



OPEN

## Differential chromatin accessibility in peripheral blood mononuclear cells underlies COVID-19 disease severity prior to seroconversion

Nicholas S. Giroux<sup>1</sup>, Shengli Ding<sup>1</sup>, Micah T. McClain<sup>2,3,4</sup>, Thomas W. Burke<sup>2</sup>, Elizabeth Petzold<sup>2</sup>, Hong A. Chung<sup>1</sup>, Grecia O. Rivera<sup>1</sup>, Ergang Wang<sup>1</sup>, Rui Xi<sup>1</sup>, Shree Bose<sup>5</sup>, Tomer Rotstein<sup>1</sup>, Bradley P. Nicholson<sup>6</sup>, Tianyi Chen<sup>7</sup>, Ricardo Henao<sup>2</sup>, Gregory D. Sempowski<sup>8</sup>, Thomas N. Denny<sup>8</sup>, Maria Iglesias De Ussel<sup>2</sup>, Lisa L. Satterwhite<sup>9</sup>, Emily R. Ko<sup>2</sup>, Geoffrey S. Ginsburg<sup>2</sup>, Bryan D. Kraft<sup>2,3</sup>, Ephraim L. Tsalik<sup>2,3,4</sup>, Xiling Shen<sup>1</sup> & Christopher W. Woods<sup>2,3,4</sup>✉

SARS-CoV-2 infection triggers profound and variable immune responses in human hosts. Chromatin remodeling has been observed in individuals severely ill or convalescing with COVID-19, but chromatin remodeling early in disease prior to anti-spike protein IgG seroconversion has not been defined. We performed the Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) and RNA-seq on peripheral blood mononuclear cells (PBMCs) from outpatients with mild or moderate symptom severity at different stages of clinical illness. Early in the disease course prior to IgG seroconversion, modifications in chromatin accessibility associated with mild or moderate symptoms were already robust and included severity-associated changes in accessibility of genes in interleukin signaling, regulation of cell differentiation and cell morphology. Furthermore, single-cell analyses revealed evolution of the chromatin accessibility landscape and transcription factor motif accessibility for individual PBMC cell types over time. The most extensive remodeling occurred in CD14+ monocytes, where sub-populations with distinct chromatin accessibility profiles were observed prior to seroconversion. Mild symptom severity was marked by upregulation of classical antiviral pathways, including those regulating IRF1 and IRF7, whereas in moderate disease, these classical antiviral signals diminished, suggesting dysregulated and less effective responses. Together, these observations offer novel insight into the epigenome of early mild SARS-CoV-2 infection and suggest that detection of chromatin remodeling in early disease may offer promise for a new class of diagnostic tools for COVID-19.

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), manifests with highly variable symptom severity<sup>1</sup>. Infected subjects demonstrate clinical trajectories that range from remaining asymptomatic to developing life-threatening illness. A growing body of evidence suggests that the range of clinical manifestations is the result of different immune responses, and that analysis of chromatin accessibility and gene expression in circulating leukocytes can define underlying molecular mechanisms<sup>2</sup>. Studies have shown suppressed immune responses in subjects with only mild symptoms, as indicated by deficient

<sup>1</sup>Department of Biomedical Engineering, Pratt School of Engineering, Duke University, Durham, NC 27708, USA. <sup>2</sup>Center for Applied Genomics and Precision Medicine, Duke University School of Medicine, Durham, NC 27710, USA. <sup>3</sup>Durham Veterans Affairs Health Care System, Durham, NC 27705, USA. <sup>4</sup>Division of Infectious Diseases, School of Medicine, Duke University Medical Center, 40 Duke Medicine Circle, Durham, NC 27710-4000, USA. <sup>5</sup>Department of Pharmacology and Cancer Biology, School of Medicine, Duke University, Durham, NC 27710, USA. <sup>6</sup>Institute for Medical Research, Durham, NC 27705, USA. <sup>7</sup>Department of Molecular Genetics and Microbiology, School of Medicine, Duke University, Durham, NC 27710, USA. <sup>8</sup>Duke Human Vaccine Institute and Department of Medicine, School of Medicine, Duke University, Durham, NC 27710, USA. <sup>9</sup>Department of Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham, NC 27708, USA. ✉email: chris.woods@duke.edu

Characteristics	Healthy controls	Close contacts	Mild disease	Moderate disease
<b>A. Bulk assays for ATAC-seq and RNA-seq</b>				
Number of subjects	7	7	8	7
Age Mean (range), years	39.7 (25–61)	42.9 (17–61)	33.0 (27–60)	34.0 (20–52)
Sex (male/female)	3/4	4/3	5/3	3/4
Max severity score ( $\pm$ SE)	N/A	9.9 $\pm$ 2.9	12.8 $\pm$ 1.9	33.6 $\pm$ 2.4
Characteristics	Healthy controls	Close contacts	Mild disease	Moderate disease
<b>B. Single cell assays for ATAC-seq/RNA-seq</b>				
Number of subjects	5	3	5	5
Age Mean (range), years	45.2 (28–61)	53.2 (43–61)	35.6 (27–60)	37.7 (29–52)
Sex (male/female)	1/4	2/1	5/0	1/4
Max severity score ( $\pm$ SE)	N/A	3.3 $\pm$ 2.8	12.2 $\pm$ 2.8	36.0 $\pm$ 2.3

**Table 1.** Demographic characteristics and mean symptom severity scores of study participants.

expression of Type I and III interferons<sup>3</sup>. Subjects with more severe disease demonstrated up-regulation of pro-inflammatory factors, including IL-6 and TNF- $\alpha$ <sup>4</sup>. The number of monocytes and associated IL-6, CCL2, and CCL8 production in the peripheral blood also are elevated in subjects with severe COVID-19<sup>5,6</sup>. Regions of regulatory chromatin that contain transcription factor motifs become accessible prior to downstream gene expression and may offer the potential for even earlier detection of these evolving biological responses<sup>7</sup>.

Analyzing time-dependent changes at the epigenetic level is complicated by the variable time course of COVID-19 across patients, which even in the outpatient setting can range from a few days of asymptomatic viral shedding to prolonged febrile illness. Most studies have used clock time (days since illness onset), a measure that is broadly related to immune response stages in many subjects. Another option is to utilize direct measures of immune maturation to characterize where during the host response to infection a given subject exists at a point in time. The development of specific antibodies (IgG) against SARS-CoV-2 marks an inflection point in a COVID-19 patient's disease progression, indicating a transition from innate immunity to acquired immunity<sup>8,9</sup>. IgG seroconversion typically occurs within two weeks of symptom onset and roughly coincides with the time that patients without critical illness will see clinical improvement<sup>10,11</sup>. However, limited data are available examining the peripheral blood immune responses early in disease progression and prior to seroconversion. To test the hypothesis that the landscape of chromatin accessibility contains biomarkers that define early molecular mechanisms that underpin divergent immunologic responses in SARS-CoV-2 infection, we performed bulk and single-cell Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) and RNA-seq on longitudinal PBMC samples from patients before and after seroconversion.

