iScience

Article

Epigenetic and transcriptional responses in circulating leukocytes are associated with future decompensation during SARS-CoV-2 infection

Micah T. McClain,
Ilya Zhbannikov,
Lisa L. Satterwhite,
..., Elizabeth A.
Petzold, Xiling
Shen, Christopher
W. Woods

micah.mcclain@duke.edu

Highlights

Detectable immune
response patterns precede
clinical events in COVID-19

Clinical decompensation
involves conserved
immune response
elements

Combined genomic
approaches offer insights
into the biology of COVID-
19

Article

Epigenetic and transcriptional responses in circulating leukocytes are associated with future decompensation during SARS-CoV-2 infection

Micah T. McClain,[1,2,9,*] Ilya Zhbannikov,[3] Lisa L. Satterwhite,[4] Ricardo Henao,[5] Nicholas S. Giroux,[6] Shengli Ding,[6] Thomas W. Burke,[1] Ephraim L. Tsalik,[7] Christina Nix,[1] Jorge Prado Balcazar,[6] Elizabeth A. Petzold,[1] Xiling Shen,[8] and Christopher W. Woods[1,2]

SUMMARY

To elucidate host response elements that define impending decompensation during SARS-CoV-2 infection, we enrolled subjects hospitalized with COVID-19 who were matched for disease severity and comorbidities at the time of admission. We performed combined single-cell RNA sequencing (scRNA-seq) and single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) on peripheral blood mononuclear cells (PBMCs) at admission and compared subjects who improved from their moderate disease with those who later clinically decompensated and required invasive mechanical ventilation or died. Chromatin accessibility and transcriptomic immune profiles were markedly altered between the two groups, with strong signals in CD4$^+$ T cells, inflammatory T cells, dendritic cells, and NK cells. Multiomic signature scores at admission were tightly associated with future clinical deterioration (auROC 1.0). Epigenetic and transcriptional changes in PBMCs reveal early, broad immune dysregulation before typical clinical signs of decompensation are apparent and thus may act as biomarkers to predict future severity in COVID-19.

INTRODUCTION

Our understanding of immune mechanisms driving the varied acute, recovery, and post-infectious manifestations of COVID-19 continues to evolve.[1] A growing body of evidence exists that delineates innate and adaptive immune responses to SARS-CoV-2 infection and in particular defines the divergence between immune responses occurring at the time of mild/moderate illness and those occurring at the time of severe/critical disease.[2–7] Coronaviruses have developed intricate, coordinated processes that regulate the host epigenome and control host innate immune antiviral defenses in a manner that supports robust viral replication and pathogenesis; however, these processes mostly occur in primarily infected cells.[8] Indeed, extensive work has described the delicate interplay between severity-associated genetic risk factors and immune activation in the lungs during COVID-19.[9,10] Additional work has focused on how host immune responses across the spectrum of disease manifest through epigenetic and transcriptional networks in circulating or resident leukocytes of various types.[11–15] Single-cell RNA sequencing studies of peripheral blood mononuclear cells from patients infected with SARS-CoV-2 suggest a distinct peripheral immune transcriptional signature of severe COVID-19 consisting of alteration of interferon (IFN)-stimulated genes in numerous cell types, as well as antigen presentation and proinflammatory pathways.[2,16,17] Populations of memory-like natural killer (NK) T cells have been described that exhibit elevated type I interferon signaling, are enriched in elderly COVID-19 patients, and are positively correlated with COVID-19 disease severity.[10,18] Both CD14$^+$ and CD16$^+$ monocytes and dendritic cells have been shown to exhibit activation of gene expression modules that suggest tolerogenic properties in severe and critical COVID-19.[15,19] These studies and others have been critical to build foundational understanding of the pathogenesis underlying the varying clinical manifestations of this novel disease.[17,19,20]

A limitation of much of the prior work is a focus on sampling at the time of the illness of interest, such as sampling patients while already in the ICU and comparing them with outpatients or those requiring only low-level supplemental oxygen. This is highly effective when the goal is to compare disease pathogenesis underlying each of these active states.[5,13,15,18] However, analysis of immunologic changes at earlier

[1]Division of Infectious Diseases, Duke University Medical Center, Durham, NC 27710, USA
[2]Durham Veterans Affairs Medical Center, Durham, NC 27705, USA
[3]Department of Medicine, Clinical Research Unit, Duke University Medical Center, Durham, NC 27710, USA
[4]Department of Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham, NC 27708, USA
[5]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC 27710, USA
[6]Department of Biomedical Engineering, Pratt School of Engineering, Duke University, Durham, NC 27708, USA
[7]Danaher Diagnostics, Washington, DC 20037, USA
[8]Terasaki Institute for Biological Innovation, Los Angeles, CA 90024, USA
[9]Lead contact
*Correspondence: micah.mcclain@duke.edu
https://doi.org/10.1016/j.isci.2023.108288

Table 1. Cohort demographic characteristics

	Controls (N = 7)	Worsening (N = 14)
Age (Years)	68 (51–82)	65 (53–90)
Sex (%)		
Female	5 (71)	5 (36)
Male	2 (29)	9 (64)
Race (%)		
White	5 (71)	8 (57)
Black/African American	2 (29)	5 (36)
Mixed Race/Other	0 (0)	1 (7)
Ethnicity (%)		
Non-Hispanic/Non-Latino	6 (86)	14 (100)
Hispanic/Latino	1 (14)	0 (0)
Comorbidities		
BMI (Range)	33.6 (21–41)	35.9 (20–48)
Tobacco Use (%)	3 (43)	4 (29)
CKD (%)	2 (29)	5 (36)
COPD (%)	1 (14)	3 (21)
Diabetes (%)	5 (71)	5 (36)
HIV (%)	0 (0)	0 (0)
CVA (%)	2 (29)	4 (29)
CAD (%)	3 (43)	5 (36)
CHF (%)	2 (29)	4 (29)
Hypertension (%)	4 (57)	9 (64)
Immunosuppression (%)	1 (14)	3 (21)
Cancer (%)	1 (14)	3 (21)
COVID-19 Severity		
WHO Ordinal Score at Admission (Mean [Range])	4.4 (4–5)	4.4 (4–5)
Max WHO Score Achieved	4.4 (4–5)	7.4 (6–8)
Days Since Symptom Onset	6.6 (4–10)	6.2 (1–9)

times—before severe disease actually manifests—offers the potential to elucidate critical early steps in the development of this dysregulated immunity. Not only does this approach allow for identification of potential checkpoints where early immune intervention might have increased value, but it also allows us to mine these foundational changes for early prognostic signals in circulating cells that predict future clinical decompensation.

To identify transcriptomic and epigenetic biomarkers capable of predicting the severity of future illness, we prospectively enrolled a matched cohort of 21 patients with similar comorbidities and disease severity at the time of hospitalization and initial sampling. We performed combined single-cell RNA sequencing (scRNA-seq) and single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) on peripheral blood mononuclear cells (PBMCs) at the time of admission and identified signals that identified those who would go on to clinically decompensate later in their hospitalization.

