	Qiao et al. 2013[79]	GEO: GSE43036
Software and algorithms		
STAR	Dobin et al., 2013[80]	http://code.google.com/p/rna-star/
DESeq2	Love et al. 2014[81]	http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HOMER	Heinz et al. 2010[64]	http://homer.ucsd.edu/homer/
MEIRLOP	Santos et al. 2020[82]	https://github.com/npdeloss/meirlop
Metascape	Zhou et al. 2019[83]	Metascape.org
Regulatory Element Locus Intersection (RELI)	Harley et al. 2018[54]	https://github.com/WeirauchLab/RELI
Uwot (0.1.11)	https://doi.org/10.48550/arXiv.1802.03426	https://github.com/jlmelville/uwot
Leiden (0.4.2)	Traag et al. 2019[84]	https://github.com/TomKellyGenetics/leiden
Igraph (1.3.4)	https://igraph.org/r/	https://github.com/igraph
FNN (1.1.3.1)	CRAN	https://cran.r-project.org/package%20=%20FNN
CMAP	Subramanian et al., 2017[52]	https://clue.io/cmap
LEGENDplex Data Analysis Software		BioLegend
SCANPY	Wolf et al. 2018[85]	https://github.com/theislab/Scanpy
R (4.1.2)	R Core Team	https://www.r-project.org
Prisms (version 9)	GraphPad	https://www.graphpad.com
Other		
Gelbreaker Tubes	IST Engineering Inc	Cat. No. 3388-100
Novex ™ TBE gel, 10%, 12 well	Thermo Fisher Scientific	Cat. No. EC62752BOX
Novex™ TBE-Urea Gels, 15%, 12 well	Invitrogen	Cat. No. EC68852BOX
UltraFree MC Centrifugal Filter, 0.45 μm pore size, hydrophilic PVDF, 0.5mL volume	Millipore Sigma	Cat. No. UFC30HVNB

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Michael T. Lam (MTLam@health.ucsd.edu).

Material availability

This study did not generate new unique reagents.

Data and code availability

- De-identified patients' csRNA-seq (GEO: GSE221065) and total RNA-seq data (GEO: GSE221066) have been deposited at NCBI Gene Omnibus and are publicly available as of the publication date. Accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patient cohorts

Cohort 1: Patients diagnosed with COVID-19 admitted to the University of California, San Diego (UCSD) affiliated hospitals, including Hillcrest Medical Center, Jacobs Medical Center, and Rady Children's Hospital, were prospectively recruited from April to June 2020. Patients, or a designated surrogate, provided informed consent to participate in this study. Blood was drawn on hospitalization days 1, 3, 5, 7, 9, 11, and discharge/death for analysis. Medical records were reviewed, and patient demographics, laboratory values, clinical characteristics, and outcomes were extracted using the Research Electronic Data Capture (REDCap) electronic data capture tool hosted at UCSD. The age, gender, and clinical characteristics of participants are detailed in Table S1. Longitudinal peripheral leukocyte cistromic activity was profiled from 17 patients, ranging from day 1–25 of admission, with a median of seven-time points per patient (n = 92 disease samples). Tables S3–S5 detail the hospital time course and the clinical severity at each time point of cistromic assessment, evaluated by Sequential Organ Failure Assessment (SOFA), Acute Physiology and Chronic Evaluation

(APACHE II), and Modified Murray Score. For control, five non-hospitalized, healthy, asymptomatic individuals provided peripheral blood for cistromic, transcriptomic, and cytokine analysis. The study was approved by the Institutional Review Board at the University of California, San Diego (UCSD IRB#190699).

Cohort 2: public data of 100 COVID-19 patients admitted to the Albany Medical Center between April 6th to May 1st in 2020.[55] Blood was drawn at enrollment and processed for leukocytes-enriched total RNA-seq. Clinical severity was characterized by "hospital-free day at day 45" (HFD-45). Further information about this cohort can be found in.[55]

Cohort 3: public data of 21 COVID-19 patients admitted to the University of California San Francisco between April 8th and May 1st of 2020.[14] Whole blood was processed for single-cell RNA-sequencing from 21 COVID-19 patients and 14 healthy controls. Further information about this cohort, including the definition of disease severity, can be found in.[14]

