Dissecting the epigenome dynamics in human immune cells upon viral and chemical exposure by multimodal single-cell profiling

Irem B. Gündüz[1]*, Bei Wei[2]*, Derek C. Chen[3], Wenliang Wang[4], Manoj Hariharan[4], Todd Norell[5], Timothy J Broderick[5], Micah T. McClain[6,7], Lisa L. Satterwhite[8], Thomas W. Burke[6], Elizabeth A. Petzold[6], Xiling Shen[9], Christopher W. Woods[6,7], Vance G. Fowler, Jr.[6,10], Felicia Ruffin[6], Parinya Panuwet[11], Dana B. Barr[11], Aaron J Wilk[12,13,14], Madeline J Lee[13,14], Catherine Blish[12,14,15], Flora Castellino[16], Anna Maria Walley[17], Thomas Evans[17], Joseph R. Ecker [4,18], Fabian Müller[1]^, William J. Greenleaf [3]^

1. Department of Integrative Cellular Biology and Bioinformatics, Saarland University, 66123 Saarbrücken, Germany
2. Medical Research Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, 510120, PR China
3. Department of Genetics, Stanford University, Stanford, CA 94305, USA
4. Genomic Analysis Laboratory, The Salk Institute for Biological Studies, 10010 N Torrey Pines Rd, La Jolla, CA 92037, USA
5. Healthspan, Resilience, and Performance, Florida Institute for Human and Machine Cognition, 40 S Alcaniz St, Pensacola, FL 32502, USA
6. Center for Infectious Disease Diagnostics and Innovation, Division of Infectious Diseases, Duke University Medical Center, Durham, NC 27710, US
7. Durham Veterans Affairs Medical Center, Durham, NC 27705 USA
8. Department of Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham, NC 27708, USA
9. Terasaki Institute for Biomedical Innovation, Los Angeles, CA 90024, USA
10. Duke Clinical Research Institute, Durham, NC 27701 USA
11. Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA 30322 USA
12. Stanford Medical Scientist Training Program, Stanford University School of Medicine, Stanford, CA

13. Stanford Immunology Program, Stanford University School ofMedicine, Stanford, CA
14. Department of Medicine, Stanford University School of Medicine, Stanford, CA
15. Chan Zuckerberg Biohub, San Francisco, CA
16. U.S. Department of Health and Human Services, Administration for Strategic Preparedness and Response, Biomedical Advanced Research and Development Authority, Washington, DC, USA.
17. Vaccitech plc, Unit 6-10, Zeus Building, Rutherford Avenue, Harwell OX11 0DF, United Kingdom
18. Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 N Torrey Pines Rd, La Jolla, CA 92037, USA

* Equal contribution (co-first)

^Co-corresponding authors: fabian.mueller@uni-saarland.de,wjg@stanford.edu

Abstract

Pathogen and chemical exposures lead to profound remodeling of the gene-regulatory landscape across human immune cell populations. Here, we present a single-nucleus chromatin accessibility atlas of human immune cells of individuals exposed to HIV-1, COVID-19, Influenza virus, organophosphates as well as healthy controls that provides insights into gene regulation driven by these cell-extrinsic stimuli. This atlas comprises 271,299 cells and 319,420 candidate regulatory elements that exhibit dynamic accessibility associated with gene expression across immune cell states. Our longitudinal HIV cohort reveals epigenetic signatures of T cell exhaustion, manifested in changes in the accessibility of binding sites for the FOXP family transcription factors. We further identified changes in the accessibility of candidate regulatory elements in CD14 monocytes upon SARS-CoV-2 exposure that are associated with a switch in NF-κB to AP-1-based regulation of cytokine networks. By integrating single-cell profiles of DNA methylation from matched samples we created a multimodal epigenome atlas of human immune cells across exposure states using the accessibility-derived candidate regulatory elements. Both modalities exhibit complementary epigenetic signatures at transcription factor binding sites associated with cell state, as exemplified in the process of memory formation in T-cells. Finally, by linking potentially regulatory DNA methylation signatures to changes in

chromatin accessibility in monocytes, we identify AP1 motifs exhibiting epigenetic dynamics, indicating selective remodeling in TF networks in severe cases of COVID-19.

Keywords

Single-cell epigenomics, pathogen exposure, human immune cells, infection, chromatin accessibility, DNA methylation, gene regulation

Introduction

Recent advances in sequencing and single-cell technologies have allowed the detailed molecular characterization of exposure-associated changes in the gene expression of human immune cell populations, revealing potential markers for diagnosis and therapeutic interventions [1,2]. For example, single-cell RNA sequencing profiling of peripheral blood mononuclear cells (PBMCs) from patients exposed to Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) highlighted downregulated expression of pro-inflammatory cytokines in peripheral monocytes associated with the severity phenotype [3]. In contrast, other viral exposures, such as Influenza virus, lead to upregulated cytokine expression in peripheral monocytes [2]. The expression patterns of cytokine receptor genes and transcription factors also differ in Influenza virus exposure from SARS-CoV-2 exposure [4]. Mild cases of COVID-19 exhibit a more limited activation of NK and monocyte cells compared to mild influenza cases. [5]. Exposure to the Human Immunodeficiency Virus type 1 (HIV-1) induces changes in cytokine and chemokine expression, similar to SARS-CoV-2 exposure, leading to shifts in cell composition and increased signatures of inflammation [6]. However, while similarities and differences in molecular responses between COVID-19 and other viral exposures are well studied at the transcriptomic level, the underlying gene regulatory programs that contribute to these responses remain less well characterized.

Exposure to chemicals, such as organophosphate pesticides (OP), can lead to toxic effects similar to those observed with exposure to infectious diseases [7]. For example, low-dose chronic OP exposure is associated with increased expression of pro-inflammatory cytokines, such as IL-1β [8]. Population-level reports link chronic OP exposure to differential methylation levels in cellular defense-associated genes, such as the HLA family [9]. However, the gene regulatory programs of human immune cells in response to OP exposure remain to be characterized at high resolution.

Single-nucleus epigenomic profiling-based analysis using the assay for transposase-accessible chromatin with sequencing (ATAC-Seq) [10] has been used to characterize gene regulatory programs of human peripheral blood cells in healthy individuals and across various disease states [5, 11, 12, 13]. A small number of studies with small cohort sizes have reported *in vivo* accessibility datasets of human immune cells upon exposure to infectious agents [14, 15]. However, a more comprehensive and integrative analysis of gene regulation across exposures to infectious agents as well as environmental chemicals could provide a more holistic view of common and distinct regulatory programs the immune system uses to respond to environmental stimuli.