## Results

**A clinical cohort to study early COVID-19 with mild or moderate symptoms.** To establish a cohort for the longitudinal study of early SARS-CoV-2 infection, subjects were selected from a large ongoing community-based prospective cohort, the Molecular Epidemiological Study of Suspected Infection (MESSI) in four experimental groups: (1) pre-pandemic healthy controls, (2) subjects exposed to SARS-CoV-2 as close contacts but negative for SARS-CoV-2 by qPCR and serology, (3) outpatients positive for SARS-CoV-2 by qPCR whose maximal illness was mild and (4) outpatients positive for SARS-CoV-2 by qPCR whose maximal illness was moderate. Demographic characteristics of study participants are shown in Table 1. Symptoms associated with COVID-19 illness were self-reported to clinic personnel using a COVID-19 symptom survey developed by Duke University Medical Center consisting of 38 symptoms. Each subject reported a symptom and perceived severity of that symptom from 0 (none) to 4 (very severe). Subjects with mild symptoms exhibited a mean score of 12.8  $\pm$  1.9 and 12.2  $\pm$  2.8 for total PBMC assays (bulk) and assays of individual PBMCs (single-cell assays), respectively, which corresponded to the World Health Organization (WHO) Ordinal 8-point scale 1 (OS1, ambulatory and with no limitation of activities). Subjects with moderate symptoms exhibited a mean score of 33.6  $\pm$  2.4 and 36.0  $\pm$  2.3 respectively which corresponded to the WHO Ordinal 8-point scale 2, (OS2, ambulatory and with limitation of activities). Interestingly, despite lacking any evidence of infection (negative SARS-CoV-2 qPCR and serological testing for at least two months after exposure), close household contacts who were exposed to SARS-CoV-2 but remained negative for infection showed broadly similar quantitative symptom scores to subjects with mild disease, with a mean score of 9.9  $\pm$  2.9 and 3.3  $\pm$  2.8 respectively (Table 1). The sum of symptom severity across all symptoms reported was able to distinguish moderate subjects from mild or close contacts but did not distinguish mild from close contacts (Fig. S1A,B). However, the specific symptoms experienced, and their severity, were different between all three groups (Fig. S1C,D). Loss of taste and smell, headache, malaise, and fatigue were most common in the moderate group. Coughing was most common in the mild group and runny nose was most common in the close contacts.

**Transcriptional profiles of PBMCs distinguish mild or moderate COVID-19 from healthy controls.** To identify underlying molecular mechanisms of disease severity in mild or moderate COVID-19 both

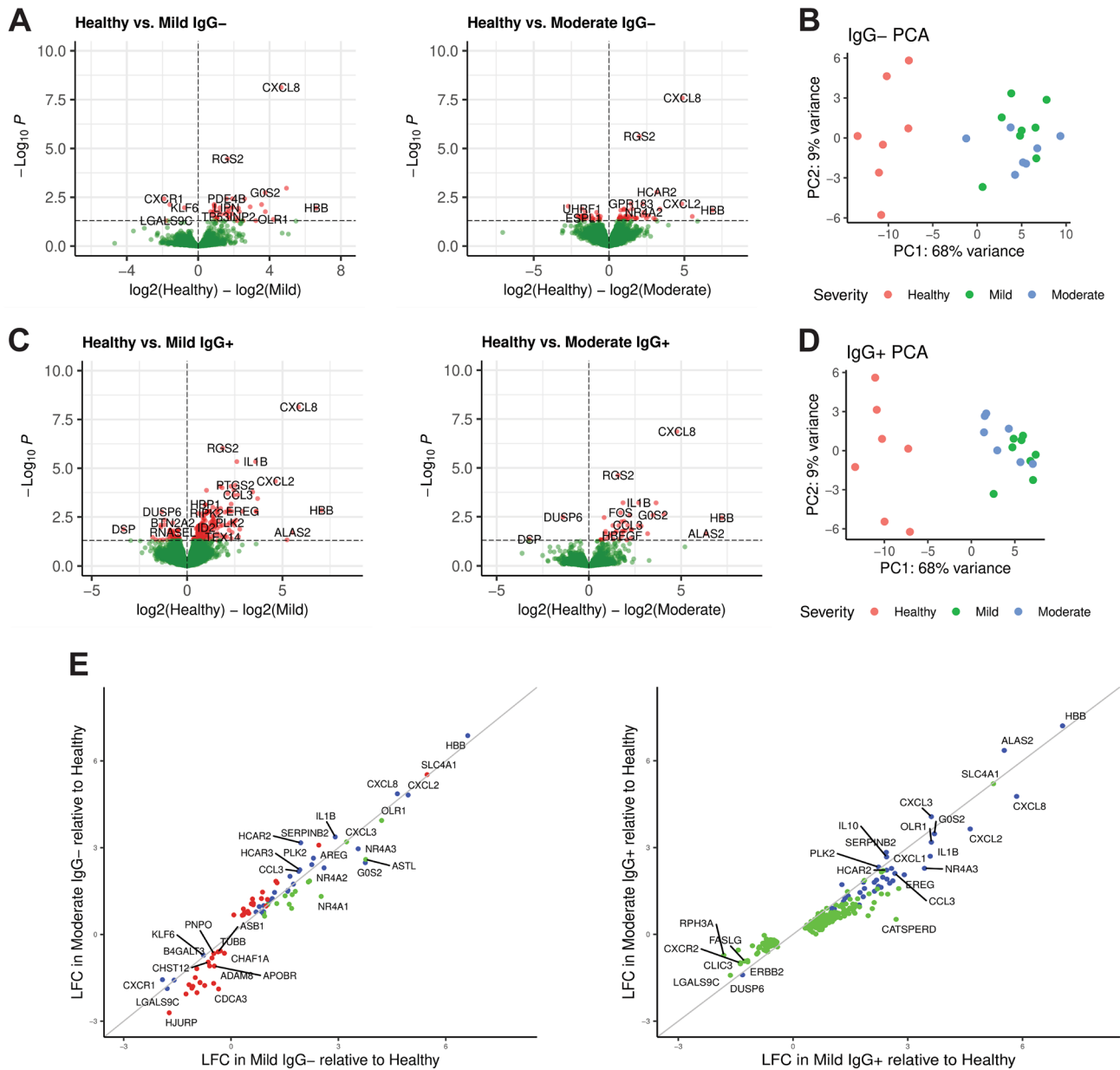
prior and after seroconversion, PBMCs were isolated and profiled as total PBMCs (bulk assays) or as single nuclei isolated from individual PBMCs (single cell assays). Longitudinal sample collection from each subject spanned the transition from seronegative (IgG<sup>-</sup>) to seropositive (IgG<sup>+</sup>) for SARS-CoV-2 spike protein in the mild and moderate symptom groups (Supplementary Table S1 and S2). For all pooled analysis of IgG<sup>-</sup> or IgG<sup>+</sup> subjects, the dataset corresponding to the last IgG<sup>-</sup> timepoint or the first IgG<sup>+</sup> timepoint, respectively, was used. Differential gene expression analysis of bulk RNA-seq data from mild IgG<sup>-</sup> COVID-19 subjects compared to healthy controls identified a transcriptional signature of 45 genes with  $p \leq 0.05$ ; moderate IgG<sup>-</sup> COVID-19 subjects compared to healthy controls identified 71 genes with  $p \leq 0.05$  (Fig. 1A). Using principal component analysis (PCA), differential gene expression between healthy controls and subjects with mild symptoms was sufficient to distinguish the groups (Fig. 1B). Similar comparisons using RNA-seq from mild IgG<sup>+</sup> subjects compared to healthy controls or moderate IgG<sup>+</sup> subjects compared to healthy controls identified 250 and 43 genes, respectively, with  $p \leq 0.05$  (Fig. 1C). Using PCA, differential gene expression between healthy controls and subjects with moderate symptoms (Fig. 1D) was sufficient to distinguish the groups. However, both mild and moderate subjects clustered together and could not be easily distinguished. The gene symbols, gene names, fold change and  $p$  values adjusted for multiple hypothesis testing for the four comparisons are found in Supplemental Table S3. Considering the level of gene expression (log fold change, LFC) in mild and moderate subjects compared to healthy controls, differentially expressed genes were identified in only mild, only moderate or both symptom cohorts (Fig. 1E). Comparing IgG<sup>-</sup> subjects, we identified 30 genes observed in both mild and moderate cohorts. In contrast, we did not identify any genes solely associated with moderate IgG<sup>+</sup> subjects and identified a further 43 genes observed in both IgG<sup>+</sup> mild and moderate cohorts. Healthy controls exhibited higher expression of the interleukin-8 precursor CXCL8 compared to other groups. Conversely, the interleukin-8 receptor, CXCR1, had higher expression in mild and moderate IgG<sup>-</sup> subjects. Lower expression of chemokine CCL3 and cytokine IL1B<sup>12</sup>, and higher expression of the myeloid cell plasticity regulator KLF6<sup>13</sup> were observed in both IgG<sup>-</sup> and IgG<sup>+</sup> COVID-19 subjects.