METHOD DETAILS

Characterizing lung injury with modified Murray Score

The Murray Score was developed to characterize the level of lung injury in acute respiratory distress syndrome.[28] This system assigned a score of 0-4 to the following 4 categories. 1) the extent of lung involvement on chest radiograph; 2) level of hypoxemia using PaO_2 to FiO_2 ratio; 3) the range of positive end expiratory pressure (PEEP); and 4) range of lung compliances based on tidal volume, peak inspiratory pressure, and PEEP. For this study, in order to stratify lung injury of patients prior to mechanical ventilation, after liberation from mechanical ventilation as well as requirement for advance therapy on mechanical ventilation, we added the mode of respiratory support. Patients on room air will be given 0 point; 1 point for 1-6 L (L) of supplemental oxygen through nasal cannula; 2 points for non-rebreather mask at 10-15L of supplemental oxygen; 3 points for mechanical ventilation; and 4 points for mechanical ventilation with proning and paralysis. The modified Murray Score was tabulated by averaging the score from these five categories. A score of 0.1–2.5 was considered mild-moderate disease. Severe ARDS is >2.5. Non-invasive positive pressure ventilations including bilevel and heated high flow nasal cannula were not included in the modified Murray score because at the time of recruitment, the safety of these modalities for exposing medical staff was not well understood, and their use was generally discouraged.

Multiorgan dysfunction assessment with SOFA and Apache-II scores

The Sequential Organ Failure Assessment (SOFA)[32] and the Acute Physiology and Chronic Health Evaluation (APACHE-II)[33] are clinical scores that predict ICU mortality based on multiorgan function. For longitudinal assessments, these clinical scores were tabulated based on the vital signs, laboratory values, level of respiratory support, and vasopressor dose within the 24-h window of sample collection for csRNA-seq, total RNA-seq, and cytokine profiling.

Blood sample processing

Blood (3-10mL) was collected in a Sodium Heparin (BD Vacutainer, ref. 366,480) or Potassium EDTA (BD Vacutainer, ref. 367861) tubes. To prevent coagulation, the tubes were inverted 10 times prior to transport at room temperature. Blood was processed within 4 h of collection and kept at room temperature throughout the protocol. Whole blood from EDTA tubes and Heparin tubes was processed for isolation of plasma and whole white blood cells (WBC).

Peripheral leukocytes isolation

To isolate WBCs, whole blood in EDTA tubes was centrifuged at 300xg for 20 min. Plasma was first removed, leaving a cell pellet containing WBCs and red blood cells (RBCs). RBCs were lysed by resuspending the cellular pellet in 1X RBC Lysis Buffer (ammonium chloride (8.02 g/L), sodium bicarbonate (0.84 g/L), and EDTA (0.37 g/L) in deionized water) and incubated for 10 min. The cell suspension was then centrifuged at 600xg for 5 min and the pellet was again resuspended in RBC lysis buffer for 5 min. The reaction was quenched with 3 times the volume of 1X HBSS (Gibco, ref. 14175-095). After a sample was collected for a cell count, the isolated WBCs were pelleted at 600xg for 5 min. The WBC pellet was lysed in Trizol Reagent (Life Technologies, ref. 15596018) with a target concentration of 5-10 million cells/mL and stored at −80C prior to RNA extraction.

Plasma isolation

Plasma for cytokine analysis was collected from Sodium Heparin tube. Plasma was removed from blood separated by Polymorphprep per manufacturer's instructions (Progen). Plasma was transferred to new microcentrifuge tubes and centrifuged at 3731xg for 5 min at room temperature to remove any cellular debris. Supernatant was transferred to new tubes and flash frozen in dry ice and 95% ethanol. Plasma was stored at −80C for further analysis.

RNA extraction and purification

Total RNA was extracted from WBCs using TRIzol(tm) reagent (Cat. No. 15596018, ThermoFisher Scientific) as per manufacturer's instructions. Half of the total RNA was submitted for capped-small RNA-seq library generation. The remaining RNA was treated with TURBO(tm) DNase (AM1907, ThermoFisher Scientific) as per manufacturer's instructions and used for bulk total RNA-sequencing.