To chart the regulatory changes conferred by dynamic chromatin accessibility, we performed snATAC-seq [10] on 92 samples collected from individuals exposed to SARS-CoV-2, HIV-1 virus, Influenza A virus (IAV), and organophosphate pesticide (OP) exposures, along with healthy donors, and generated chromatin accessibility profiles of 271,299 peripheral blood cells. We analyzed how these exposure-specific alterations affect the activity of *cis*-regulatory elements (cCREs) in each exposure state and cell type. Individuals exposed to SARS-CoV-2 exhibited significantly altered chromatin accessibility profiles in CD14+ monocytes, linked to a potential shift in TF-based regulation of cytokine signaling networks, such as the downregulation

of the NF-kB motif family. Moreover, we identified chromatin signatures of T cell exhaustion in prolonged HIV-1 infection, characterized by changes in accessibility of regulatory networks involving FOXP transcription factors. By integrating single-cell methylation profiles from matching cohorts [16,17], we further characterized the interplay between DNA methylation and chromatin accessibility in potentially regulatory DNA regions and found anticorrelated patterns of chromatin accessibility and DNA methylation in these regions that indicate coordinated epigenetic changes across cell populations and epigenomic modalities. Our analysis revealed TF signatures such as BATF, AP1, and ETS TF motif families, exhibiting joined dynamics in accessibility and DNA methylation in memory T-cells compared to naive states, suggesting active regulatory roles in T-cell memory formation. Finally, we identified changes in DNA methylation that are concordant with accessibility changes observed upon SARS-CoV-2 exposure in CD14+ monocytes, particularly in regulatory networks involving Interferon regulatory factors (IRFs) and AP-1 factors.

Results

A single-cell epigenome atlas of human immune cells across viral and chemical exposures

We assayed the single-cell chromatin accessibility landscape of human immune cells of healthy individuals and individuals with chemical and viral exposures. We profiled 92 human PBMC samples across 4 different cohorts and exposure groups (Figure 1A; Supplementary Table 1): (i) a SARS-Cov2 cohort consisting of healthy control samples (n=7) and individuals with mild (WHO score 1–3; n=8), moderate (WHO score 4-5; n=4) and severe (WHO score 6-8; n=6) COVID-19 indications; (ii) a longitudinal cohort of four HIV infected individuals in pre-infection (n=4), acute infection (n=4) and chronic infection (n=4), in which antiretroviral therapy (ART) was administered between the acute and chronic infection stages; (iii) an influenza cohort of 5

individuals sampled 2 days prior to live influenza virus challenge (control group) as well as 3 days, 6 days and 28 days after virus challenge; and (iv) a cohort of individuals with varying degrees of exposure to organophosphates (OP), grouped by urinary TCPY (3,5,6-trichloropyridinol) levels: low exposure (TCPY < level of detection; n=9), medium exposure (0 < TCPY <= 2.4 µg/L; n=6) and high exposure (TCPY > 2.4 µg/L; n=20).

We performed a 10X-based single-nucleus assay of transposases accessible chromatin sequencing (snATAC-seq) and obtained chromatin accessibility profiles of 285,276 cells, of which 271,299 cells were retained after quality control and filtering (Supplementary Figure 1A; Methods). We integrated our four cohorts to characterize immune cell populations across pathogen exposures while correcting for batch effects using Harmony [18]. We identified 22 clusters using graph-based clustering (see Methods, Supplementary Figure 1B) and 319,420 reproducible peaks that represent candidate *cis*-regulatory elements (cCREs) across the identified clusters (Supplementary Figure 1C).

To annotate these clusters with corresponding cell types, we integrated and mapped cell type annotations from a publicly available single-cell transcriptome atlas of cross-tissue immune cells [19] (Supplementary Figure 1D). We grouped our 22 clusters into 12 broad immune cell populations, including NK cells, $CD14^+$ classical and $CD16^+$ non-classical monocytes, dendritic cells (DC), as well as naive and mature subpopulations of B cells, plasma cells, and T cells (Figure 1B). To validate these annotations, we examined the chromatin-derived gene activity and transcription factor (TF) activity, which serve as a proxy for gene expression and regulatory potential, of canonical cell type markers (Figure 1C, D, and Supplementary Figure 1E, F) [11]. Gene loci of corresponding marker genes exhibited characteristic chromatin accessibility profiles around promoter and cCRE regions in their respective cell types (Figure 1D and Supplementary Figure 1E), e.g., IRF8 exhibited increased activity in B cells and monocytes,

while the GATA3 locus and its target binding sites were mostly accessible in T and NK cell populations.

To better understand the role of transcription factors in the regulation of cell state, we aggregated accessibility across the genomic occurrences of a consolidated set of clustered TF binding motifs (Figure 1E) [20,21]. Comparing the derived TF motif activities of cells with each other (Supplementary Table 2), we observed increased NF-kB activity in $CD14^+$ classical and $CD16^+$ non-classical monocytes, dendritic cells (DC), B naive, B memory, and plasma cells compared to T and NK cell populations. We also observed increased chromatin activity at candidate binding sites of transcription factors known to regulate cellular differentiation and function in T cells (RUNX and ETS) and B cells (PAX) compared with other cell types.

We assessed the impact of each exposure condition on cell-type composition by comparing it with the respective control groups (Figure 1F, Supplementary Figure 1G). We observed a relative expansion of the $CD8^+$ T cell population (padj=0.029) and a consistent but statistically insignificant depletion of $CD4^+$ T cells in individuals chronically exposed to HIV (Supplementary Figure 1F). In individuals with severe COVID-19 conditions, we observe a proportionally larger population of plasma cells (padj=0.016) and depletion of NK cells (padj=0.035). Other cell types did not exhibit significant changes in proportions in severe cases of COVID-19, consistent with reports by Wilk 2021 [5]. Furthermore, we observed a depletion of naive T-cells (padj=0.04) and DC cells (padj=0.022) in mild cases of COVID-19 compared to the control group. Although the cell-type compositions do not change significantly during OP and influenza exposures, we observed nominal composition changes. For example, on the 6th day after Influenza exposure, we observed a depletion in T-naive cell composition and an increase in classical CD14+ and non-classical CD16+ monocytes compared to controls (Figure 1F, padj = 0.056, 0.15, respectively).