### Chromatin remodeling prior to seroconversion can distinguish mild from moderate COVID-19.

To test whether chromatin accessibility can distinguish COVID-19 disease severity in individuals early in their disease course (prior to seroconversion) an Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)<sup>14</sup> was used to quantify differential chromatin accessibility in three groups: subjects with mild symptoms, subjects with moderate symptoms and pre-pandemic healthy controls. A set of 455 differentially accessible chromatin regions (DARs) was found to distinguish IgG<sup>-</sup> subjects with mild symptoms from those with moderate symptoms at  $p \leq 0.05$  (Fig. 2A). Of these markers, 73 regions shared differential accessibility in mild and moderate IgG<sup>-</sup> subjects, demonstrating both severity-associated evidence of chromatin remodeling and a smaller conserved epigenetic response. Additionally, a set of 375 regions of differentially accessible chromatin was identified in IgG<sup>+</sup> subjects (Fig. 2B). The effect sizes of each DAR were compared across healthy controls, mild, and moderate IgG<sup>-</sup> subjects to determine the overlap features associated with either cohort. We determined that many of the markers that distinguish mild subjects from healthy controls also distinguished moderate subjects from healthy controls with little difference in effect size. In contrast, the markers that distinguished mild and moderate subjects from each other had higher accessibility in healthy controls and the lowest accessibility in moderate subjects. DARs specific to the comparisons in 2A–B as heat map depictions (Fig. 2C,D) for prior and post seroconversion shows overlap between mild and moderate symptom subjects compared to healthy controls. Functional enrichment was then performed for each set of differentially accessible regions to identify an association with transcription factor motifs and downstream pathways. In IgG<sup>-</sup> subjects with mild symptoms, we found an enrichment of transcription factors and pathways related to viral infection (Fig. 2E). Fewer enriched pathways were observed in the IgG<sup>-</sup> subjects with moderate symptoms. Transcription factors found to be enriched in accessible chromatin from IgG<sup>+</sup> subjects with mild symptoms were associated with dysregulation of myeloid development via HOXA3, C/EBP-B and C/EBP-D<sup>15</sup> (Fig. 2F).

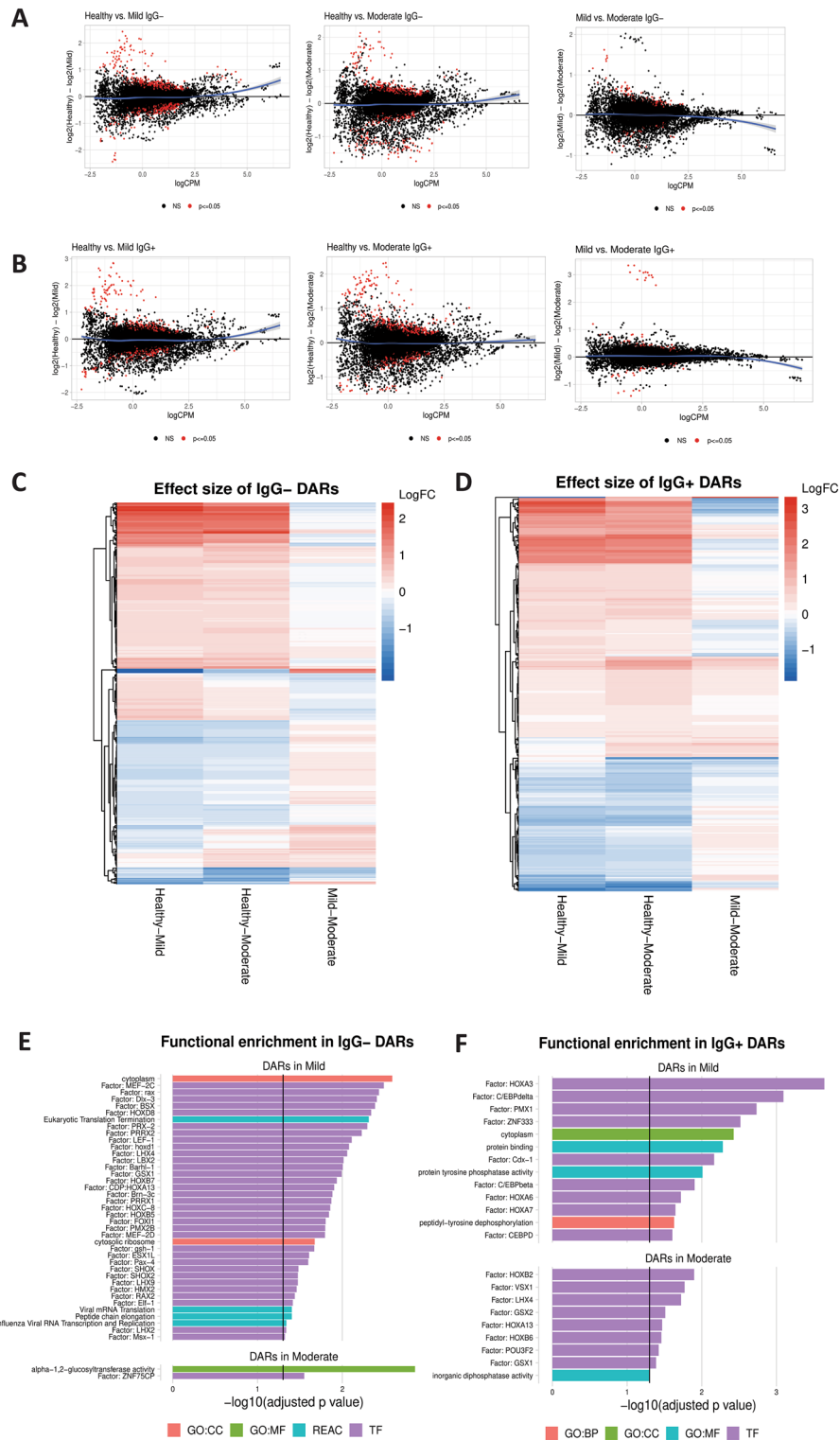
**CD14<sup>+</sup> monocytes are the primary cell type differentially activated in IgG<sup>-</sup> subjects.** The chromatin accessibility signature detected in the bulk ATAC-seq datasets identified candidate epigenetic biomarkers associated with COVID-19 severity and showed that PBMCs undergo extensive chromatin remodeling in response to SARS-CoV-2 infection. To understand how each cell type contributed to this signature and to track the evolution of the chromatin landscape of each cell type during seroconversion, we performed single-cell (sc) ATAC-seq profiling of the PBMCs isolated from subjects with mild or moderate symptoms and healthy controls (Table 1 and Supplemental Table S2). The relative abundance of cell types collected from each subject cohort were similar between IgG<sup>-</sup> and IgG<sup>+</sup> timepoints (Fig. 3A; Supplemental Tables S4 and S5). Additional details on how cell type labels were transferred from paired scRNA-seq datasets for each sample are available in the supplemental methods (Fig. S2A–D; Supplemental Table S6). Analysis of transcription factor motif accessibility was applied to PBMCs from IgG<sup>-</sup> mild or moderate subjects to identify shared regulatory mechanisms. Transcription factors with highly accessible motifs and low gene expression were identified, consistent with epigenetically primed chromatin remodeling without corresponding transcriptional activity (Fig. 3B). These motifs, including those of the CREB and ETS protein families, were characterized by increased binding activity as measured by footprint depth and relatively low average per-cell gene expression. Inflammatory regulators including JUN/FOS, proteins from the KLF family, and IRF1 showed elevated gene expression. A complete list of transcription factors associated with accessible regions and low gene expression is found in Supplemental Table S7.