Capped small RNA-sequencing

csRNA-seq was performed as described in previously[26] using protocol version 5.2.[86] Briefly, small RNAs of 20-65 nt were size selected from 0.3-1.0 μg of total RNA by denaturing gel electrophoresis using 15% TBU gels (Invitrogen). A 10% input sample was taken aside, and the remainder enriched for 5′-capped RNAs with 3′-OH representing RNAPII initiated RNAs. Monophosphorylated RNAs were selectively degraded by Terminator 5′-phosphate-dependent exonuclease (Lucigen). Subsequent 5′ dephosphorylation by CIP (NEB) followed by decapping with RppH (NEB) augments Cap-specific 5′ adapter ligation by T4 RNA ligase 1 (NEB). The 3′ adapter was ligated using truncated T4 RNA ligase 2 (NEB) without prior 3′ repair to select against degraded RNA fragments. Following cDNA synthesis, libraries were amplified for 11-14 cycles and sequenced SE75 on the Illumina NextSeq 500 sequencer.

Sequencing reads were trimmed for 3′ adapter sequences using HOMER ("homerTools trim −3 AGATCGGAAGAGCACACGTCT -mis 2 -minMatchLength 4 -min 20") and aligned to the human GRCh38/hg38 genome using STAR[80] with default parameters. Only reads with a single, unique alignment (MAPQ ≥ 10) were considered in the downstream analysis. Furthermore, reads with spliced or soft clipped alignments were discarded (the latter often removes erroneous snRNA alignments). Transcription Start Regions (TSRs), representing loci with significant transcription initiation activity (i.e. 'peaks' in csRNA-seq), were defined using HOMER's findcsRNATSS.pl tool,[26] which uses short input RNA-seq, traditional RNA-seq, and annotated gene locations to eliminate loci with csRNA-seq signal arising from non-initiating, high abundance RNAs that nonetheless are captured and sequenced by the method (full description is available in Duttke et al.[26]). To lessen the impact of outlier samples across the data collected for this study, csRNA-seq samples were first pooled into a single META-experiment, and TSRs where then identified using findcsRNATSS.pl with a minimal TSR detection threshold of 1 read per 10 million mapped reads ("-ntagThreshold 1"), yielding 93,465 TSRs total. The resulting TSRs were then quantified in all samples by counting the 5′ ends of reads aligned at each TSR on the correct strand. The raw read count table was then normalized using DESeq2's rlog variance stabilization method.[81] The resulting normalized data was used for all downstream analysis. Normalized genome browser visualization tracks were generated using HOMER's makeMultiWigHub.pl tool. TSR genomic DNA extraction, nucleotide frequency analysis relative to the primary TSS, general annotation, other basic analysis tasks were performed using HOMER's annotatePeaks.pl function. Overlaps between TSRs and other genomic features (including peaks from published ATAC-seq studies, and annotation to the 5′ promoter of annotate GENCODE(v34) transcripts), was performed using HOMER's mergePeaks tool.

Total RNA sequencing

Libraries were prepared using Illumina's TruSeq Stranded Total RNA Library Prep Gold according to manufacturer's instructions. In brief, rRNA was depleted from total RNA (0.35 mg) by using subtractive hybridization. The RNA was then fragmented by metal-ion hydrolysis and subsequently converted to cDNA using Super-Script II. The cDNA was then end-repaired, adenylated, and ligated with Illumina sequencing adapters. Finally, the libraries were enriched by PCR amplification. All sequencing libraries were then quantified, pooled, and sequenced paired-end 150 base-pair (bp) on Illumina Novaseq at the Salk Next Generation Sequencing Core. Each library was sequenced on average 30 million reads. Sequencing reads were aligned to the human GRCh38/hg38 genome using STAR. STAR was also used to quantify read counts per gene using transcripts defined by GENCODE (version 34). RNA-seq read counts were then normalized using DESeq2's rlog variance stabilization method[81] for all downstream analyses.

For total RNA-seq from Overmyer et al.,[55] sequencing reads were downloaded from GEO: GSE157103 and were processed in the same fashion (i.e. mapped with STAR, rlog normalized with DESeq2).