HIV-induced T-cell exhaustion dynamics in different stages of the infection

To further characterize the expansion of the $CD8^+$ T cell population during HIV infection, we sub-clustered $CD8^+$ T cells into different subtypes. Of note, we identified a population of exhausted T cells (Tex), marked by higher gene activity scores (defined as aggregated locus accessibility, see Methods) of exhaustion markers such as CTLA4, NCAM1, and HAVCR2 (Figure 2A-C, Supplementary Table 3) [22]. We also found other gene activity scores elevated in this subset, such as ROBO1/2, which are not traditional exhaustion markers, suggesting specific roles of these genes in HIV-infection-related exhausted T cells [23,24].

When we assessed the proportional changes in T-cell populations within the HIV cohort, we observed a depletion of $CD4^+$ T cells. In contrast, the proportion of exhausted $CD8^+$ T cells increased after the HIV exposure ("acute" stage, Figure 2D). Changes in cell abundance for $CD4^+$ T cells and exhausted CD8+ T cells are anti-correlated throughout the progression of infection. For example, subject 1 exhibited the most severe depletion of $CD4^+$ T cells along with an increase of exhausted $CD8^+$ T cells during the infection trajectory; for subjects 1 and 2, the proportions of $CD4^+$ T cells decreased in the "acute" stage and recovered in "chronic" stage after antiretroviral therapy (ART), consistent with a response in viral load to ART and regeneration of the $CD4^+$ compartment (Figure 2E). Conversely, the proportions of exhausted $CD8^+$ T cells decreased in the "chronic" stage compared to the "acute" stage; For subjects 3 and 4, the proportions of $CD4^+$ T cells continued to decrease in the "chronic" stage after ART, while the proportions of exhausted $CD8^+$ T cells also increased, highlighting that cell type compositions relevant to disease progression and therapeutic response can be inferred from epigenetic signatures.

Next, we identified differentially accessible regions in exhausted T-cells (Tex) compared to other $CD8^+$ T-cell populations in the HIV cohort (Figure 2F, Supplementary Table 4). We then determined the activity changes of TF motifs based on the enrichment of corresponding motifs in hypo- and hyper-accessible cCREs. In regions with increased accessibility in exhausted cells, we observed the enrichment of FOXP2 and other FOX family motifs, while elements with decreased accessibility were enriched in SP and SMAD motifs (Figure 2G). We then examined the TF motifs present around accessible elements near marker genes for T cell exhaustion; for example, for gene CTLA4, we found that FOXP motifs are located mostly within cCREs with increased accessibility in Tex compared with other subtypes/subclusters $CD8^+$ T cells (Figure 2H), consistent with the hypothesis that CTLA4 was activated by FOXP family binding in Tex.

SARS-CoV-2 Reshapes Chromatin and Cytokine Regulation in $CD14^+$ Monocytes

To assess the impact of each exposure on cell-type-specific accessibility profiles, we aggregated the accessibility of each cell type for donors and exposure conditions into pseudo-bulk samples. We performed pairwise differential accessibility analysis between control and disease groups for each cohort and cell type across cCREs. Our analysis revealed limited differentially accessible regions (DARs) upon influenza and organophosphate exposures (Supplementary Figure 2A). In contrast, we observe substantial differences in accessibility associated with COVID-19, particularly in severe cases within the population of $CD14^+$ classical monocytes (Figure 3A, Supplementary Figure 3A).

We identified 923 differentially accessible regions (DARs) in CD14+ classical monocyte profiles in severe COVID-19 and 373 DARs in moderate conditions (Figure 3A). We found that in the union of these DARS, accessibility fold-changes positively correlate between moderate and severe conditions, indicating that these underlying differential regions are consistent across these levels of disease severity, differing primarily in the magnitude of the observed accessibility

changes (Figure 3B, C). Next, we performed TF motif enrichment analysis in DARs, revealing that hypoaccessible regions in severe cases of COVID-19 were enriched for NF-kB and SPI motif clusters. In contrast, hyperaccessible regions were enriched for AP1 and CREB factors (Figure 3D). Both NF-kB and AP1 are key regulators of cytokine signaling; however, while NF-kB is predominantly associated with acute inflammatory responses, AP1 factors are linked to late-stage inflammation and stress responses. A shift from NF-kB to AP1-dominated regulation may reflect an underlying immune paralysis, consistent with previous reports (Wilk 2021, Wimmers 2023, and Lam 2023 [25,26].

To evaluate how chromatin accessibility changes relate to gene expression changes, we re-analyzed publicly available, matched single-cell transcriptome data from 34,548 cells obtained from the same COVID-19 cohort (see Methods) [3]. We compared differential gene expression in $CD14^+$ classical monocytes between severe COVID-19 and control groups from scRNA-seq and gene activity derived from chromatin accessibility (Figure 3E, Supplementary Table 5, see Methods). Consistent with the findings of Wimmers 2023 [26], we observed an increased expression and gene activity in interferon-stimulated genes (ISGs), including the IFIT family (Figure 3E, Supplementary Figure 2B). In contrast, decreased expression and gene activity were observed in primarily inflammatory genes, including IL-1B. We also identified genes near differential peaks that align with the findings of Wilk 2021, such as IL6 and CD4, where they loose accessibility. We observed decreased accessibility in CD4, a key regulator of macrophage differentiation and pro-inflammatory cytokine induction, under severe conditions [5,27].

We observed decreased accessibility in HLA Class II genes, expression of which is lost by myeloid-derived suppressor cells (MDSC), which suggests that CD14+ monocytes in severe COVID-19 exhibit MDSC-like molecular features (Supplementary Figure 3B) [5,28]. Furthermore, our analysis revealed that genes associated with gaining accessibility are enriched in monocyte apoptosis and the negative regulation of type I interferon-mediated signaling, suggesting a potential shift toward immune suppressive and pro-apoptotic programs

(Supplementary Figure 3C). Conversely, genes with decreased accessibility are enriched with positive regulation of monocyte differentiation, indicating repression of the normal differentiation process. Together, these findings reveal a chromatin reprogramming pattern in severe disease, implicating decreased accessibility at antigen presentation and differentiation genes, and increased accessibility at immune suppressive and pro-apoptotic loci, as drivers of altered monocyte function.