RESULTS

Clinical cohort

Twenty-one subjects matched for disease severity and duration of illness were enrolled at the time of initial hospital admission with moderate COVID-19 (WHO ordinal score 4 or 5, Table 1). Fourteen of these patients would experience later clinical deterioration after enrollment and progress to requiring invasive mechanical ventilation (n = 14), six of whom died (mean increase in WHO ordinal score of 3.0). The remaining seven patients remained stable and never developed increased oxygen requirements or required transfer to the ICU (WHO ordinal score never increased from admission value of 4 or 5). Samples for the study were drawn at the initial time of presentation for hospital admission for all subjects.
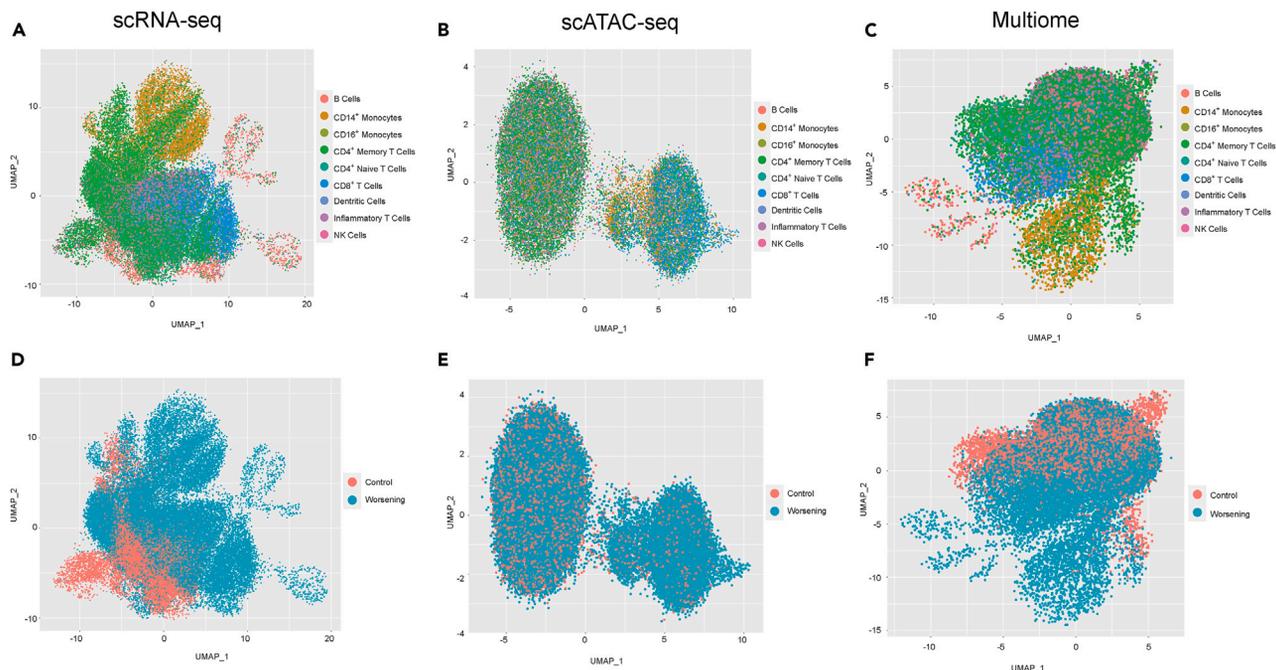
Figure 1. (Top row) Uniform manifold approximation and projection (UMAP) plots show relative distribution of immune cell types for the two cohorts, worsening and stable, based on (left to right)

(A) scRNA-seq, (B) scATAC-seq, and (C) multiome composite data. Immune cell types were color coded and include B cells, CD14+ monocytes, CD16+ monocytes, CD4+ memory T cells, CD4+ naive T cells, CD8+ T cells, dendritic cells, inflammatory T cells, and natural killer (NK) cells. (Second row) UMAP plots of relative distribution of combined immune cell types for the two cohorts, worsening and stable, based on experimental group (left to right).

(D) scRNA-seq data, (E) scATAC-seq data, and (F) multiome composite data. MicroRNAs, non-coding RNAs, mitochondrial genes, ribosomal RNAs, and low-expressed genes were excluded.

A multiomic approach using combined scRNA-seq and scATAC-seq

To examine early transcriptional and epigenetic events that occur at the level of circulating leukocytes and define diverging clinical states early in the disease course, we utilized a composite assay of scATAC-seq and scRNAseq. The 10X Genomics Chromium Single Cell Multiome ATAC + Gene Expression platform was utilized to analyze chromatin accessibility and transcriptional profiles in peripheral blood mononuclear cells (PBMCs) from these patients before clinical decompensation occurred. This approach enabled the production of high-quality, single-cell accessible chromatin and transcriptional profiles across the major immune cell types in PBMCs and facilitated analysis of how memory or effector cells are established in both the innate and adaptive immune responses of individuals with diverging patterns of clinical illness.[15]

We first visualized the distribution of immune cell populations based on the transcriptional or epigenetic scores of canonical lineage markers using uniform manifold approximation and projection (UMAP) plots, grouped by cell type or experimental cohort. Despite being clinically indistinguishable at the time of enrollment and sampling, patients who would go on to decompensate due to COVID-19 already exhibited marked alterations in cell-type-specific gene expression and chromatin accessibility compared with patients who would remain stable or improve (Figure 1; Table S1). Patients who would later decompensate exhibited a reduction in the number of B cells and CD4+ memory and naive T cells, but increased numbers of CD8+ T cells and monocytes (Table S1). Of the three modalities examined, scRNA-seq and the combined multiomic approach demonstrated the most prominent differences between the two cohorts.