Cytokine measurements

The Human Anti-Virus Response Panel (13-plex; BioLegend, San Diego, CA) was used to quantitate human plasma cytokines (IL-1β, IL-6, IL-8, IL-10, TNFα, and IP-10). Plasma samples were stored at ˉ80°C until use. For the cytokine assay, plasma was freshly thawed at room temperature, centrifuged at 1,000 x g for 5 min, and run at a 2-fold dilution in Assay Buffer per the manufacturer's instructions. Samples were acquired on a Canto II flow cytometer (BD) using a high throughput sampler. Samples were run in duplicate unless plasma volume was inadequate, and standards were run on all plates. Cell signaling technology (CST) was run prior to all flow cytometry runs to ensure low detector CVs and set laser delay. LEGENDplex Data Analysis Software (BioLegend) was used for analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Principal component analysis

We performed Principal Component Analysis (PCA) on the rlog normalized csRNA-seq libraries (n = 97) using prcomp() function in R 4.1.2.

Unsupervised machine learning for TSR cluster identification

We used hierarchical clustering and Uni-form Mani-fold Approximation and Projection (UMAP) for unbiased clustering of TSRs based on csRNA-seq activity at the level of each TSR. First, patterns of csRNA-seq regulation were identified with unbiased hierarchical clustering using HOMER ("homerTools cluster"). The rlog normalized read counts across all TSRs were first row centered by the average read count for each TSR, and the data was subsequently hierarchically clustered using average linkage and the Pearson correlation coefficient between TSR profiles as the distance metric. Due to the size of the dataset, 10,000 random TSRs were first

selected for hierarchical clustering. After completion, the remaining TSRs were assigned to their location in the hierarchical tree based on their nearest neighbor. The final clusters were defined as the maximum-sized subclusters with an average correlation coefficient no greater than 0.30 and a minimum of 500 TSRs (to exclude small, highly variable clusters), yielding a total of 26. To visualize the data, we performed UMAP independently using R packages uwot, leiden,[84] igraph, and FNN, with the following setting: n_component = 2, n_neighbors = 16, a = 2.5, b = 0.575, and metric = 'correlation'. We used ggplot2[87] to visualize the UMAP projection and to overlay information, including cluster IDs, relative csRNA-seq activity, chromatin accessibility, and correlation coefficients for Modified Murray Score, SOFA, APACHE-II, and Hospital Time using the 'color' anesthetic mapping parameter of ggplot2.

Motif enrichment analysis

To identify TF motifs in TSRs that are associated with clinical scores and other quantitative phenotypes, we applied MEIRLOP, a tool that uses logistic regression to model the presence of motifs in a set of scored DNA sequences while accounting for simple nucleotide composition bias, such as that introduced by CpG Islands.[82] TSRs were first scored by how well their activity profile across samples correlated (Pearson) with the Modified Murray Score, such that TSRs with relatively high activity in samples from very sick patients (i.e. high Modified Murray Score) yield high correlation coefficient values, while TSRs with relatively high activity in healthy patients yield low values. Sequences from −150 to +50 bp relative to the primary (mode) TSS within each TSR were then extracted, and these sequences and their associated correlation scores were then analyzed using MEIRLOP. 438 TF motifs in HOMER's known motif database were evaluated and the motifs yielding the most extreme enrichment coefficients with significant p values were reported.

To identify motifs associated with discrete TSR clusters, we used HOMER[64] to scan for DNA motifs in each TSR (−150,+50) using HOMER's known motif database, assigning the presence of a motif to each TSR if the motif was detected at least once. Motif enrichment for each cluster was calculated by comparing motif occupancy rates in each cluster versus all other clusters to calculate the log2 enrichment and significance using the Fisher Exact test. The top enriched motif(s) were then used to label the clusters, accounting for highly similar motifs from large families.

For motif instances and positional analysis, we analyzed the frequency of motif instances from HOMER's known motif database in 10 bp increments relative to the primary TSS within a −500 to 500 bp window using the annotatePeak.pl script from HOMER (options "-size 1000 -hist 10"). For background reference, we tabulate the motif instances using the entire active cistrome identified by csRNA-seq (n = 93,465).