Coordinated Chromatin and DNA Methylation Signatures of T Cell Memory Formation

We integrated our single-cell chromatin accessibility atlas with single-cell DNA methylation data that were separately generated for a subset of 39 samples from our organophosphate and HIV cohorts using snmC-seq2 [16]. These datasets comprise single-cell methylomes from 7 major immune cell types obtained by fluorescence-activated cell sorting (FACS) (see Methods). We aggregated DNA methylation signals across the cCREs identified by scATAC-seq. After preprocessing and quality control (Methods), we retained 22,488 cells with DNA methylation profiles, which we integrated with corresponding chromatin accessibility profiles of 110,376 cells from matching samples using canonical correlation analysis (Methods) (Figure 4A). This integration on the level of cCREs allowed the cross-projection between low-dimensional representations of chromatin accessibility and DNA methylation and the matching of annotated cell types. We observe high cell-type correspondence between the annotated chromatin accessibility and DNA methylation profiles on the level of broadly defined cell types (B-cells, T cells, monocytes, NK cells) while noting weaker correspondence on the level of T-cell subpopulations (Figure 4A, B).

TF binding exhibits varying degrees of sensitivity to DNA methylation in TF binding sites. DNA methylation is known to promote TF binding in some motifs and prevend binding in others [29]. To determine the relationship between TF-associated chromatin accessibility and DNA

methylation, we derived TF regulator activity by aggregating chromatin accessibility [21] and DNA methylation profiles across putative binding sites of TFs in pseudobulk samples for each individual, disease condition, and cell type (see Methods and Supplementary Figure 4A). We observed that TF activities derived from both epigenomic modalities can discriminate broad cell type clusters, with methylation-based TF activities providing fine-grained differentiation between naive and memory T-cell clusters (Supplementary Figure 4B).

We further expanded the moderate inverse relationship between accessibility and methylation-derived TF activities by analyzing it in a cell-type-specific manner. The inverse correlation was most pronounced in B-cells (cor.=-0.48), followed by Th-Naive (cor.=-0.23), Monocyte (cor.=-0.18), Tc-Naive (cor.=-0.17), NK-cell, Th-Mem (cor.=-0.01), and Tc-Mem (cor.=0.01), respectively (Supplementary Figure 4D, Supplementary Table 6). For instance, the SPIB motif displays high accessibility with low methylation in monocytes, whereas in T-cells, it exhibits lower accessibility accompanied by higher methylation. A similar inverse relationship between accessibility and methylation was observed for NF-κB family members, including REL, RELA, and RELB (Figure 4C, Supplementary Figure 4D).

Next, we identified TF motifs exhibiting the highest variability in chromatin accessibility-based TF activities across cell types. We observed a moderate inverse relationship between accessibility and methylation-based TF activity scores, where TFs exhibiting higher accessibility tended to show lower methylation and vice versa (Figure 4C,D). To further assess the methylation sensitivity of these motifs, we examined their annotated methylation sensitivity [30,31]; 397 of JASPAR motifs correspond to TFs annotated with methylation sensitivity annotation from methyl-SELEX experiments, with certain TFs on which DNA methylation had a negative impact on binding (methylMinus), while in other binding sites DNA methylation promotes TF binding (methylPlus). We observed a broad range of correlation scores between accessibility and methylation-associated TF motif activity for methylMinus and methylPlus motifs, with methylMinus motifs exhibiting a skew towards negative correlation coefficients between methylation and accessibility (Supplementary Figure 4C).

Consistent with reports on the global loss of DNA methylation in T cell memory formation [33], we observe lower methylation levels at TF motifs in memory T cells compared to naive cell states (Figure 4C, D). We observed that naive and memory T-cells exhibit dynamic methylation activity patterns across TF motifs, accompanied by relatively stable accessibility patterns. To characterize the interplay between methylation and accessibility in TF-driven regulation of T-cell memory formation, we compared motif activities between naive and memory T-cells across both modalities (see Methods). We observed an anti-correlation between the differentially active motifs identified from the two epigenomic modalities in $CD4^+$T and $CD8^+$T cell subsets, suggesting that these epigenetic layers coordinate TF activity to drive memory formation in T-cells (Figure 4D and Supplementary Figure 5A).

Next, we assessed the chromatin "footprints" – characteristic high accessibility at TFBS compared to the genomic background and protective patterns likely due to TF binding – of TF motifs implicated with roles in T-cell memory formation. We compared these chromatin footprints to DNA methylation footprints (Figure 4E, Supplementary Figure 5B). Our findings indicate that BATF, FOSL2::JUND, and FOS motifs exhibit increased accessibility and decreased methylation levels in memory T-cells compared with naive cells, consistent with the accessibility reports of Calderon 2019 and methylation reports of Durek 2016 [32,33]. Furthermore, CEBPE, RELA, IRF4, and RUNX3 exhibit a nominal increase in accessibility while exhibiting markedly decreased methylation in T-memory cells compared to naive T-cells. Consistent with observations from Abdelsamed 2020, who showed decreased methylation levels in the EOMES TFBS in CD8 memory T-cells [34], we also observe a decrease in methylation in T-memory cells compared to T-naive cells across EOMES binding sites, coupled with increasing accessibility to $CD8^+$ T-memory cells compared to naive T-cells (Supplementary Figure 5B). MAF::NFE2 motifs show increased accessibility in memory T-cells compared to naive T-cells, with decreased in memory-specific methylation observed (Supplementary Figure 5B). FOXP1 motifs, a known marker of memory formation in T-cells, exhibit slightly decreased

methylation in memory T-cells, with nominally decreased accessibility patterns in quiescent memory T-cells compared to naive T-cells [33]. The binding motifs of ETS1, a member of the ETS factor family that regulates T-cell differentiation and function, exhibit increased accessibility and decreased methylation in memory T-cells, compared to naive T-cells [35]. Our findings indicate that memory-associated TFs are regulated by coordinated epigenetic changes at their binding sites, highlighting a layered regulatory program underlying memory formation in T-cells.