Utilizing sparse logistic regression models, cell-type-specific signatures associated with clinical worsening were developed and validated by nested leave-one-out cross-validation (LOOCV). Using a multiomic approach (combined scRNA-seq and scATAC-seq), the two cohorts could be discriminated with a high degree of accuracy (Figure 2). The strongest discriminatory signal was seen in CD4+ memory T cells (auROC 0.94) and inflammatory T cells (auROC 0.87), followed by NK cells (auROC 0.87) (Figures 2 and S1; Table 2). Aggregating across cell types utilizing the combined multiome platform showed even stronger predictive potential than individual cell types alone (auROC 1.0). However, among the worsening cohort, no significant differences were noted between individuals who died in the hospital (n = 6) and those who did not (n = 8) with any modality tested. Mechanistically, cohort-specific differences in biological activation varied across modalities, ranging from pathways governing immune function, cell surface signaling, and defense response in CD4+ T cells and inflammatory T cells to cell communication and signal transduction and development in dendritic cells and CD8+ T cells (Figures 3 and S2)).
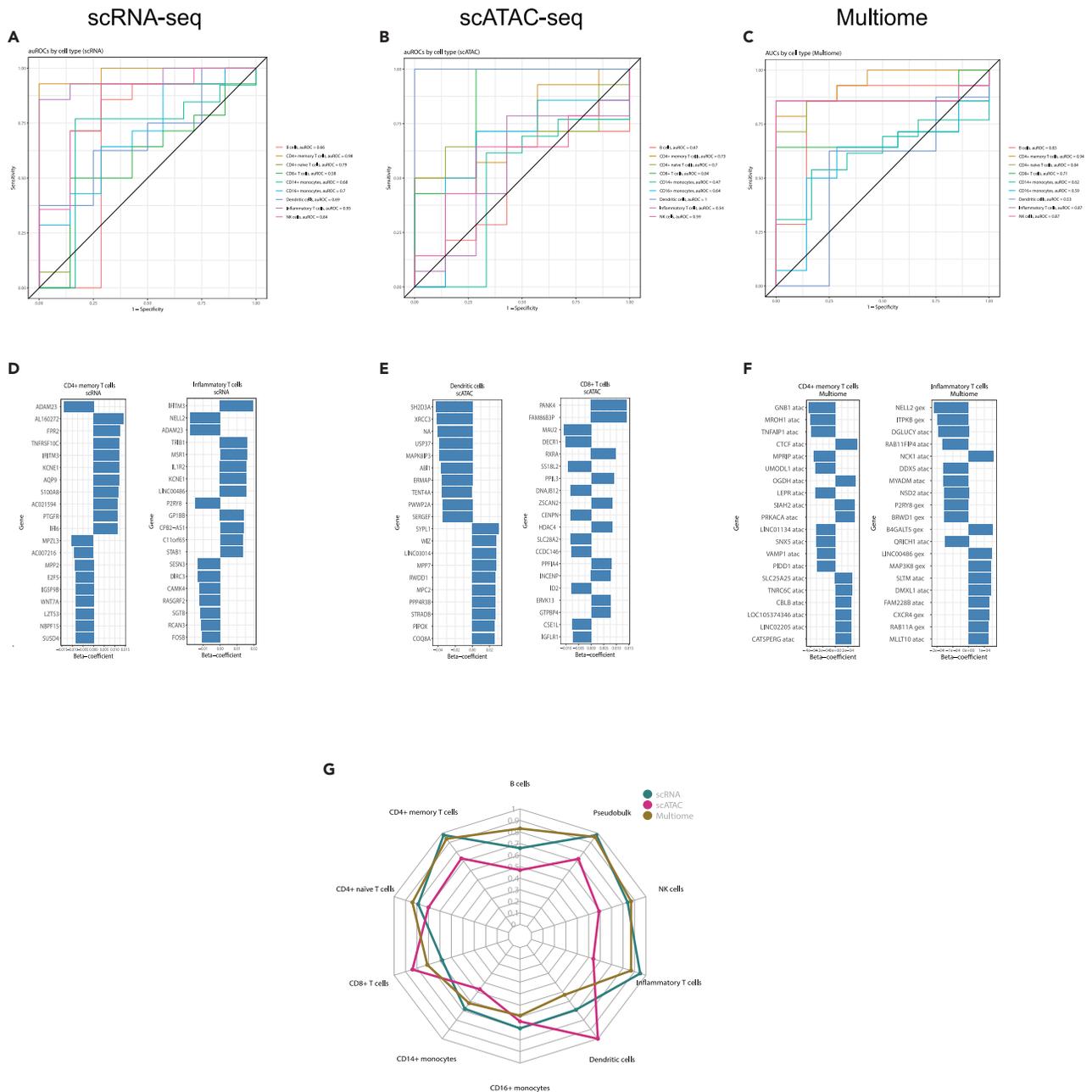
Figure 2. Performance and features of top diagnostic signatures

(Top row) Plots depicting the area under a receiver operating characteristic (auROC) curve show model performance using data for single-cell type that was normalized and log-transformed CPM (counts per million) to generate 1,000 genes with (A) the most variable differential expression (scRNA-seq), (B) regions of accessible chromatin (scATAC-seq), and (C) multiome composite data. Model performance for individual cell types (color coded), including B cells, CD4[+] memory T cells, CD4[+] naïve T cells, CD8[+] T cells, CD14[+] monocytes, CD16[+] monocytes, dendritic cells, inflammatory T cells, and NK cells, is shown in Table 2. (Second row) Top 10 genes with positive model's beta-coefficients and top 10 genes (by absolute value) with negative model's beta-coefficients that drive the model performance for the two most predictive cell types for (D) CD4[+] memory T cells and inflammatory T cells for scRNA-seq data, (E) dendritic cells and CD8[+] T cells for scATAC-seq data, and (F) CD4[+] memory T cells and inflammatory T cells for the multiome data. (G) Radar plot of relative performance of scRNA-seq data (green), scATAC-seq data (pick), and multiome composite data (brown), plotted by each immune cell type, shown in Table 2.

scRNA-seq alone predicts future decompensation

The largest quantitative changes in transcription across the two groups were seen in CD14[+] monocytes (Figure 1), which included upregulation of markers of both inflammation (*CD38, IL1R*) and tolerance (*CD163, IL8,* and others).[19] Interestingly, despite the strong

Table 2. Prognostic performance (auROC) of the top signatures for each cell type

Cell type	RNA-seq $\mathbb{E}[\,p(x)]$	ATAC-seq $\mathbb{E}[\,p(x)]$	Multiome (RNA-seq and ATAC-seq)
B cells	0.66	0.47	0.83
CD14$^+$ monocytes	0.68	0.47	0.62
CD16$^+$ monocytes	0.70	0.64	0.59
CD4$^+$ memory T cells	**0.98**	0.73	**0.94**
CD4$^+$ naive T cells	0.79	0.70	0.84
CD8$^+$ T cells	0.58	**0.84**	0.71
Dendritic cells	0.69	**1.00**	0.53
Inflammatory T cells	**0.95**	0.54	**0.87**
NK cells	0.84	0.59	0.87
Cell-type aggregation	1.00	0.71	1.00
Pseudo-bulk	0.98	0.72	0.96

severity-associated activation of CD14$^+$ monocytes in some individuals, monocyte-based signatures performed less well at predicting decompensation at early times. The most discriminatory gene expression signatures were seen in CD4$^+$ memory T cells (auROC 0.98) and a subset of inflammatory T cells that show some features of T$_H$17-like activity (expression of *IL17A*, *CCL3*, and *TLR4*, auROC 0.95),[21] followed by NK cells (auROC 0.84). NK cells are already exhibiting patterns of increased type I interferon signaling and inflammation (*IFI6*, *IFI27*), along with downregulation of tumor necrosis factor alpha (TNF-α) response pathways (TNFAIP1/2) in those who will go on to decompensate (Figures 2 and S1). Other cell types demonstrated variable strong quantitative signal differences between worsening subjects and controls but offered poor to moderate discrimination between groups at the time of hospital admission (Table 2). Nonetheless, when aggregated across cell types, gene expression signatures tested at the time of initial hospitalization offered strong prediction for future decompensation (auROC 1.0).

scATAC-seq reveals epigenetic changes associated with future decompensation

In addition to changes in gene expression, chromatin remodeling was also significantly altered in almost all immune cell compartments between the two cohorts. Dendritic cells (auROC 1.0) demonstrated the highest prognostic performance (Figure 2), although this was tempered by the relatively low numbers of these cells overall (Table S1). CD8$^+$ T cells showed the next most significant discrimination (auROC 0.84) by scATAC-seq, with all other cell types showing auROCs of 0.73 or lower (Table 2). However, as opposed to the data from scRNA-Seq and the combined multiome approach, aggregating across cell types with scATAC-seq data (auROC 0.71) did not offer improved performance compared with the top-performing individual cells.