Integration of csRNA-seq data with hematopoietic NGS epigenome profiling

Previously published bulk epigenomics data (ATAC-seq/ChIP-seq) from isolated hematopoietic cell types was used to assess the potential cell-type specificity of TSRs identified in our whole leukocyte csRNA-seq profiling experiments. ATAC-seq data from GEO: GSE118189 [29] 26 different peripheral blood cell types was supplemented with ATAC-seq from primary neutrophils from GEO: GSE150018 [30] to analyze open chromatin. H3K27ac ChIP-seq data for 21 different peripheral blood cell types were downloaded from the Blueprint Epigenome project (https://www.blueprint-epigenome.eu/)[31] to assess regions with active chromatin modifications. TF ChIP-seq data for RELA(NF-κB) in monocyte derived macrophages (MDM) (GEO: GSE100381 [39]), NFKB1 in $CD4^+$ T cells (GEO: GSE116695 [41]), STAT3 in MDM (GEO: GSE120943 [77]), GR in MDM (GEO: GSE109438 [78]), and IRF1 in MDM (GEO: GSE43036 [79]) were used to confirm the binding of TFs to predicted sites based on DNA motif analysis of TSR sequence. For ATAC-seq and TF ChIP-seq experiments, sequencing reads were downloaded from NCBI SRA, trimmed for adapter sequences, and aligned to the hg38 genome using STAR with default parameters. Replicate experiments were pooled by concatenating alignment files. Uniquely aligned reads (MAPQ>10) were then analyzed using HOMER to find peaks using "-style atac" and "-style factor" for ATAC-seq and TF ChIP-seq experiments, respectively. TF ChIP-seq peaks were found using their respective negative control input sequencing experiments, while ATAC-seq peaks were found using the pooled input from all ChIP-seq experiments as a control. HOMER was used to create genome browser tracks, quantify reads in the vicinity of TSRs to quantify enrichment, and create histograms of read distributions relative to the primary TSS (e.g. Figure S1C). H3K27ac ChIP-seq data from the Blueprint Epigenome project was downloaded as bigWig files, converted to bedGraph files using UCSC's bigWigToBedGraph utility, and quantified at TSRs using HOMER's annotatePeaks.pl program using the '-bedGraph' option.

To score TSRs by their cell type-specific ATAC-seq enrichment (i.e. Figure 2C), ATAC-seq reads (normalized to 10^7 total mapped reads) for each cell type were quantified in the vicinity of all TSR (+/−200bp from the primary TSS or each TSR). For each TSR, the enrichment for each cell type was defined as the log2 ratio of reads from that cell type divided by the average normalized read count for all cell types. The same approach was used to score ChIP-seq specific enrichment (i.e. Figures S3 and S5C) by quantifying each ChIP-seq experiment across TSRs (+/−200 bp for TF ChIP-seq, +/− 500 bp for H3K27ac ChIP-seq). Aggregate cluster cell-type enrichments were reported by calculating the average TSR cell-type specific enrichment for each of the TSRs in the cluster/TF-network (i.e. Figures S3 and S5C). Cell type enrichment patterns were further hierarchically clustered using Cluster 3.0 [88] (Pearson Correlation, average linkage) and visualized using Java TreeView.[89]

Overlap of disease-associated SNPs from COVID-19 GWAS with TSRs

GWAS meta-analysis results from the COVID-19 Human Genetics Initiative[53] corresponding the A2, B1, B2, and C2 comparisons (Version 6, hg38 version) were downloaded from the consortium website (https://www.covid19hg.org/results/r6/). Significant disease-associated SNPs were defined using a p value threshold of 5×10^{-8} as recommended in the original study.[53] To visualize

disease-associated SNPs overlapping TSRs in the TSR UMAP, we identified the SNP with the most significant p value that overlapped each TSR within −300 to +100 relative to the primary TSS.

Regulatory Element Locus Intersection (RELI)[54] was used to assess the significance of overlap between lists of TSRs (e.g. TF clusters, TSRs with modified Murray Scores >0.15, etc.) and significant GWAS SNPs. First, lists TSRs were first mapped from the hg38 to the hg19 version of the human genome using UCSC's liftOver tool. Next, lists of significant disease-associated SNPs were expanded to include additional SNPs in LD ($r^2 > 0.8$) using SNIPA (EU ancestry).[90] These data were then used to run RELI using default parameters, looking for SNPs that overlapped TSRs from −300 to +100 relative to the primary TSS, reporting the corrected p value and overlapping SNPs.