SARS-CoV-2 Induces NF-κB and AP-1-Linked Epigenetic Remodeling in CD14^{+} Monocytes

We next aimed to identify DNA methylation signals in cCREs with exposure-specific chromatin accessibility profiles and to relate exposure-related chromatin accessibility changes to DNA methylation changes. We aggregated DNA methylation signals in cCREs defined by scATAC peak regions into pseudo-bulk samples per donor and exposure condition for each cell type and identified differential DNA methylation in cCREs by comparing healthy control and COVID-19 pseudo-bulk samples. Our results reveal that CD14+ classical monocytes exhibit the most pronounced differentially methylated regions (DMRs) in severe cases of COVID-19 compared to other cell types (Figure 5A, Supplementary Figure 6A). In contrast, mild cases exhibit fewer DMRs, and other cell types show limited changes upon SARS-CoV-2 exposure. We observed 2073 DMRs in monocyte profiles in severe COVID-19 cases compared to healthy controls. We found a weak negative correlation (correlation coefficient = -0.07) between mean methylation difference and accessibility fold-change, indicating concordant epigenetic alterations across these modalities (Figure 5B).

Next, we performed TF enrichment analysis in the accessibility-defined DMRs hypomethylated and hypermethylated in severe cases compared to controls, revealing enrichment of IRF1 motifs in hypermethylated DMRs consistent with reported expression changes from Wilk 2021 (Figure 5C) [5,36] and chromatin accessibility changes in these regions (Figure 3D). We

observed the enrichment of IRF factors in hypermethylated DMRs consistent with methylation-based TF enrichment findings of Godoy□Tena 2022 [36]. We also found enrichment of CREB/ATF motifs in hypomethylated DMRs, which have been shown to inhibit the NF-κB-driven pro-inflammatory cytokine expression [37,38]. Consistent with the findings of Godoy□Tena 2022, we found enrichment of CEBP motifs in hypomethylated DMRs [36]. Moreover, consistent with the findings of Wimmers 2023 showing altered chromatin accessibility, we found enrichment of SPI motifs in hypermethylated DMRs [26].

To further characterize how the binding of these TFs could be altered through the epigenome upon SARS-CoV-2 exposure, we focused on differentially active TF motifs (pad.j < 0.05) from accessibility and their corresponding methylation TF activity (Figure 5D). We found that SPI factors lose accessibility in patients with severe COVID-19 compared to the control group, accompanied by increased methylation. Moreover, AP1 factors, such as FOSL2::JUN and JUN, gain accessibility accompanied by decreased methylation upon severe conditions. In contrast, while NF-κB factors lose accessibility, we found that RELA and REL also lose methylation, whereas NF-κB1 and NF-κB2 gain methylation under severe conditions (Figure 5D).

Finally, we analyzed the motif footprints of differentially active TF motifs to validate our TF activity findings (Figure 5E, Supplementary Figure 6B). Supporting our activity findings, SPI motifs—including SPIB and SPIC—exhibited reduced accessibility alongside a modest increase in methylation. For AP-1 factors such as FOSL2::JUN, JUN, and FOSB::JUNB, we observed no clear differences in accessibility or methylation compared to the control group, although slight trends toward increased accessibility and decreased methylation were noted. In contrast, we observed nominally decreased accessibility in RELA and REL, NF-kB factors, in the severe group, accompanied by unchanged methylation levels. For CREB and CEBP factors, such as CREB1, CEBPD, and CEBPA, we observed nominally increased accessibility and a nominal decrease in methylation compared to the control group (Figure 5E, Supplementary Figure 6B). Besides the differentially active motifs, we also assessed BATF motifs, which belong to the AP1

factors, since we observed them enriched in hypermethylated DMRs (Figure 5C). We observed a nominal decrease in accessibility in BATF::JUN and BATF, accompanied by increased methylation in severe cases for both motifs (Supplementary Figure 6B).

Discussion

We profiled a comprehensive single-cell atlas of chromatin accessibility in human immune cells across various environmental stimuli, including viral and chemical exposures, charting the regulatory responses of different immune cell types to different environmental stimuli. By high-resolution mapping of chromatin accessibility of candidate *cis*-regulatory elements (cCREs), we identified shared and stimulus-specific activity patterns that shape immune cell function in response to external perturbations. Furthermore, by integrating exposure-associated changes in chromatin accessibility with matched single-cell DNA methylation profiles, we defined the epigenetic patterns associated with transcription factors (TFs) involved in defining general cell types and during immune memory formation. We further characterized altered signatures of multimodal epigenetic activity of TFBS associated with SARS-CoV-2 exposure, including NF-kB and AP1 factors. These findings highlight the utility of our single immune cell epigenome atlas for identifying regulatory mechanisms modulated by TF activity across cell states and in response to stimuli.

We identified an association between FOXP gene activity and an associated enrichment of its potential binding motifs and the upregulation of exhaustion markers, particularly CTLA4, within a subset of CD8+ T cells from HIV-infected individuals. This parallels the well-characterized role of FOXP3 in CD4+ regulatory T cells (Tregs), where FOXP3 binding directly activates CTLA4 transcription to enforce immunosuppressive functions [54,55]. In the context of CD8+ T cell exhaustion, our data suggest a conserved regulatory mechanism wherein FOXP family members may drive CTLA4 expression to reinforce an exhausted phenotype. Although FOXP3 is not typically expressed in CD8+ T cells, the observed accessibility at FOXP motifs suggests

that other FOXP family members may be active in this context. However, our current dataset lacks the resolution to pinpoint the specific FOXP member(s) involved. Future studies employing single-cell transcriptome and TF binding profiling in HIV-specific CD8+ T cells could clarify this mechanism.

We found that classical CD14+ monocytes exhibited markedly altered chromatin accessibility at cCREs in COVID-19 and that the magnitude of these changes was correlated with disease severity. Herein, target binding motifs of NF-kB factors exhibited decreased chromatin accessibility in severe COVID-19 cases, and we observed an enrichment of CREB/ATF factors among differentially accessible cCREs, potentially contributing to the inhibition of NF-kB activity, consistent with previous reports [5,26]. We also found an enrichment of AP1 factors, suggesting a shift in NF-kB-mediated regulation of the pro-inflammatory cytokine network to AP1-mediated regulation in severe COVID-19. Previous studies indicate that silencing NF-kB factors turns CD14+ monocytes into a cell state resembling myeloid-derived suppressor cells (MDSCs) during severe COVID-19, resulting in decreased expression of HLA class II genes [5]. In line with this observation, we found decreased accessibility in the HLA class II gene vicinity.