Comparison of PBMCs collected from subjects with mild or moderate symptoms identified specific peaks enriched on the day of study enrollment (IgG<sup>-</sup>) versus a timepoint 2–4 weeks later (IgG<sup>+</sup>) for all cell types (Fig. 3C). A total of 5988 peaks were identified as uniquely accessible at one of the three longitudinal (IgG<sup>-</sup>,



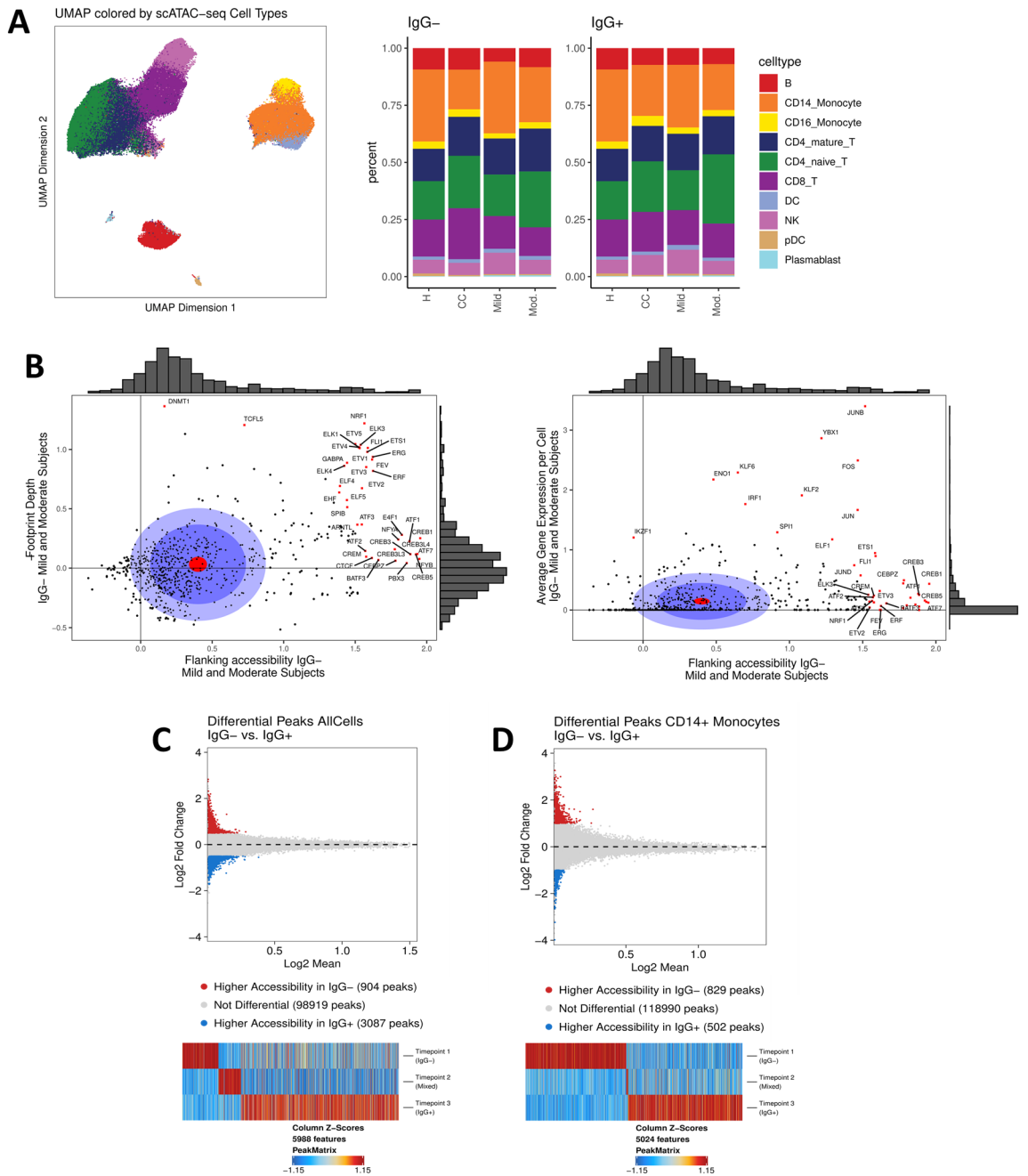
**Figure 1.** Differential gene expression in peripheral blood mononuclear cells (PBMCs) distinguishes COVID-19 symptom severity. **(A)** Volcano plot depictions of differential gene expression from comparisons of healthy controls to IgG- subjects with mild symptoms (left), and healthy controls to IgG- subjects with moderate symptoms (right). Differential expression in genes with  $p \leq 0.05$  are plotted in red. **(B)** Principal component analysis (PCA) using differentially expressed genes identified in IgG- subjects with mild symptoms (green) and moderate subjects (blue) show distinct separation from healthy controls (red). **(C)** Volcano plot depictions of differential gene expression from comparisons of healthy controls to IgG+ subjects with mild symptoms (left), and healthy controls to IgG+ subjects with moderate symptoms (right). Genes with  $p \leq 0.05$  are plotted in red. **(D)** PCA using differentially expressed genes identified in IgG+ subjects with mild symptoms (green) and moderate subjects (blue) show distinct separation from healthy controls (red). **(E)** Log fold change (LFC) in differentially expressed genes in IgG- subjects with mild symptoms as a function of IgG- subjects with moderate symptoms (left) and likewise in IgG+ subjects (right). Genes differentially expressed in mild compared to healthy are green, moderate compared to healthy are red and genes differentially expressed in mild or moderate compared to healthy controls are blue.

mixed serology, and IgG+) timepoints. A similar experimental approach was applied to each major cell type individually to determine the extent of chromatin remodeling prior to seroconversion. The cells with enrichment of significant peaks at the early, IgG- timepoint were primarily CD14+ monocytes (Fig. 3D).



**Figure 2.** Remodeling of the chromatin landscape identifies gene regulatory markers associated with symptom severity. (A) Differential chromatin accessibility, comparing healthy controls to mild symptom IgG- subjects (left) and healthy to moderate symptom IgG- subjects (center), and IgG- mild to moderate symptoms (right). Regions that are significantly different in each comparison at  $p \leq 0.05$  are red; regions not significantly different are black. A loess curve fit to the results is plotted in blue. (B) Differential chromatin accessibility, comparing healthy controls to mild symptom IgG+ subjects (left), healthy controls to moderate symptom IgG+ subjects (center), and mild to moderate IgG+ subjects (right). A loess curve fit to the results is plotted in blue. (C,D) Heat map depiction of differentially accessible regions (DARs) that compares healthy to mild symptoms (left column), healthy to moderate symptoms (center column) and mild to moderate symptoms (right column) in IgG- subjects (C) and in IgG+ (D). (E,F) Functional enrichment analysis of DARs in the TRANSFAC database (purple), the REACTOME database (red), and gene ontologies (GO) in categories of biological processes (BP, orange), cellular component (CC, blue), and molecular function (MF, green). DARs identified comparing healthy to mild symptoms (top) or healthy compared to moderate symptoms (bottom) are shown for IgG- (E) and IgG+ (F) subjects.





**Figure 3.** Epigenetic signatures in cells collected from COVID-19 subjects evolve with disease progression. **(A)** Uniform manifold approximation and projection (UMAP) plot of single-cell ATAC-seq data generated from healthy controls, uninfected close contacts, and COVID-19 subjects with mild or moderate symptoms (left). Relative abundance of cell types represented in the scATAC-seq data are plotted for each group at IgG- and IgG+ timepoints for COVID-19 subjects and collection day 0 and 14 for close contacts (right). **(B)** Transcription factor motif accessibility and binding activity measured by footprint depth (left) or average per-cell gene expression (right) in COVID-19 subjects with mild or moderate symptoms. Distribution of accessibility and gene expression are plotted as histograms along the axes (gray). The red circle is the average of all points; the dark blue circle is 50% of all data; the light blue circle is 75% of all data. Motifs with the top 5% change in flanking accessibility are plotted in red. **(C)** Differentially accessible chromatin between IgG- and IgG+ timepoints in all cells and **(D)** CD14+ monocytes. Differentially accessible chromatin observed across the IgG-, mixed and IgG+ timepoints in COVID-19 subjects is shown where significance is defined as  $p \leq 0.05$  and absolute LFC  $\geq 0.5$  (top); Changes in differentially accessible chromatin are shown across time (bottom).



with promoter-proximal transcription factor motifs (Supplemental Methods). CD14<sup>+</sup> monocyte subpopulations from subjects with mild symptoms exhibited significant up-regulation of the regions that regulate IRF7, IRF1, and STAT1 transcription factors, whereas the CD14<sup>+</sup> monocyte subpopulations from subjects with moderate symptoms were characterized by up-regulation of the CEBPB and KLF3 regulatory regions (Fig. 5B).