To define the biology underlying these accessibility changes, we analyzed differentially expressed peaks to determine the associated genes for each cell type (Figure 2). Similar to the transcriptomic data, although monocyte populations exhibit strong interferon responsiveness and activity in all subjects in both cohorts as has been described,[22,23] only a subset of patients who will decompensate exhibit patterns of chromatin accessibility in CD14$^+$ monocytes that distinguish them from nonprogressors. This suggests that although monocytes are critical components of the host response to SARS-CoV-2, they do not seem to consistently act as primary indicators of early clinical decompensation in all subjects. Similar to monocytes, CD8$^+$ T cells offered suboptimal discrimination but manifested dysregulated signals that may drive some of the pathology of critical illness including strong features of a hyperinflammatory state (IFN-responsive genes including *IFI44L*), increased cytotoxicity modules (*GZMA*, *PRF1*), and early markers of T cell exhaustion (*HAVCR*, *PDCD1*, Figure S1). Dendritic cells demonstrated activation in areas that have been shown to be directly involved in RNA viral responses (*DCP1A*, *TENT4A*, *MAFB*), including *JPT2*, which is known to be critical for SARS-CoV-2 modulation and function. NK cell gene expression and chromatin accessibility peaks, however, demonstrate early activation (and dysregulation) prior to increasing disease severity with heightened induction of interferon-stimulated genes and increased expression of cytotoxic pathways, as has been shown previously in patients already exhibiting critical illness.[15,18,24] Analysis of transcription factor (TF) motif accessibility was also performed using HOMER and applied to identify shared regulatory mechanisms (Figure S3). Transcription factors with highly accessible motifs were identified, including FRA1 (a component of TF activator protein 1 [TF AP1]) and NFIL3, which have been shown to be involved in mediating lung inflammation and macrophage activation, and ZNF683, which is known to be involved in control of HIV replication.[25–27]

Evaluation of published gene expression signatures of COVID-19 severity

A number of published gene expression signatures have been defined that differentiate severe from non-severe disease (after it has already manifested) in either COVID-19 or sepsis (in the pre-COVID era).[28–30] We next utilized our dataset to examine whether these signatures of severe disease are manifesting at earlier times before severe disease develops. Several individual genes from these published signatures
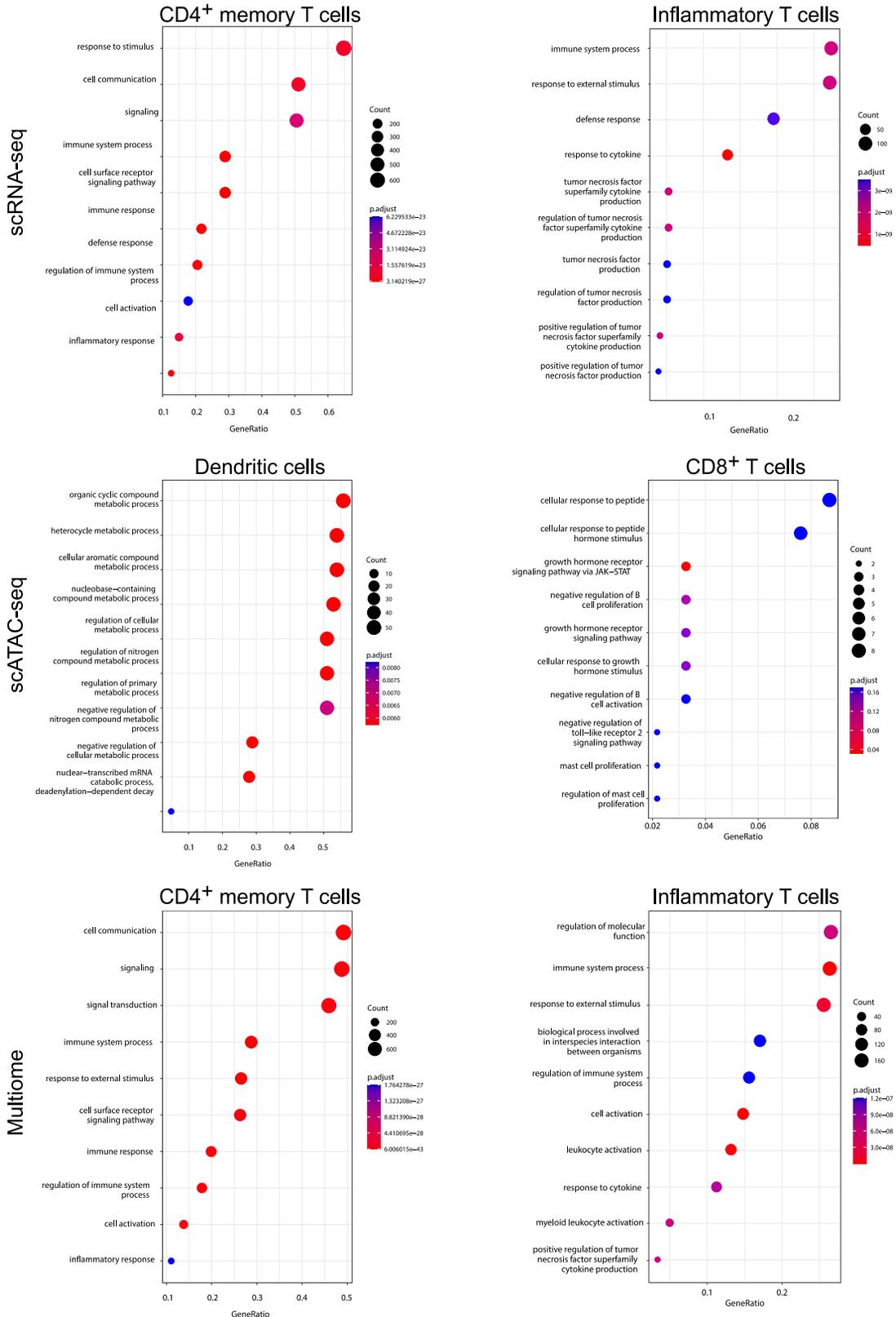
Figure 3. Dot plots show significance of biological pathways specific to the COVID-19 worsening cohort, generated using the top 1,000 most differentially expressed genes from overrepresentation tests

(Top row) CD4+ memory T cells (left) and inflammatory T cells (right) for scRNA-seq data, (center row) dendritic cells (left), and CD8+ T cells (right) for scATAC-seq data and (bottom row) CD4+ memory T cells (left) and inflammatory T cells (right) for composite multiome data.

are already differentially expressed across our two cohorts at the time of admission (Figure S4). In many of these cases, the discriminatory genes were most differentially expressed (or exhibited differential chromatin accessibility) in CD4+ T cell subsets (including *SOCS3*, *HK3*, and *BATF*). For some genes (*TGFB1*, *CD163*, and *PER1*), differential expression associated with worsening COVID-19 was observed primarily in CD14+ monocytes, reinforcing the observation that monocyte-level responses occur at early times in some individuals.