Our integration of aggregated DNA methylation signals in cCREs with scATAC-seq data in paired cohorts demonstrated (anti-)correspondence between these modalities at the broad cell-type level. This correspondence was less pronounced when considering T cell subpopulations of cells, indicating complementary information content of both epigenetic modalities. This discrepancy suggests the transition from naive to memory state in T-cells may involve regulatory mechanisms captured not by accessibility alone, with DNA methylation providing an additional layer of information defining these cell states. DNA methylation and chromatin accessibility may be dynamic at different rates during the process of memory formation. Specifically, methylation-sensitive transcription factors (TFs), as revealed by methyl-SELEX [30], can exhibit distinct binding behaviors; DNA methylation suppresses TF binding in

methylMinus motifs, whereas DNA methylation promotes TF binding in methylPlus motifs. We developed a method to infer transcription factor (TF) activity from the DNA methylation footprints in TFBS, providing a complementary perspective on accessibility-based measurements of TF binding dynamics. We then systematically compared TF activities inferred from DNA methylation with those derived from accessibility. Our evaluation revealed that while accessibility-based TF activity provides a robust method to distinguish T-cell clusters, methylation-based TF activity offers a more specific readout of TF-driven regulation of T-cell memory formation. FOXP1 and IRF4, known memory formation markers in T-cells, exhibited epigenetic concordance across modalities. In memory T-cells, IRF4 motifs displayed increased accessibility and decreased methylation compared to naive T-cells. In contrast, FOXP1 displayed the opposite pattern with nominal changes, consistent with previous studies reporting accessibility changes in these TFs [33]. We found that AP1 factors, such as BATF, exhibited starkly altered activity in memory formation, indicated by increased accessibility and decreased methylation. In contrast, NF-κB factors were predominantly influenced by methylation changes with minimal accessibility shifts. This suggests that different TF families are governed by distinct epigenetic mechanisms that regulate T-cell memory. Further experimental and computational validation is needed to examine the interactions between these TFs, particularly how varying accessibility and methylation levels among different TF networks coordinate together to define memory cell identity.

Our TF activity analysis also revealed that differences in accessibility between naive and CD8 memory T-cells are more pronounced than in CD4 memory T-cells (Figure 4D, Supplementary Figure 5A), consistent with the previous observations in the literature Wang et al.,2023 [40], indicating more substantial TF-associated chromatin remodeling. For example, EOMES exhibited increased accessibility with only a minimal decrease in DNA methylation in CD8 T cells, whereas CD4 memory T-cells showed only a nominal change. However, since a portion of our data for matching multimodal single-cell epigenomes was obtained from our HIV exposure

cohort, this may be the effect of HIV induced exhaustion of CD8 T memory cells and a general depletion of the CD4 population. Further experimentation is necessary to validate the TF binding dynamics of EOMES upon HIV virus exposure. We also observed that the changes in accessibility and methylation of TF binding regions are not uniformly pronounced across all TFs. While some TFs show significant variations between memory and naive T-cells, others display less variation. This emphasizes the need for additional experiments on why some TFs are susceptible to epigenetic regulation while others maintain epigenetic stability during the transition from naive to memory state in T-cells.

Consistent with our accessibility findings, the most pronounced differentially methylated cCREs were observed in monocytes from severe COVID-19 cases compared to healthy controls. TF enrichment analysis in severe DMRs identified a highly overlapping set of TF motifs enriched in accessibility analysis, suggesting that both epigenomic modalities contribute to a TF-driven shift in cytokine network regulation. Moreover, multimodal TF activity analysis confirmed that these TFs enriched in cCREs, such as AP1 and NF-kB factors, were also differentially active, highlighting their role in severe COVID-19.

We further supported these findings through TF motif footprinting, which revealed that under severe COVID-19 conditions, monocyte regulatory programs undergo selective remodeling. Specifically, AP-1 actors such as JUN, FOSB::JUNB, and FOSL2::JUN displayed nominally increased accessibility and decreased methylation in severe cases compared to controls, suggesting activation of AP-1 mediated networks. In contrast, NF-κB factors such as REL and RELA display a nominal decrease in accessibility without accompanying methylation changes in TF footprints. Although NF-κB factors show only modest changes in TF footprinting, their motifs are significantly enriched in DARs. This suggests that chromatin at NF-κB target sites becomes temporarily less accessible, even though NF-κB remains bound. The absence of accompanying DNA methylation changes further supports a model of short-term or reversible repression, rather than stable silencing of NF-κB–driven regulatory programs.

Consistent with the previous reports of Wimmers 2023 and Wilk 2021, our findings indicate stable activation of AP1 factors in severe COVID-19 [5,26]. This implies that AP1 remodeling may underlie dysregulated monocyte responses and sustained inflammation. While our inferring TF activity of monocytes identified the pivotal AP1 factors that are activated and NF-κB factors transiently modulated in severe COVID-19, we note several limitations. First, motif-based footprinting infers changes in TF activity from accessibility and methylation; however, it does not directly measure TF binding. Second, our analysis focused on monocytes from patients with established severe COVID-19; thus, it remains unclear whether the observed epigenetic remodeling is the cause of disease progression. Longitudinal studies are necessary to track the accessibility and methylation dynamics from early infection through clinically worsening or recovery states to assess the relationship between TF-based remodeling and disease severity.

In summary, our comprehensive multimodal atlas of human immune cell epigenomic remodeling across diverse environmental stimuli offers novel insights into T-cell exhaustion dynamics in chronic HIV infection, the regulatory TFs involved in T-cell memory formation, and epigenetic deregulation in monocyte in severe COVID-19 disease. By integrating chromatin accessibility with DNA methylation at single-cell resolution, this work established a framework for dissecting stimulus-specific transcriptional regulation and provides a valuable resource for future studies on immune response and dysfunctions associated with environmental exposures.