Single-cell ATAC-seq data for CD14<sup>+</sup> monocytes were processed using a similar approach to identify clusters with correlated gene expression regulatory activity (Fig. 5C). Activity at the transcription factor motifs was estimated by measuring genome-wide motif accessibility, and the correlations between scRNA-seq and scATAC-seq clusters were calculated in a pairwise manner (Fig. 5D). Differential gene expression between the mild and moderate sub-populations identified interferon-stimulated gene activation as a distinguishing feature of symptom severity. In addition to these transcriptomic markers, gene expression measurements of enriched monocyte-related pathways were calculated to quantify the average transcriptomic response in each cell subset (Fig. 5F, Supplemental Tables S8, S9). To perform this analysis, scRNA-seq clusters containing CD14<sup>+</sup> monocytes from both mild and moderate subjects were merged where appropriate (Supplemental Methods) and two published datasets were incorporated to enable comparison with more severe COVID-19 disease phenotypes<sup>16,17</sup>. We identified an enrichment of interferon response activity in the mild and moderate monocytes from Mild 1, Mild 2 and Moderate 1 CD14<sup>+</sup> monocyte subsets (scRNA-seq clusters 2, 4 and 7) which was consistent with our bulk RNA-seq data on the same cohorts. CD14<sup>+</sup> monocytes from subjects with severe or critical disease did not share the same strong gene expression signature of interferon signaling. Furthermore, monocytes from these clusters have elevated expression of TNF- $\alpha$ , NF- $\kappa$ B, toll-like receptor (TLR) signaling, and monocytic recruitment factor CCR2 which helps explain why these cells cluster separately in Fig. 5A. Expression of these pathways is also elevated in monocyte subsets from subjects with more severe disease, including the ICU subjects treated with TLR agonists<sup>16</sup>. Pathways identified centered on those of the adaptive immune response. Differential regulation of pathways related to the transition from innate to adaptive immunity, marked by dectin-1 and the toll-like receptor cascade, suggested that activation of T and B cell receptor signaling were observed in these sub-populations<sup>18–20</sup>.

## Discussion

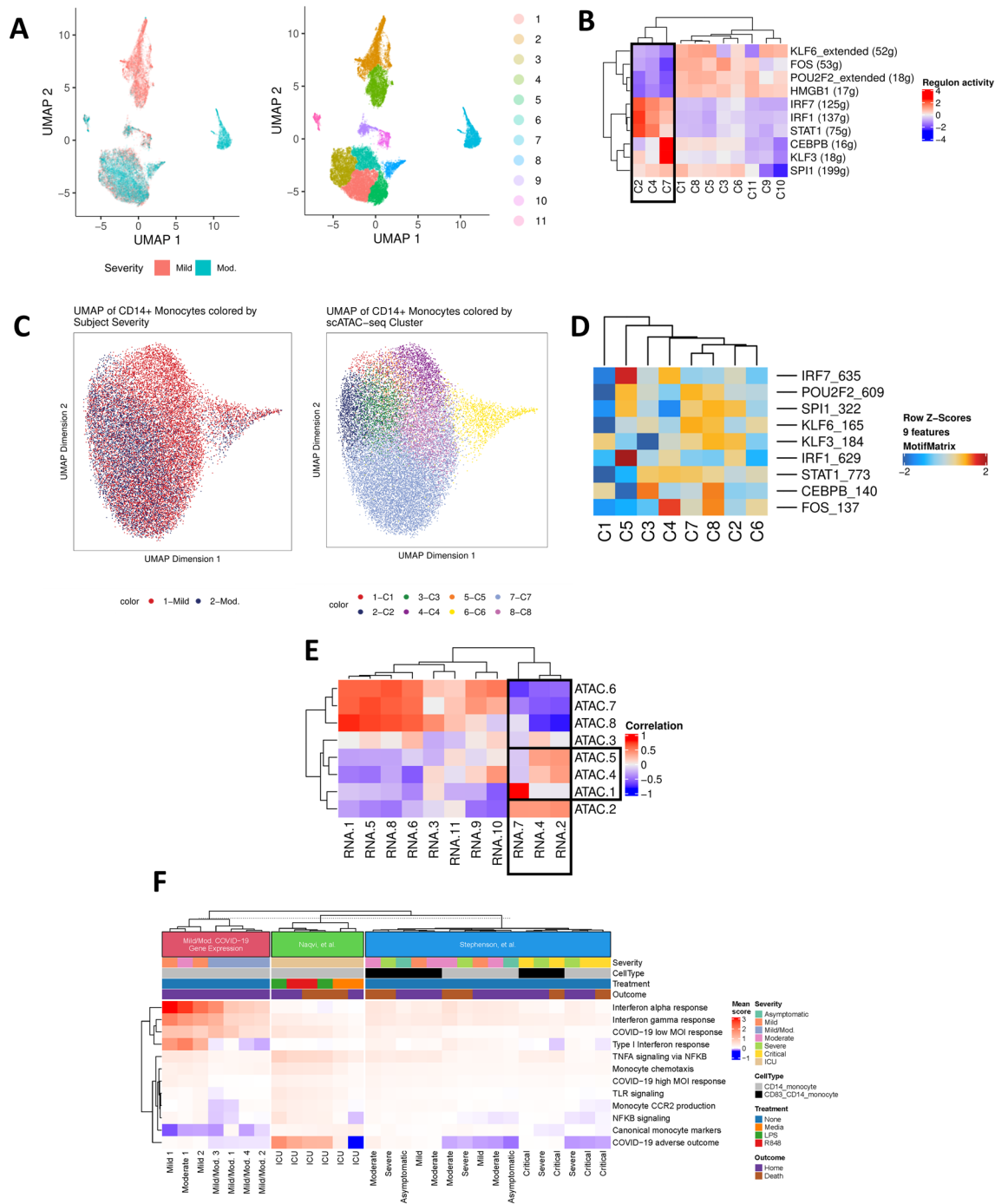
This study demonstrated that the epigenomes of critical subsets of PBMCs were remodeled extensively early in SARS-CoV-2 infection and reflected disease severity in IgG<sup>-</sup> subjects with mild and moderate symptoms. Specifically, differential activity of transcription factors and chromatin accessibility prior to anti-spike IgG seroconversion were observed to distinguish disease severity, preceding the later transcriptional response. The fact that these signals were robust even in a small cohort of subjects with subtle differences in disease severity (WHO OS 1 vs 2) suggests the potential power of epigenetic approaches to classify subsets of SARS-CoV-2 infection. This study complements previous work that identified signatures of accessible chromatin in severe or convalescing COVID-19<sup>21–24</sup>.

We found that transcription factor motifs enriched in peaks accessible prior to seroconversion had the highest occupancy in cells from the myeloid lineage, specifically CD14<sup>+</sup> monocytes, dendritic cells, and CD16<sup>+</sup> monocytes. KLF and CREB transcription factor families, which are known to regulate monocyte-macrophage polarization, had elevated motif accessibility in subjects with mild or moderate symptoms, suggesting an epigenetic priming mechanism controlling cell fate<sup>25,26</sup>. This distinction is important as not only are CD14<sup>+</sup> monocytes generally more activated in early COVID-19, as has been described, but here were differentially activated across subjects with variable disease severity. These differences included modulation of IFN- $\gamma$ , as reported previously. Specifically, mild symptom severity (or effective control) was marked by upregulation of classical antiviral pathways including those regulating IRF1 and IRF7. With increased severity, these antiviral signals diminished, suggesting that dysregulated and less effective responses underlied moderate disease. These early epigenetic changes occurred prior to transcriptional manifestations and in this cohort were also more stable and predictive of disease severity than gene expression alone. These findings suggest that epigenetic approaches focusing on chromatin accessibility may offer even greater promise than other host-based molecular diagnostics and predictive tools, as has been suggested for DNA methylation-based profiles<sup>27</sup>. Our work is consistent both with previous studies that identified DNA-methylation signatures, which included viral response and interferon signaling that will predicted SARS-CoV-2 infection and clinical outcome<sup>27</sup>, and with genome-wide DNA methylation signatures associated with severe COVID-19 that highlight hypermethylation in IFN-related genes, hypomethylation in inflammatory genes and increased epigenetic age<sup>28</sup>.