DISCUSSION

Collectively, these data demonstrate that severity-associated alterations in immune response patterns at the transcriptional and epigenetic level in circulating leukocytes are already present before clinical signs of decompensation become manifest. Thus, patients who are clinically indistinguishable at the time of hospital admission are shown to have varying patterns of underlying immune dysregulation, and these patterns are associated with, and predictive of, future clinical illness. Remarkably, alterations present in subjects who will later decompensate are present to varying degrees in most major cell type subsets, indicating broad immune dysregulation that may drive clinical worsening. Thus, these findings add to the growing body of evidence defining the nature of the profound immune dysregulation that underlies critical illness following SARS-CoV-2 infection. However, we also observed strong inter-individual variability in some signals, particularly in monocyte populations, strongly suggesting that multiple patterns or pathways of immune dysregulation may exist, which can lead to progressive clinical decompensation following SARS-CoV-2 infection. This has critical implications for pathway-focused immune-based treatment options and highlights that future work targeting individualized treatment options based on the underlying immunopathology of a specific patient may offer even better outcomes from immunomodulatory therapeutics. The significance of differences in cell-type predilection for alterations at the gene expression level compared with regions with enrichment in chromatin accessibility is unclear. One hypothesis is that for some signals, alterations in chromatin remodeling preceded those manifested at the gene expression level. This would occur in B cells, dendritic cells, and monocytes, where chromatin accessibility patterns outperform gene expression at early detection of pending decompensation. At later times (when clinical decompensation has occurred), corresponding changes in gene expression patterns downstream of these loci become apparent, which has been suggested in other works.[2,15] However, future analysis of longitudinal samples as disease severity worsens will be necessary to better define these temporal relationships.

Across modalities, the most consistent prognostic signal was seen in the T cell subsets studied and, overall, suggested a combination of early hyperinflammatory T cell responses combined with signals of impending T cell exhaustion in subjects who are developing more critical illness, consistent with phenotypes reported previously.[31] This is in contrast to other works that have examined markers of COVID-19 severity (at the time of severe illness) and identified neutrophil-driven signatures highlighted by interferon response elements.[32,33] DNA methylation signatures (from whole blood) that focus on genomic regions involved in antiviral responses and interferon signaling were associated with downstream clinical outcomes in hospitalized adults,[32] whereas purified neutrophils demonstrate inflammatory signatures that are also associated with disease severity.[33] Herein, we utilized a complimentary approach (single-cell analysis of purified PBMCs, which lack a significant neutrophil component) and found that multiple additional inflammatory cell types are heavily involved in progression to, and the pathogenesis of, severe/critical COVID-19. Comparatively, the discriminatory performance of the multiomic approach reported here is higher than that reported with methylation profiles and similar to that seen with combined clinical variables and neutrophil-specific gene expression signatures, although this is potentially confounded by the size of the current study. If validated in larger studies, our findings would suggest the potential power of merging cell-type-specific gene expression and chromatin accessibility for prognostic purposes.

While not providing the best discrimination at early times in this study, NK cells are already manifesting some of the inflammatory changes that have been described at later times once severe/critical disease is manifest.[5,34] This includes increased expression of a number of interferon-stimulated gene pathways as well as gene that drive IFN-α signaling in those who will develop severe/critical COVID-19, coupled with a marked downregulation of TNF-related genes, as has also been described.[18] The presence of dysfunctional NK cells at early times prior to decompensation in many patients suggests that these responses are potentially more closely involved in the onset and pathogenesis of clinical worsening than has previously been thought.[24]

Similarly, at earlier times when patients are not yet critically ill, some subjects are already exhibiting strong activation of CD14+ monocytes, but these alterations were not consistently expressed in those who would later decompensate. Interestingly, once critical illness has already manifested, CD14+ and CD16+ monocytes, along with dendritic cells, have also been shown to contain signals that allow for prediction of mortality.[35] Our study suggests that during early times prior to decompensation, multiple genomic regions, particularly those related to accelerated inflammation and immune exhaustion in T cell subsets, offer more reliable predictions of future clinical worsening.

In conclusion, SARS-CoV-2 infection triggers robust and often dysregulated responses in PBMCs across phases of clinical illness. The current work shows that epigenetic and transcriptomic signals in these circulating cells are detectable well before the clinical changes associated with them become apparent. The potential for a prognostic approach based on these findings is profound, suggesting that measurement of expression levels of a small set of transcriptional and epigenetic features could both detect SARS-CoV-2 infection and simultaneously provide information about impending clinical decompensation, thus allowing for earlier and more targeted application of therapeutic interventions.

Limitations of the study

The findings of this study are limited by the relatively small sample size (n = 21). One result of this is the finding of numerical (but not significant) differences in comorbidities between the two groups, including higher levels of immunosuppression and malignancy in the worsening cohort and more CAD and diabetes in the control cohort (Table 1). However, despite this limitation, the multiomic approach identifies transcriptional and epigenetic signatures that significantly discriminate patients by eventual clinical outcome. Many of the immune signals demonstrated prior to decompensation are consistent with those reported in other studies to be active later during active critical illness and further support the veracity of these findings.[2,3,11] Additionally, most cases in the study occurred in mid-late 2020 and thus represent primarily pre-Delta strains (mostly Alpha/Beta by epidemiology). However, although the rates of disease severity have varied widely across strains, once severe illness has occurred, similarities in dysregulated immune profiles have been reported over time. This holds true even as circulating strains have evolved, suggesting that biological pathways associated with severe illness seem to be more conserved than incidence of disease severity itself.[2,4,36,37] The study was also unable to evaluate whether any of these signals offer more broad applicability to other respiratory viral infections. Validation studies with larger cohorts, emerging variants, and even other viruses will be necessary to conform these findings and evaluate broad applicability of these biomarker signatures.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Institutional review board approvals
  - Clinical cohort enrollment
- METHOD DETAILS
  - PBMC cell nuclei preparation
  - 10X ATAC+GEX library preparation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Single-cell (sc) sequencing data processing
  - Predictive modeling
  - Transcriptome analysis
  - Gene Enrichment and Pathway Analysis
  - Motif analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108288.

AUTHOR CONTRIBUTIONS

M.T.M., T.W.B., X.S., and C.W.W. conceived and implemented the study. All authors helped acquire, analyze, or interpret data. R.H., I.Z., and N.S.G. performed statistical analyses. M.T.M., L.L.S., R.H., E.L.T., and C.W.W. drafted the manuscript, which was critically revised by all other authors, I.Z., N.S.G., S.D., T.W.B., C.N., J.P.B., E.A.P., and X.S.

DECLARATION OF INTERESTS

M.T.M., E.L.T., T.W.B., and C.W.W. have patents pending on Molecular Methods to Diagnose and Treat Respiratory Infections.

REFERENCES

1. Tay, M.Z., Poh, C.M., Rénia, L., MacAry, P.A., and Ng, L.F.P. (2020). The trinity of COVID-19: immunity, inflammation and intervention. Nat. Rev. Immunol. *20*, 363–374.