Methods

Sample acquisition and patient consent

Human peripheral blood mononuclear cells (PBMCs) were collected and isolated from different subjects in different cohorts exposed to various chemical or biological pathogens, including organophosphates, Influenza vaccines (followed by live Influenza virus challenge), HIV, and

SARS-CoV-2 viruses. The subjects or surrogates have informed and approved all patient consents in different cores.

snATAC-seq library generation, sequencing, and alignment

The snATAC-seq assay was performed by using the Chromium Next GEM Single Cell ATAC Reagent Kits version 1.1 (10x Genomics; PN-1000175) with the demonstrated protocol (nuclei isolation for scATAC-seq from PBMCs) provided by 10x Genomics. Briefly, cryopreserved PBMCs with around 100,000 cells were thawed and resuspended in RPMI medium, washed with PBS twice, and then lysed with 100 µl lysis buffer for 5 min. The lysed nuclei were centrifuged after washing with 1 ml washing buffer and then resuspended in diluted nuclei buffer. The nuclei concentration was then determined with a TC20 Automated Cell Counter (Bio-Rad Laboratories; 1450102), and around 8,000-10,000 single nuclei were used for downstream steps: Tn5 transposition, single nuclei barcoding, and library preparation following the instructions in the kit. The final DNA libraries were sequenced with the NovaSeq 6000 system (Illumina), leading to around 200 million read pairs per sample. Data processing was performed with cellranger-atac software (10x Genomics; version 1.2.0). The demultiplexed sequencing reads were aligned to the GRCh38 (hg38) human reference genome, resulting in the related BAM files and fragment files for downstream analysis.

snATAC-seq quality control, dimensionality reduction, clustering, and batch correction

Following preprocessing, we utilized the ArchR package (version 1.0.2) [40] to process snATAC-seq fragments. We added human reference genome (hg38) annotation with ArchR's "addArchRGenome" function. Transcription start site (TSS) enrichment and the number of unique fragments were computed to assess the quality of the cells. Cells with TSS enrichment above 4 and log(nFragments) above 3 were retained. ArchR's "addDoubletScores" was used to

identify doublets, and “filterDoublets” with a ratio of 1.0 was used to remove doublets. 12,684 cells were identified and removed as doublets. 271,299 cells were retained after the initial filtering and quality control processes.

ArchR’s implementation of iterative LSI used dimensional reduction by “addIterativeLSI" with 25,000 variable features (tiling regions) and 30 dimensions. We identified clusters using “addClusters” with 0.8 resolution and obtained 22 clusters. For integration across exposure cohorts, Harmony batch correction was applied using “addHarmony” with “iterativeLSI” as reduced dimensions. Then, “addUMAP” was used to obtain the two-dimensional representation of the data with reducedDims = "Harmony"**,** nNeighbors = 40, minDist = 0.5, and metric = "cosine". Batch-corrected data is only used for visualization purposes.

Gene activity score, weighted chromatin accessibility around each gene, of marker genes obtained using “getMarkerFeatures”. We then visualize these markers' smoothened gene activity score on UMAP, where smoothening is applied using the MAGIC algorithm through ArchR’s “addImputeWeights**”**. Data smoothing was only used for visualization and not for downstream analyses. “addMotifAnnotations” adds the “JASPAR2020” motif set annotation. Deviation scores are computed using “addDeviationsMatrix”, and the deviation Z-scores of marker motifs were plotted on UMAP. Finally, track plots of marker genes are plotted using “plotBrowserTrack” from the ArchR package.

Cell-type annotation, peak-calling, and generating pseudo bulk replicates

Before peak calling, pseudo-bulk replicates are generated per cluster using “addGroupCoverages”. Then, we performed peak-calling with MACS2, using ArchR’s “addReproduciblePeakSet” with default settings. Canonical correlation analysis-based integration was performed to integrate the scRNA-seq ImmuneAtlas dataset [citation] and

scATAC-seq through ArchR's "addGeneIntegrationMatrix" with default settings. We computed the Jaccard Index of scATAC cluster labels and their corresponding predicted scRNA labels. The final cell types were annotated to clusters based on the Jaccard similarity assessment following the integration (Supplementary Figure 1D). We assessed the significance of cell-type proportions by exposure condition as follows: The cell-type proportions per sample were obtained using the "prop.table" function. To compare disease vs control conditions, we generated boxplots with ggplot2 [42] and applied "stat_compare_means" from the ggpubr package [43] using the Wilcoxon rank-sum test and Bonferroni correction for multiple testing.

For pseudobulk analyses, we summarized fragment counts per cell type and sample as follows: Cell-type-specific fragments were extracted from Arrow files using "getFragmentsFromArrow". Samples that didn't contain at least 50 fragments per cell type were discarded from the analysis. In total, 688 pseudobulk samples were retained for use in the analysis. Cell-type specific fragments are then exported as bed files using "export.bed" from the rtracklayer R package [44].

TF motif cluster annotations [45] were added to the ArchR project using "addPeakAnnotations" from the ArchR package. Deviation scores for Viestra motifs were computed using "addDeviationsMatrix". The added deviation Z-score matrix with annotated cell types was then extracted using the "getGroupSE" function with the following parameters: groupBy =" ClusterCellTypes", divideN=TRUE. The selected variable motifs were filtered based on Z-scores, retaining those with absolute Z-scores greater than 0.01 or less than or equal to 5.

Cell-type-specific differential accessibility analysis using pseudo-bulk samples

Pseudo-bulk samples per cell type and sample were processed, and differential analysis was performed using the "run_chraccr" function from the ChrAccR package [46]. The configuration of vanilla ChrAccR analysis is as follows: differentialColumns =sample_exposure_group,

filteringCovgCount=1L, filteringSexChroms=TRUE, filteringCovgReqSamples=0.005, differentialCutoffL2FC=0.5. By default, ChrAccR uses the DESeq2 package for differential analysis, which requires at least 3 samples per comparison group to perform pairwise differential analysis. For the COVID-19 cohort, the data processing date was used as the batch confounder by setting differentialAdjColumns = “processing_date” in the ChrAccR configuration. TF motif enrichment was done by LOLA implementation in ChrAccR by configuring the LOLA database to “lolaDbHg38”. Enrichment for TF motifs in differentially accessible regions for both hyper- and hypo-accessible regions was based on a Q-value threshold (q =< 0.05).

Differential gene expression analysis of CD14+ Monocytes in COVID-19 using scRNA-seq and cross-modality comparisons

Samples from our COVID-19 cohort have paired publicly available single-cell transcriptomics data [3] for 34,548 cells. We subsetted $CD14^+$ classical monocytes from the available Seurat object (see Supplementary Data) and then retained only protein-coding genes from the subsetted Seurat object. Next, we ran a pairwise differential analysis between the control and severe groups of COVID-19 with “DEseq2” using the “FindMarkers” function from Seurat [48]. Simultaneously, we subsetted $CD14^+$ classical monocytes coming from the COVID-19 cohort with the same set of genes from the chromatin-derived gene-activity matrix in ArchR and ran the Wilcox test between the control and severe groups. We plotted a scatter plot of fold changes from the ArchR and Seurat analysis described above while highlighting the differential protein-coding genes (adj. p-value < 0.05).