Target genes selection and pathway analysis

Traditionally target genes are identified as the nearby genes to regulatory elements.[91] Because csRNA-seq simultaneously profiles transcriptional initiation of protein-coding genes and *cis*-regulatory elements, the initiation activities of the target genes should correlate with the *cis*-regulatory elements. We defined target genes if they have promoter TSR that are in the same TF-network cluster. These target genes were submitted for Gene Ontology and Pathway Analysis using Metascape.[83]

Calculating TF-network activity from csRNA-seq

The aggregate TSR csRNA-seq signal represents the activity of the TF-network program. For each TSR cluster, the median csRNA-seq normalized count (log2) for all TSRs in a given cluster was calculated to represent TF-network activity.

Target gene selection to calculate TF-network activity from transcriptomes

To compute TF-network activity from transcriptomic data, we first need to identify target genes whose steady state RNA levels reflect regulation at the transcription level. The target genes RNA level from total RNA-seq should match the transcriptional initiation activity from csRNA-seq. We used matching samples to process both total RNA-seq and csRNA-seq (n = 55) to identify target genes within the E2F/MYB, STAT/BCL6, and T1ISRE/STAT network. We computed a Pearson correlation coefficient for each target gene's RNA level to the activity pattern of the cistromic TSR cluster across the 55 samples. Target gene was then selected by the following criteria: 1) the gene has a TSR resides in the annotated promoter and clusters into the E2F/MYB, STAT/BCL6, or T1ISRE/STAT network; 2) the gene is unique to one TF-network; 3) the gene is expressed in the transcriptomic analysis, defined by having a median normalized read count greater than 4 FKPM across all samples; 4) the target genes RNA level is positively correlated with the activity of the TSR cluster (correlation coefficient >0.2). The last criteria remove genes whose steady-state RNA level can be influenced by post-transcriptional mechanisms. We identified 106, 176, and 58 target genes for E2F/MYB, STAT/BCL6, and T1ISRE/STAT respectively. For validation, the TF-network activity determined by the median of TSRs (csRNA-seq) and target genes (total RNA-seq) for the three TF-networks were analyzed in a 3 × 3 correlation matrix with Pearson's correlation and two-tail test for statistical significance (Prism 9).

Integrated analysis with single-cell RNA-seq data

Single-cell RNA-seq data from Combes et al.[14] was analyzed using the SCANPY toolkit.[85] Cells with greater than 20% mitochondrial or 50% rRNA reads were excluded, as were cells with fewer than 100 genes detected. In some cases, previous annotation labeling cells as possible multiplets was available and used to filter out non-singlets. Gene set signatures were tabulated per cell from raw read counts. For cell type identification, read counts per cell were normalized and log-transformed before applying the regress_out function to the total counts. The counts were then scaled to unit variance and zero mean. The cells were run through PCA, neighbor graph generation, and UMAP with default parameters. Cell clusters were identified using Leiden clustering with a resolution parameter of 0.25. Marker genes from these clusters were used to identify five neutrophil clusters.

Drug repurposing & connectivity mapping

CMap (https://clue.io/cmap) provides expression similarity scores for a specific expression profile with other drug-induced transcriptional profiles, including consensus transcriptional signatures of 2,837 drugs grouped into 83 drug classes.[52] The connectivity score from CMap is calculated based on the observed enrichment scores in the queried gene lists relative to transcriptional signatures in the L1000 reference database. The score incorporates a nominal p value calculated based on the comparison between the query and reference signatures relative to a null distribution of random queries, using the Kolmogorov-Smirnov enrichment statistic, which is then corrected for multiple testing using the false discovery rate method.[52,92–94]

For drug repurposing, the connectivity map scores were computed based on the target genes for each TF-network cluster. We hypothesized that the gene expression pattern resulting from the perturbation by a therapeutic compound should negatively correlate with the COVID-19 transcriptional signature as previously shown.[93,94] Therefore, we selected those compounds that had significant negative connectivity map scores (i.e. compounds with the connectivity scores < −90, predicted to reverse our input signature[52]). For each cluster, we grouped predicted drugs into: (1) individual drug lists with connectivity scores (cs) <-90; (2) Drugs for each cluster based on the predicted cluster targets with cs < -90; and (3) Pharmacological drug classes (i.e. JAK inhibitors) to determine which broader classes of drugs may be predicted by each cluster to reverse the COVID-19 transcriptional signature (cs < - 90).