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

3. Yao, C., Bora, S.A., Parimon, T., Zaman, T., Friedman, O.A., Palatinus, J.A., Surapaneni, N.S., Matusov, Y.P., Cerro Chiang, G., Kassar, A.G., et al. (2021). Cell-Type-Specific Immune Dysregulation in Severely Ill COVID-19 Patients. Cell Rep. *34*, 108590.

4. Carapito, R., Li, R., Helms, J., Carapito, C., Gujja, S., Rolli, V., Guimaraes, R., Malagon-Lopez, J., Spinnhirny, P., Lederle, A., et al. (2022). Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort. Sci. Transl. Med. *14*, eabj7521.

5. Zhang, S., Cooper-Knock, J., Weimer, A.K., Shi, M., Kozhaya, L., Unutmaz, D., Harvey, C., Julian, T.H., Furini, S., Frullanti, E., et al. (2022). Multiomic analysis reveals cell-type-specific molecular determinants of COVID-19 severity. Cell Syst. *13*, 598–614.e6.

6. Wang, S., Yao, X., Ma, S., Ping, Y., Fan, Y., Sun, S., He, Z., Shi, Y., Sun, L., Xiao, S., et al. (2021). A single-cell transcriptomic landscape of the lungs of patients with COVID-19. Nat. Cell Biol. *23*, 1314–1328.

7. Wen, W., Su, W., Tang, H., Le, W., Zhang, X., Zheng, Y., Liu, X., Xie, L., Li, J., Ye, J., et al. (2020). Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. Cell Discov. *6*, 31.

8. Kgatle, M.M., Lawal, I.O., Mashabela, G., Boshomane, T.M.G., Koatale, P.C., Mahasha, P.W., Ndlovu, H., Vorster, M., Rodrigues, H.G., Zeevaart, J.R., et al. (2021). COVID-19 Is a Multi-Organ Aggressor: Epigenetic and Clinical Marks. Front. Immunol. *12*, 752380.

9. Onoja, A., Picchiotti, N., Fallerini, C., Baldassarri, M., Fava, F., GEN-COVID Multicenter Study, Colombo, F., Chiaromonte, F., Renieri, A., Furini, S., and Raimondi, F. (2022). An explainable model of host genetic interactions linked to COVID-19 severity. Commun. Biol. *5*, 1133.

10. Guo, C., Wu, M., Huang, B., Zhao, R., Jin, L., Fu, B., Wang, P., Wang, D., Zheng, M., Fang, J., et al. (2022). Single-cell transcriptomics reveal a unique memory-like NK cell subset that accumulates with ageing and correlates with disease severity in COVID-19. Genome Med. *14*, 46.

11. Corley, M.J., Pang, A.P.S., Dody, K., Mudd, P.A., Patterson, B.K., Seethamraju, H., Bram, Y., Peluso, M.J., Torres, L., Iyer, N.S., et al. (2021). Genome-wide DNA methylation profiling of peripheral blood reveals an epigenetic signature associated with severe COVID-19. J. Leukoc. Biol. *110*, 21–26.

12. Daamen, A.R., Bachali, P., Owen, K.A., Kingsmore, K.M., Hubbard, E.L., Labonte, A.C., Robl, R., Shrotri, S., Grammer, A.C., and Lipsky, P.E. (2021). Comprehensive transcriptomic analysis of COVID-19 blood, lung, and airway. Sci. Rep. *11*, 7052.

13. Li, S., Wu, B., Ling, Y., Guo, M., Qin, B., Ren, X., Wang, C., Yang, H., Chen, L., Liao, Y., et al. (2021). Epigenetic Landscapes of Single-Cell Chromatin Accessibility and Transcriptomic Immune Profiles of T Cells in COVID-19 Patients. Front. Immunol. *12*, 625881.

14. McClain, M.T., Constantine, F.J., Henao, R., Liu, Y., Tsalik, E.L., Burke, T.W., Steinbrink, J.M., Petzold, E., Nicholson, B.P., Rolfe, R., et al. (2021). Dysregulated transcriptional responses to SARS-CoV-2 in the periphery. Nat. Commun. *12*, 1079.

15. Wilk, A.J., Lee, M.J., Wei, B., Parks, B., Pi, R., Martínez-Colón, 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*, e20210582.

16. Lee, J.S., Park, S., Jeong, H.W., Ahn, J.Y., Choi, S.J., Lee, H., Choi, B., Nam, S.K., Sa, M., Kwon, J.-S., et al. (2020). Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. Sci. Immunol. *5*, eabd1554.

17. Ren, X., Wen, W., Fan, X., Hou, W., Su, B., Cai, P., Li, J., Liu, Y., Tang, F., Zhang, F., et al. (2021). COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. Cell *184*, 1895–1913.e19.

18. Krämer, B., Knoll, R., Bonaguro, L., ToVinh, M., Raabe, J., Astaburuaga-García, R., Schulte-Schrepping, J., Kaiser, K.M., Rieke, G.J., Bischoff, J., et al. (2021). Early IFN-α signatures and persistent dysfunction are distinguishing features of NK cells in severe COVID-19. Immunity *54*, 2650–2669.e14.

19. Gómez-Rial, J., Currás-Tuala, M.J., Rivero-Calle, I., Gómez-Carballa, A., Cebey-López, M., Rodríguez-Tenreiro, C., Dacosta-Urbieta, A., Rivero-Velasco, C., Rodríguez-Núñez, N., Trastoy-Pena, R., et al. (2020). Increased Serum Levels of sCD14 and sCD163 Indicate a Preponderant Role for Monocytes in COVID-19 Immunopathology. Front. Immunol. *11*.

20. Cappuccio, A., Chawla, D.G., Chen, X., Rubenstein, A.B., Cheng, W.S., Mao, W., Burke, T.W., Tsalik, E.L., Petzold, E., Henao, R., et al. (2022). Multi-objective optimization identifies a specific and interpretable COVID-19 host response signature. Cell Syst. *13*, 989–1001.e8.

21. Hu, D., Notarbartolo, S., Croonenborghs, T., Patel, B., Cialic, R., Yang, T.-H., Aschenbrenner, D., Andersson, K.M., Gattorno, M., Pham, M., et al. (2017). Transcriptional signature of human pro-inflammatory TH17 cells identifies reduced IL10 gene expression in multiple sclerosis. Nat. Commun. *8*, 1600.

22. 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 *182*, 1419–1440.e23.

23. Vanderbeke, L., Van Mol, P., Van Herck, Y., De Smet, F., Humblet-Baron, S., Martinod, K., Antoranz, A., Arijs, I., Boeckx, B., Bosisio, F.M., et al. (2021). Monocyte-driven atypical cytokine storm and aberrant neutrophil activation as key mediators of COVID-19 disease severity. Nat. Commun. *12*, 4117.

24. Bi, J. (2022). NK cell dysfunction in patients with COVID-19. Cell. Mol. Immunol. *19*, 127–129.

25. He, Y.Y., Zhou, H.F., Chen, L., Wang, Y.T., Xie, W.L., Xu, Z.Z., Xiong, Y., Feng, Y.Q., Liu, G.Y., Li, X., et al. (2022). The Fra-1: Novel role in regulating extensive immune cell states and affecting inflammatory diseases. Front. Immunol. *13*, 954744.