Interestingly, CD14<sup>+</sup> monocytes underwent the most extensive chromatin remodeling over time and exhibited epigenetic profiles at early seronegative times that distinguished disease severity. In contrast, after seroconversion, motifs enriched in accessible peaks were not seen in monocytes, instead showing the highest activity in B cells and plasmablasts, consistent with the transition from inflammatory signaling to adaptive immune development in this phase. Our work is consistent with epigenetic profiles of immune cells of individuals convalescing from COVID-19 that show establishment of immunological memory<sup>24</sup>.

There are limitations to this study, including the relatively small sample size, nonstandard enrollment and sampling times that vary across individuals, and a lack of subjects with critical illness to examine how these epigenetic changes manifest in more severe disease. However, development of a host-response assay that leverage the highly sensitive epigenetic biomarkers established early during infection has the potential to fill a clear unmet clinical need in the care of patients with COVID-19<sup>29,30</sup>. For example, in addition to the well characterized respiratory system damage of COVID-19, the deleterious effects on the central nervous system (CNS) can be devastating and include headache, anosmia (loss of smell), hyposmia (loss of taste), disturbance of smell, taste or vision, epileptic seizures, Guillain-Barré syndrome and intracerebral hemorrhage<sup>31</sup>. Our future studies will determine association between epigenetic biomarkers and specific symptoms that reflect damage to the CNS or peripheral nervous systems.





**Figure 5.** CD14+ monocytes undergo extensive chromatin remodeling prior to seroconversion and harbor severity-specific epigenetic biomarkers. (A) scRNA-seq UMAP of all CD14+ monocytes collected from mild and moderate IgG– subjects. UMAP colored by disease severity (left) and cluster number (right). (B) Heat map depiction of regulon activity computed for each scRNA-seq cluster using SCENIC. Activity of the top 10 regulons (right) is plotted for each cluster of cells from the UMAP (bottom). The black box indicates clusters of interest. (C) scATAC-seq UMAP of all CD14+ monocytes collected from IgG– mild and moderate subjects. UMAP colored by disease severity (left) and cluster number (right). (D) Heat map depiction of transcription factor motif enrichment of identified regulons (right) plotted for each scATAC-seq cluster (bottom). (E) Correlation plot using regulon activity to link clusters between scRNA-seq (columns) and scATAC-seq (rows). Black box indicates clusters of interest. (F) Enrichment of gene expression in monocyte-related pathways for CD14+ monocyte subsets from IgG– mild and moderate subjects. Pathway scores were computed for cell subsets from two published datasets for comparison. Hierarchical clustering was applied to samples within each dataset.

In summary, we found that the evolution of the chromatin landscape in circulating leukocytes during COVID-19 primes host immunological responses at early times, is mediated primarily by CD14<sup>+</sup> monocytes and correlates with an observed divergence in disease severity. These changes temporally precede transcriptional manifestations of pathways related to adaptive immunity. Together, these observations offer novel insight into severity-associated variations in host responses to SARS-CoV-2 infection and suggest that detection of critical components of chromatin remodeling in early disease may offer promise for a new class of diagnostic tools for COVID-19. The COVID-19 pandemic has highlighted with excruciating clarity the need for prognostic biomarkers to triage patients and personalize their treatment relevant to their likelihood to experience disease progression. Our findings offer a potential pathway via translation to diagnostic platforms capable of detecting markers of chromatin accessibility. Furthermore, improved understanding of the transcriptional and epigenetic response may reveal novel therapeutic opportunities. We envision that the rapid development of nanotechnology that supported mRNA vaccines can be harnessed to support detection of peripheral blood-based epigenetic biomarkers of early pre-seroconversion COVID-19<sup>32</sup>.

## Methods

**Cohort recruitment, biological sample collection, and initial phenotyping.** The study was approved by the Duke University Institutional Review Board. Protection of human subjects was in accordance with research protocols approved by the Duke University Institutional Review Board, consistent with the Declaration of Helsinki. Written informed consent was obtained from all research subjects or their legally authorized representatives. Subjects with confirmed or suspected SARS-CoV-2 infection or their close contacts were identified in the outpatient setting and enrolled into the Molecular and Epidemiological Study of Suspected Infection protocol (MESSI, IRB Pro00100241). All close contacts and subjects with mild or moderate COVID-19 were longitudinally sampled from enrollment to convalescent phase. Biological samples and demographic information were collected prospectively at first visit (Day 0) and at weekly intervals on Day 7 and Day 14. At each visit, infection with SARS-CoV-2 was confirmed using quantitative polymerase chain reaction (qPCR) of nasopharyngeal (NP) swab samples, and serology testing was performed for IgG against the SARS-CoV-2 spike domain. All subjects with mild or moderate COVID-19 progressed from seronegative (IgG<sup>-</sup>) to seropositive (IgG<sup>+</sup>). Close contacts were qPCR negative and IgG<sup>-</sup> at all time points; healthy controls were enrolled pre-pandemic and were not tested for SARS-CoV-2 or spike protein IgG. Self-reported symptom surveys were performed at each visit. To quantify symptom severity, the sum of 38 defined symptoms, each scored 0–4 (0–none, 1–mild, 2–moderate, 3–severe, 4–very severe), was determined from symptom onset through each longitudinal sample collection. SARS-CoV-2 q-PCR tests used virus RNA extracted from NP samples in 140  $\mu$ L of viral transport medium (VTM) using QIAamp Viral RNA Mini Kit (QIAGEN, Cat# 52904) according to manufacturer's instructions. SARS-CoV-2 nucleocapsid (N1) and human RNase P (RPP30) RNA copies were determined using 5  $\mu$ L of isolated RNA in the CDC-designed kit (CDC-006-00019, Revision: 03, Integrated DNA Technologies 2019-nCoV kit). SARS-CoV-2 IgG ELISA tests for antibody response were performed using the anti-SARS-CoV-2 spike S1 domain IgG ELISA assay (EUROIMMUN Medizinische Labordiagnostika AG, Lübeck, Germany).

**Purification of PBMCs.** Peripheral blood mononuclear cells (PBMCs) were purified using the Ficoll-Hypaque density gradient method according to manufacturer's instructions. Briefly, whole blood was collected in ACD Vacutainer tubes (Becton Dickinson, Cat# 364606) and processed within 8 h by dilution 1:2 in PBS, layered onto the Ficoll-Hypaque (Sigma Aldrich, Cat# GE17-1440-02) in 50 ml conical tubes, and centrifuged at 420  $\times$  g for 25 min. Buffy coat was collected, washed twice in D-PBS (Sigma Aldrich, Cat# D8537) by centrifugation at 400  $\times$  g for 10 min to isolate PBMCs which were assessed for viability and cell count using a Vi-Cell automated cell counter (Beckman-Coulter). PBMCs were adjusted to 10  $\times$  10<sup>6</sup> cells/ml in cryopreservation media (90% FBS, 10% DMSO), frozen at –80 °C using CoolCell LX (BioCision) for 12–24 h and stored in liquid nitrogen vapor phase.