26. Kobayashi, T., Matsuoka, K., Sheikh, S.Z., Elloumi, H.Z., Kamada, N., Hisamatsu, T., Hansen, J.J., Doty, K.R., Pope, S.D., Smale, S.T., et al. (2011). NFIL3 is a regulator of IL-12 p40 in macrophages and mucosal immunity. J. Immunol. *186*, 4649–4655.

27. Lu, Y., Zhang, M.X., Pang, W., Song, T.Z., Zheng, H.Y., Tian, R.R., and Zheng, Y.T. (2022). Transcription Factor ZNF683 Inhibits SIV/HIV Replication through Regulating IFNγ Secretion of CD8+ T Cells. Viruses *14*.

28. Buturovic, L., Zheng, H., Tang, B., Lai, K., Kuan, W.S., Gillett, M., Santram, R., Shojaei, M., Almansa, R., Nieto, J.Á., et al. (2022). A 6-mRNA host response classifier in whole blood predicts outcomes in COVID-19 and other acute viral infections. Sci. Rep. *12*, 889.

29. Peterson, D.R., Baran, A.M., Bhattacharya, S., Branche, A.R., Croft, D.P., Corbett, A.M., Walsh, E.E., Falsey, A.R., and Mariani, T.J. (2023). Gene Expression Risk Scores for COVID-19 Illness Severity. J. Infect. Dis. *227*, 322–331.

30. Sweeney, T.E., Perumal, T.M., Henao, R., Nichols, M., Howrylak, J.A., Choi, A.M., Bermejo-Martin, J.F., Almansa, R., Tamayo, E., Davenport, E.E., et al. (2018). A community approach to mortality prediction in sepsis via gene expression analysis. Nat. Commun. *9*, 694.

31. Diao, B., Wang, C., Tan, Y., Chen, X., Liu, Y., Ning, L., Chen, L., Li, M., Liu, Y., Wang, G., et al. (2020). Reduction and Functional Exhaustion of T Cells in Patients With Coronavirus Disease 2019 (COVID-19). Front. Immunol. *11*, 827.

32. Konigsberg, I.R., Barnes, B., Campbell, M., Davidson, E., Zhen, Y., Pallisard, O., Boorgula, M.P., Cox, C., Nandy, D., Seal, S., et al. (2021). Host methylation predicts SARS-CoV-2 infection and clinical outcome. Commun. Med. *1*, 42.

33. LaSalle, T.J., Gonye, A.L.K., Freeman, S.S., Kaplonek, P., Gushterova, I., Kays, K.R., Manakongtreecheep, K., Tantivit, J., Rojas-Lopez, M., Russo, B.C., et al. (2022). Longitudinal characterization of circulating neutrophils uncovers phenotypes associated with severity in hospitalized COVID-19 patients. Cell Rep. Med. *3*, 100779.

34. Zhang, B., Zhang, Z., Koeken, V.A.C.M., Kumar, S., Aillaud, M., Tsay, H.C., Liu, Z., Kraft, A.R.M., Soon, C.F., Odak, I., et al. (2023). Altered and allele-specific open chromatin landscape reveals epigenetic and genetic regulators of innate immunity in COVID-19. Cell Genom. *3*, 100232.

35. Amrute, J.M., Perry, A.M., Anand, G., Cruchaga, C., Hock, K.G., Farnsworth, C.W., Randolph, G.J., Lavine, K.J., and Steed, A.L. (2022). Cell specific peripheral immune responses predict survival in critical COVID-19 patients. Nat. Commun. *13*, 882.

36. Cao, X., Li, W., Wang, T., Ran, D., Davalos, V., Planas-Serra, L., Pujol, A., Esteller, M., Wang, X., and Yu, H. (2022). Accelerated biological aging in COVID-19 patients. Nat. Commun. *13*, 2135.

37. Wargodsky, R., Dela Cruz, P., LaFleur, J., Yamane, D., Kim, J.S., Benjenk, I., Heinz, E., Irondi, O.O., Farrar, K., Toma, I., et al. (2022). RNA Sequencing in COVID-19 patients identifies neutrophil activation biomarkers as a promising diagnostic platform for infections. PLoS One *17*, e0261679.

38. 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.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human peripheral blood mononuclear cells	This paper	N/A
Chemicals, peptides, and recombinant proteins		
Lysis Buffer (10 mM Tris-HCl, pH 7.4 with 10mM NaCl, 3mM MgCl2, 0.01% Tween-20, 0.1% NP-40m 0.01% Digitonin, 1% BSA, 1mM DTT, and 1U/ul RNase inhibitor)	10X Genomics	N/A
1X Diluted Nuclei Buffer	10X Genomics	N/A
Critical commercial assays		
TruSeq ChIP Sample Prep Kit	Illumina	N/A
Chromium Single Cell Multiome ATAC + Gene Expression Kit	10X Genomics	CG000365 and CG000338
2100 Expert High Sensitivity DNA Assay	Agilent	DE54108252
Deposited data		
Raw and analyzed data	This paper	GSE215915
Data analysis code	This paper	https://doi.org/10.5281/zenodo.8305391
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
Oligonucleotides		
Primer sequences 10X ATAC + RNA seq kit (CG000338)	Illumina	CG000338
Software and algorithms		
R	R Core Team (2022)	https://www.R-project.org/
glmnet	Friedman et al.	https://www.jstatsoft.org/v33/i01/
umap	Konopka et al.	https://cran.r-project.org/package=umap
ClusterProfiler	Wu et al.	http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
HOMER	Heinz et al.[38]	http://homer.ucsd.edu/homer/motif/
Seurat 3	Hoffman et al.	https://satijalab.org/seurat/
ArchR	Granja et al.	https://www.archrproject.com
Harmony	Korsunsky et al.	https://portals.broadinstitute.org/harmony
CellRanger	10xgenomics	10xgenomics.com
Bowtie2	Langmead and Salzberg	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools	Li et al.[13]	http://samtools.sourceforge.net/
Other		
Countess II FL Automated Cell Counter	ThermoFisher	N/A
Agilent Bioanalyzer High Sensitivity DNA Chip	Agilent Technologies	DE54108252
Illumina NovaSeq 6000	Illumina	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr Micah McClain (micah.mcclain@duke.edu).

Materials availability

All unique/stable reagents generated in this study are available from the lead contact.

Data and code availability

- Raw and analyzed data can be accessible though GEO repository accession number GSE215915.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Institutional review board approvals

The relevant protocols were approved by the IRBs of participating institutions and were conducted in accordance with the Declaration of Helsinki, applicable regulations, and local policies.

Clinical cohort enrollment

Adult patients with confirmed SARS-CoV-2 infection were identified through the Duke University Health System (DUHS) and enrolled into the Molecular and Epidemiological Study of Suspected Infection (MESSI, Pro00100241). Participant demographics and clinical features are detailed in Table 1. Written informed consent was obtained from all subjects or legally authorized representatives. RT-PCR testing for SARS-CoV-2 was performed at either the North Carolina State Laboratory of Public Health or through the clinical laboratory at DUHS. Patients were adjudicated to have COVID-19 based on a positive PCR for SARS-CoV-2 along with an appropriate clinical syndrome and absence of additional acute cause of respiratory symptoms. Patients were approached upon admission for enrollment and followed longitudinally throughout their hospital stay. Samples analyzed represent blood draws from initial time of enrollment, which are matched to initial WHO Ordinal scores from that same time. Clinical deterioration was defined as an increase of 1 or more from baseline enrollment score as measured by the ordinal WHO clinical progression scale: 4, hospitalized, not requiring supplemental oxygen but requiring ongoing medical care (related to Covid-19 or to other medical conditions); 5, hospitalized, requiring any supplemental oxygen; 6, hospitalized, requiring noninvasive ventilation or use of high-flow oxygen devices; 7, hospitalized, receiving invasive mechanical ventilation or extracorporeal membrane oxygenation (ECMO); and 8, death.

METHOD DETAILS

PBMC cell nuclei preparation

Cryopreserved PBMCs were washed and counted using Countess II FL Automated Cell Counter (ThermoFisher). Nuclei were prepared from 600,000 cells by incubation and wash with chilled Lysis Buffer (10 mM Tris-HCl, pH 7.4 with 10mM NaCl, 3mM MgCl2, 0.01% Tween-20, 0.1% NP-40m 0.01% Digitonin, 1% BSA, 1mM DTT, and 1U/ul RNase inhibitor). Intact nuclei were collected by centrifugation, resuspended in 1X Diluted Nuclei Buffer (10X Genomics) with 1mM DTT and 1U/uL RNase inhibitor, and nuclei concentration and viability determined using Countess II FL with trypan blue.

10X ATAC+GEX library preparation

Simultaneous measure of gene expression (GEX) and chromatin accessibility (ATAC) from the same PBMC nuclei was performed using Chromium Single Cell Multiome ATAC + Gene Expression Kit according to manufacturer's suggested protocols (protocols CG000365 and CG000338; 10X Genomics, Pleasanton, CA). Cell nuclei were prepared at targeted concentration of up to 10,000 nuclei/sample) and incubated with Transposition Mix at 37C for 60 min. Transposed nuclei were combined with Master Mix and loaded into Chromium Next GEM Chip, along with Gel Beads strip, and transposed nuclei-GEM emulsion was prepared using Chromium Controller. Resulting single-nuclei GEMs were incubated to lyse nuclei and produce 10X Barcoded DNA from transposed DNA (for ATAC) and 10X Barcoded cDNA from poly-adenylated mRNA (for GEX), including unique molecular identified (UMI unique to each nucleus) and Illumina P5 (ATAC) or TruSeq Read 1 (GEX) sequencing primer sequences. GEMs were then broken and barcoded, transposed DNA and cDNA were extracted and pre-amplified (7 cycles) to yield input for ATAC and GEX library preparations. ATAC libraries and GEX cDNA libraries were prepared separately by PCR with library fragmentation, end repair, tailing, adaptor ligation, and sample index PCR, according to kit instructions. Library fragment size and quality assessed using Agilent Bioanalyzer High Sensitivity DNA Chip. ATAC libraries were sequenced on Illumina NovaSeq 6000 instrument with S2 or S4 flow cell, with 25,000 paired-end 50 bp reads per nuclei. GEX libraries were sequenced on Illumina NovaSeq 6000 instrument with S2 or S4 flow cell, with 20,000 paired-end 100 bp reads per nuclei.

QUANTIFICATION AND STATISTICAL ANALYSIS

Single-cell (sc) sequencing data processing

Libraries were then pooled and sequenced using Illumina NovaSeq 6000 platform with the goal of reaching saturation or 20,000 unique reads per cell on average. Sequencing data were used as input to the 10x Genomics Cell Ranger pipeline to demultiplex BCL files, generate FASTQs, and generate feature counts for each library. For dimensionality reduction and cell type annotation, gene-barcode matrices generated using Cell Ranger count were analyzed using Seurat 3 with the default parameters. Cells with > 5% of reads mapping to the mitochondrial genome or > 2,500 genes detected were removed from the analysis. Counts were log-normalized, and the top 2,000 variable features were identified. Principal component analysis was performed using these variable genes, and the top 20 principal components were used for downstream analysis. UMAP dimensionality reduction was performed using the top 20 principal components identified using the Harmony package. Graph-based clustering was performed with resolution = 0.5. Cell types were inferred by using the DatabaseImmuneCellExpression-Data() function from the SingleR package. Labels were confirmed by identification of differentially expressed genes using the FindAllMarkers() function from Seurat. For scRNA-seq and scATAC-seq integration, fragment file outputs generated using CellRanger ATAC count were analyzed using ArchR following the standard workflow and with default parameters unless otherwise specified. Cells with a transcription start site enrichment score < 4, cells with fewer than 1000 detected fragments, and putative doublets were removed from downstream analysis. Dimensionality reduction was computed using iterative latent semantic indexing (LSI), and batch effect correction was applied using Harmony. Graph-based clustering was performed using the FindClusters() method from Seurat 3 with resolution = 0.8. UMAP embeddings were calculated with the top 30 principal components from either LSI or Harmony. scRNA-seq cell type annotations were directly transferred to label the scATAC-seq data. Pseudo-scRNA-seq gene expression profiles were estimated for each cell type using the addGeneIntegrationMatrix() function from ArchR during the constrained label transfer procedure.

Predictive modeling

Cell-type level signatures were obtained via sparse logistic regression. To select the regularization parameters of the regression model controlling sparsity (signature size), and to estimate performance characteristics, we used nested leave-one-out cross-validation (LOOCV), in which the inner and outer LOOCV routines are used for regularization parameter estimation and performance estimation, respectively. For the latter use used the area under the receiving operating characteristic (auROC). Models for each modality (GEX and ATAC) and their combination (GEX+ATAC) were built and evaluated separately. For each modality, sequencing data was normalized by library size to account for differences in sequence depth across samples. To reduce the effects of small sample size on overfitting, we restricted models to the top 1000 transcripts or accessibility sites by variance. The combined model used the concatenation of transcripts and accessibility sites, each of which was normalized separately. All analyses were performed in R 4.10 (The R Foundation for Statistical Computing).

Transcriptome analysis

Transcriptome analysis was performed separately on scRNA-seq, scATAC-seq and multiome (combined scRNA-seq and scATAC-seq) data. Quality control included filtering low-expressed genes and trimmed mean normalization. Association tests were performed with R-packages limma and edgeR. The Benjamini-Hochberg (BH) procedure was used for multiple testing correction throughout the study.

Gene Enrichment and Pathway Analysis

Gene Enrichment and Pathway Analysis using overrepresentation tests were performed with R-package ClusterProfiler on selected top 1,000 highly-differentially expressed genes obtained from univariate analysis and Gene Ontology database. Analysis was performed separately for scRNA-seq, scATAC-seq and for multiome data.

Motif analysis

Motif analysis was performed using HOMER v4.11 (Heinz et al., 2010). Specifically, we focused on upstream regions and used the original scATAC-seq peaks as the background in the analysis. Other parameters were left to their default.