**RNA extraction, total RNA-seq, and analysis.** RNA was extracted from 300 K cells using the Zymo Direct-zol RNA Miniprep Kit (Zymo Research, Cat# R2051) and RNA quality assessed using the Agilent DNA Screentape assay. The RNA Integrity Number (RIN) scores for all samples were >7.0. Total RNA libraries were generated using the NuGEN Ovation<sup>®</sup> SoLo RNA-Seq Library Preparation Kit (Tecan Life Sciences, Cat# 0500-96). Libraries were sequenced using Illumina NovaSeq 6000 instrument with S4 flow cell and 150 base pair paired-end reads (Illumina, Cat# 20012866). FASTQ files were generated from the NovaSeq BCL outputs and quality was assessed with FASTQC<sup>33</sup>. Differentially expressed genes were identified between subjects with different disease severity using the limma package and voom to model variance<sup>34</sup>.

**Nuclei purification, ATAC-seq and analysis.** Nuclei were extracted from frozen PBMCs. Briefly, 100 K cells were spun down at 300  $\times$  g for 5 min at 4 °C. The supernatant was removed, and cells were mixed with 100  $\mu$ L of lysis buffer (10 mM NaCl, 3 mM MgCl<sub>2</sub>, 10 mM Tris-HCl pH7.4, 0.1% Tween-20, 0.1% Nonidet<sup>™</sup> P40) and lysed on ice for 4 min. Wash buffer (1 mL; 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 10 mM Tris-HCl pH7.4, 0.1% Tween20) was added before centrifuging at 500  $\times$  g for 5 min at 4 °C. ATAC-seq libraries were generated as described<sup>14</sup>. Briefly, transposition mix (25  $\mu$ L 2  $\times$  TD buffer, 2.5  $\mu$ L transposase (Tn5, 100 nM final), 22.5  $\mu$ L water) (Illumina, Cat# 20031198) was added to the nuclear pellets, incubated at 37 °C for 30 min, and DNA purified using the QIAGEN MinElute PCR Purification Kit (QIAGEN, Cat#28004). DNA fragments were PCR amplified for a total of 10–11 cycles and resulting libraries purified using the QIAGEN MinElute PCR Purification Kit. The libraries were sequenced with an Illumina NovaSeq 6000 S4 flow cell using 100 bp paired-end reads (Illumina, Cat# 20027466). FASTQ files were generated from the NovaSeq BCL outputs and used as input to the ENCODE

ATAC-seq pipeline (<https://github.com/ENCODE-DCC/atac-seq-pipeline>) using the MACS2 peak-caller with all default parameters. Differential accessibility was calculated between groups of subjects with different disease severity using the csaw package<sup>35</sup>.

**Single-cell (sc)RNA-seq and analysis.** Frozen PBMCs were thawed, and count and cell viability were measured by Countess II. The cell viability exceeded 80% for all samples except PBMC samples from CC subjects, which had viability between 70 and 80%. For single cell (sc) RNA-seq, 200 K cells were aliquoted, spun down, resuspended in 30  $\mu$ L PBS + 0.04% BSA + 0.2U/ $\mu$ L RNase inhibitor, and counted using Countess II. GEM generation, post GEMRT cleanup, cDNA amplification, and library construction were performed following 10X Genomics Single Cell 5' v1 chemistry and quality was assessed using Agilent DNA Screentape assay. 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. Dimensionality reduction and cell type annotation was accomplished using gene-barcode matrices generated using Cell Ranger count were analyzed using Seurat 3 with the default parameters unless otherwise specified<sup>36</sup>. Regulatory network inference was accomplished by converting the scRNA-seq Seurat object into a SingleCellExperiment and used as input to analysis with the SCENIC package<sup>37</sup>.

**scATAC-seq and analysis.** PBMCs were thawed and nuclei were extracted as for ATAC-seq. The single-cell suspensions of scATAC-seq samples were converted to barcoded scATAC-seq libraries using the Chromium Single Cell 5' Library, Gel Bead and Multiplex Kit, and Chip Kit (10x Genomics). The Chromium Single Cell 5' v2 Reagent (10x Genomics, Cat# 120237) kit was used to prepare single-cell ATAC libraries according to the manufacturer's instructions. Quality was assessed using Agilent DNA Screentape assay. Libraries were then pooled and sequenced using Illumina NovaSeq 6000 platform with the goal of reaching saturation or 25,000 unique reads per nuclei on average. Sequencing data were used as input to the 10x Genomics Cell Ranger ATAC pipeline to demultiplex BCL files, generate FASTQs, and generate feature counts for each library. scRNA-seq and scATAC-seq were integrated using fragment file outputs generated using Cell Ranger ATAC count were analyzed using ArchR following the standard workflow and with default parameters unless otherwise specified<sup>38</sup>. Feature and motif enrichment analysis (peak calling) was performed using MACS2 via the addReproduciblePeakSet() method in ArchR which uses pseudo-bulk replicates of cells grouped on a specific design variable. The correlations between chromVAR transcription factor deviation scores and scRNA-seq derived gene expression data were calculated using the ArchR method correlateMatrices() to identify activators and repressors. Topic-based clustering was performed for the CD14+ monocytes from the mild or moderate subject cohorts using the R package cisTopic<sup>39</sup>.

Detailed methods for qPCR and antibody tests, and epigenetic and genomic profiling, including RNA extraction, sequencing, differential expression analysis and differential chromatin accessibility analysis for both bulk and single cell analytic approaches can be found in Supplemental Information.

## Data availability

The sequencing datasets and related clinical metadata tables are available via the NIH/NCBI Gene Expression Omnibus (GEO) repository using accession number GSE206284.

Received: 23 March 2022; Accepted: 28 June 2022

Published online: 09 July 2022

## References

- Blanco-Melo, D. *et al.* Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**, 1036–1045 e1039. <https://doi.org/10.1016/j.cell.2020.04.026> (2020).
- Helling, B. A. *et al.* Altered transcriptional and chromatin responses to rhinovirus in bronchial epithelial cells from adults with asthma. *Commun. Biol.* **3**, 678. <https://doi.org/10.1038/s42003-020-01411-4> (2020).
- Park, A. & Iwasaki, A. 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> (2020).
- Del Valle, D. M. *et al.* An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat. Med.* **26**, 1636–1643. <https://doi.org/10.1038/s41591-020-1051-9> (2020).
- Schulte-Schrepping, J. *et al.* Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell* <https://doi.org/10.1016/j.cell.2020.08.001> (2020).
- Zhou, Z. *et al.* Heightened innate immune responses in the respiratory tract of COVID-19 patients. *Cell Host Microbe* **27**, 883–890. e882. <https://doi.org/10.1016/j.chom.2020.04.017> (2020).
- Ma, S. *et al.* Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116 e1120. <https://doi.org/10.1016/j.cell.2020.09.056> (2020).
- Bournazos, S., Gupta, A. & Ravetch, J. V. The role of IgG Fc receptors in antibody-dependent enhancement. *Nat. Rev. Immunol.* **20**, 633–643. <https://doi.org/10.1038/s41577-020-00410-0> (2020).
- Ibarro, F. J. *et al.* Rapid decay of anti-SARS-CoV-2 antibodies in persons with mild Covid-19. *N. Engl. J. Med.* **383**, 1085–1087. <https://doi.org/10.1056/NEJMc2025179> (2020).
- Long, Q. X. *et al.* Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat. Med.* **26**, 845–848. <https://doi.org/10.1038/s41591-020-0897-1> (2020).
- Zhao, J. *et al.* Antibody responses to SARS-CoV-2 in patients with novel coronavirus disease 2019. *Clin. Infect. Dis.* **71**, 2027–2034. <https://doi.org/10.1093/cid/cia344> (2020).
- Amati, A. L. *et al.* Chemokines (CCL3, CCL4, and CCL5) inhibit ATP-induced release of IL-1beta by monocytic cells. *Mediators Inflamm.* **2017**, 1434872. <https://doi.org/10.1155/2017/1434872> (2017).
- Goodman, W. A. *et al.* KLF6 contributes to myeloid cell plasticity in the pathogenesis of intestinal inflammation. *Mucosal Immunol.* **9**, 1250–1262. <https://doi.org/10.1038/mi.2016.1> (2016).

14. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9. <https://doi.org/10.1002/0471142727.mb2129s109> (2015).
15. Aldrahe, S. *et al.* Dysregulation of macrophage development and phenotype in diabetic human macrophages can be rescued by Hoxa3 protein transduction. *PLoS ONE* **14**, e0223980. <https://doi.org/10.1371/journal.pone.0223980> (2019).
16. Naqvi, I. *et al.* DAMPs/PAMPs induce monocytic TLR activation and tolerance in COVID-19 patients; nucleic acid binding scavengers can counteract such TLR agonists. *Biomaterials* **283**, 121393. <https://doi.org/10.1016/j.biomaterials.2022.121393> (2022).
17. Stephenson, E. *et al.* Single-cell multi-omics analysis of the immune response in COVID-19. *Nat. Med.* **27**, 904–916. <https://doi.org/10.1038/s41591-021-01329-2> (2021).
18. Beutler, B. A. TLRs and innate immunity. *Blood* **113**, 1399–1407. <https://doi.org/10.1182/blood-2008-07-019307> (2009).
19. Brown, G. D. Dectin-1: A signalling non-TLR pattern-recognition receptor. *Nat. Rev. Immunol.* **6**, 33–43. <https://doi.org/10.1038/nri1745> (2006).
20. Niu, X. *et al.* Longitudinal analysis of T and B cell receptor repertoire transcripts reveal dynamic immune response in COVID-19 patients. *Front. Immunol.* **11**, 582010. <https://doi.org/10.3389/fimmu.2020.582010> (2020).
21. Covid-19 Multi-omics Blood Atlas (COMBAT) Consortium. A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell* **185**, 916–938 e958. <https://doi.org/10.1016/j.cell.2022.01.012> (2022).
22. Li, S. *et al.* Epigenetic landscapes of single-cell chromatin accessibility and transcriptomic immune profiles of T cells in COVID-19 patients. *Front. Immunol.* **12**, 625881. <https://doi.org/10.3389/fimmu.2021.625881> (2021).
23. Wilk, A. J. *et al.* Multi-omic profiling reveals widespread dysregulation of innate immunity and hematopoiesis in COVID-19. *J. Exp. Med.* **218**, e20210582. <https://doi.org/10.1084/jem.20210582> (2021).
24. You, M. *et al.* 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> (2021).
25. Kapoor, N. *et al.* Transcription factors STAT6 and KLF4 implement macrophage polarization via the dual catalytic powers of MCPiP. *J. Immunol.* **194**, 6011–6023. <https://doi.org/10.4049/jimmunol.1402797> (2015).
26. Sweet, D. R., Fan, L., Hsieh, P. N. & Jain, M. K. Kruppel-like factors in vascular inflammation: Mechanistic insights and therapeutic potential. *Front. Cardiovasc. Med.* **5**, 6. <https://doi.org/10.3389/fcvm.2018.00006> (2018).
27. Konigsberg, I. R. *et al.* Host methylation predicts SARS-CoV-2 infection and clinical outcome. *Commun. Med. (London)* **1**, 42. <https://doi.org/10.1038/s43856-021-00042-y> (2021).
28. Corley, M. J. *et al.* Genome-wide DNA methylation profiling of peripheral blood reveals an epigenetic signature associated with severe COVID-19. *J. Leukoc. Biol.* **110**, 21–26. <https://doi.org/10.1002/JLB.5HI0720-466R> (2021).
29. Altschul, D. J. *et al.* A novel severity score to predict inpatient mortality in COVID-19 patients. *Sci. Rep.* **10**, 16726. <https://doi.org/10.1038/s41598-020-73962-9> (2020).
30. Cote, A., Ternacle, J. & Pibarot, P. Early prediction of the risk of severe coronavirus disease 2019: A key step in therapeutic decision making. *EBioMedicine* **59**, 102948. <https://doi.org/10.1016/j.ebiom.2020.102948> (2020).
31. Singh, V., Allawadhi, P., Khurana, A., Banothu, A. K. & Bharani, K. K. Critical neurological features of COVID-19: Role of imaging methods and biosensors for effective diagnosis. *Sens. Int.* **2**, 100098. <https://doi.org/10.1016/j.sintl.2021.100098> (2021).
32. Khurana, A. *et al.* Role of nanotechnology behind the success of mRNA vaccines for COVID-19. *Nano Today* **38**, 101142. <https://doi.org/10.1016/j.nantod.2021.101142> (2021).
33. Andrews, S. FastQC: A quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2015).
34. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29. <https://doi.org/10.1186/gb-2014-15-2-r29> (2014).
35. Lun, A. T. & Smyth, G. K. csaw: A Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* **44**, e45. <https://doi.org/10.1093/nar/gkv1191> (2016).
36. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e1821. <https://doi.org/10.1016/j.cell.2019.05.031> (2019).
37. Aibar, S. *et al.* SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086. <https://doi.org/10.1038/nmeth.4463> (2017).
38. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411. <https://doi.org/10.1038/s41588-021-00790-6> (2021).
39. Bravo Gonzalez-Blas, C. *et al.* cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* **16**, 397–400. <https://doi.org/10.1038/s41592-019-0367-1> (2019).

## Acknowledgements

We thank Eric van Gieson and Thomas Thomou for their advice throughout the project, and Andrew Macintyner, Thomas Oguin, and their teams at the Duke Regional Biocontainment Laboratory for COVID IgG serology and PCR viral load testing. We thank Raul Louzao, Christina Nix, Carolyne Whiting, Tyffany Locklear, Pam Isner, Anna Mazur, Jack Anderson, Maria Miggs, Aleah Bowie, Julie Steinbrink, Robert Rolfe, and Jorge Prado Balcazar for technical expertise. We wish to express our deep appreciation to the participants and their families for their generosity in collaborating with us in this work.

## Author contributions

N.S.G., M.T.M., T.W.B., G.S.G., X.S., C.W.W. designed the study; N.S.G., S.D., M.T.M., T.W.B., E.P., H.A.C., G.O.R., R.X., S.B., T.R., T.C., B.P.N., E.R.K. collected samples and generated data; N.S.G., E.W., R.H., G.D.S., T.N.D., B.D.K., E.L.T., L.L.S. analyzed data; N.S.G., T.W.B., L.L.S., M.I.D.U., C.W.W. wrote the manuscript.

## Funding

This work was supported by the U.S. Defense Advanced Projects Agency (DARPA, N66001-09-C-2082 and HR0011-17-2-0069) to XS and CWW, the Veterans Affairs Health System and Virology Quality Assurance (VQA) 75N93019C00015 to CWW, NIH/NHLBI K08HL130557 to BDK, DARPA/USMRAA W911NF1920111 to XS, NIH/NIGMS R35GM122465 to XS, and NIH/NIAID UC6AI058607 to the Duke Biocontainment Laboratory. This work used a high-performance computing facility supported by grants 2016-IDG-1013 and 2020-IIG-2109 from the North Carolina Biotechnology Center.

## Competing interests

MTM reports grants on biomarker diagnostics from the Defense Advanced Research Projects Agency (DARPA), National Institutes of Health (NIH), Sanofi, and the Department of Veterans Affairs. TWB reports grants from

DARPA and is a consultant for and holds equity in Predigen and Biomeme. MTM, TWB, ELT, GSG, and CWW report patents pending on Molecular Methods to Diagnose and Treat Respiratory Infections. ELT reports grants on biomarker diagnostics from DARPA, the NIH/Antibacterial Resistance Leadership Group (ARLG); an ownership stake in Predigen and Biomeme; consultancy fees from bioMerieux; GSG reports an ownership stake in Predigen and Biomeme; CWW reports grants on biomarker diagnostics from DARPA, NIH/ARLG, Predigen, and Sanofi; and has received consultancy fees from bioMerieux, Roche, Biofire, Giner, and Biomeme. All other authors NSG, SD, EP, HAC, GOR, EW, RX, SB, TR, BPN, TC, RH, GDS, TND, MIDU, LLS, ERK, BDK and XS have no competing interests to report.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-15668-8>.

**Correspondence** and requests for materials should be addressed to C.W.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022, corrected publication 2023