Processing and integrating snmC-seq with snATAC-seq

snmC-seq data for 38 samples and 22,488 cells corresponded to 110,376 cells from the scATAC-seq dataset. In total, 16,835 cells profiles were obtained for the COVID-19 cohort. A fluorescence-activated cell sorting (FACS) antibody cocktail was used to sort the immune cell

types into 384-well plates based on their antibody labeling. Raw ALLC files from snmC-seq [16] were processed by summing CpG methylation and total counts were over peak regions obtained from ArchR's "getPeakSet" function. Sex chromosomes and regions that weren't covered by at least 20 percent of all cells were removed from the dataset. To obtain a feature matrix for integration with chromatin accessibility data, the inverted binarization procedure was applied to the retained DNA methylation dataset by setting regions with DNA methylation values below 0.4 to 1 and all other values to 0, for subsequent imputation. Importantly, missing values (NA) were preserved during this step and not coerced to 0, in order to allow for downstream imputation. Then, iterative LSI and clustering were applied to the binarized dataset with the following parameters: 50 dimensions, imputeMode="pcimpute", and 20,000 features. The details of the imputation process are described below. Reduced dimensions were then visualized on UMAP using the uwot package [47]. Imputed data is only used for iterative LSI and clustering, not for downstream processes.

For integration with scATAC data, the imputation of the sparse DNA methylation matrix is as follows: in the first iteration, the input methylation matrix is centered (but not scaled), and initially, each missing value in the input matrix is replaced with zero. Then PCA is performed using the irlba package [49]. The matrix is then reconstructed from the principal components by multiplying the scores (PCA-transformed values) with the transposed loadings (eigenvectors). This reconstructed matrix provides updated estimates for the missing values, which are then used to replace the initially imputed zeros. This process is repeated iteratively, updating the missing values with the estimated ones in each iteration until convergence. The convergence criterion is based on the mean squared difference between consecutive imputations and stops when the estimated imputation values change insignificantly. This procedure is iteratively performed until convergence. The resulting imputed data is only used for visualization purposes.

To integrate chromatin accessibility data with the DNA methylation dataset, the chromatin accessibility data were binarized by setting the regions with at least one Tn5 insertion and the

rest as 0. Term frequency-inverse document frequency (TF-IDF) transformation was applied to binarized chromatin accessibility and inverted-binarized DNA methylation data of the shared samples. The resulting matrices were then used to create Seurat objects [48]. Canonical correlation analysis (CCA)-based integration is performed via “FindTransferAnchors” from the Seurat package with default parameters. We then computed nearest neighbors in CCA embeddings for both datasets and visualized the mapped cell labels on UMAP for each dataset.

Cell-type-specific differential methylation analysis using pseudo-bulk samples

The custom function “createPseudoBulks" (see Supplementary Code) generates pseudo bulk samples by summing the methylation site counts per donor, cell type, and exposure condition. It takes the maximum single-base coverage (default: 99999), sample annotation, grouping variables for summarizing methylated sites, excluded chromosomes (default: sex chromosomes), and strand specificity (default: FALSE) as parameters. The resulting pseudo-bulks were saved as bed files. In total, 1015 pseudobulk samples were generated.

For the analysis of methylation pseudobulk samples, the RnBeads package version 2.12.2 is used with the following settings: region.aggregation = "sum", filtering.sex.chromosomes.removal = TRUE, and strand.specific = FALSE [50]. scATAC-seq peak regions were used as the reference regions for aggregating methylation counts. Differential analysis was conducted with RnBeads` “rnb.run.differential” wrapper function for limma-based differential analysis [51], where each exposure group was compared with the control samples per cell type. LOLA enrichment analysis was performed using the same settings as described in the ATAC-seq analysis.

Defining TF regulator activity from DNA methylation and cross-modality TF activity analysis

We computed a TF activity score based on how the DNA methylation signal deviates at the motif center from the genomic background computed at +/-250bp distance away from the motif center (Supplementary Figure 4A). To account for potential bias incurred by DNA sequence composition, we calculated both an observed and an expected deviation score and subtracted the expected deviation score from the observed one. The observed deviation score is computed by aggregating the DNA methylation signal of the sample across putative binding sites of TFs (TFBS) and flanking regions, followed by calculating the deviation score as Deviation score = (Average methylation in footprint base) / (Average methylation in the background).

To compute the expected methylation levels, we first divided the whole human genome into 30 base-paired bins and subsetted the distal enhancer regions [51]. We then calculated the GC frequency for each bin and discretized these into five frequency levels. For each frequency level, we computed the average methylation level across corresponding bins, which we refer to as the GC frequency table. These average values were then used to compute the methylation levels of the corresponding bins around TFBS in each sample and for each TF, which are referred to as the expected average methylation levels. We computed the expected deviation score using these expected methylation levels by dividing the expected methylation level at the motif centers by the average of the expected methylation levels at the motif flanking regions. We subtracted the expected deviation score from the observed deviation score to obtain the bias-corrected deviation score.

For heatmap visualizations, we applied row-wise Z-score transformation to normalize the bias-corrected TF deviation scores per TF motif. We developed an R package, methylTFR, and methylTFRAnnotationCreator to derive transcription factor (TF) activity from DNA methylation

and generate the annotations described above (See Supplementary Code). We presented pre-computed annotations for JASPAR2020 and Altius TF motif sets [31,45] for Human Genome assembly version 38 as an R package called methylTFRAnnotationHg38 (see Supplementary Code).

We used the “run_methyltfr” function from the methylTFR package to compute bias-corrected deviation scores from the DNA methylation dataset for the JASPAR2020 motif sets. This function takes a sample annotation, a genome-wide GC distribution object, transcription factor binding sites, an enhancer region, a minimum coverage threshold per site (default=5), and a precomputed GC frequency table per annotation as input parameters. To obtain bias-corrected TF deviation scores from the chromatin accessibility dataset for JASPAR2020 motifs, we used the chromVAR package [21]. We subsetted the variable motifs using Z-scores based on a threshold (>=0.01 & <= 5).

We used chromVAR’s “differentialDeviations” function with the Wilcox test to identify differentially active TF motifs based on chromatin accessibility-derived TF deviation scores, using the parameters parametric=TRUE and alternative=”two.sided”. Additionally, we adapted this function to analyze DNA methylation-derived TF deviation scores. Motifs with adjusted p-values lower than 0.05 were saved as tsv files. The correlation between the chromVAR and methylTFR deviation scores among the shared samples was assessed using a Pearson correlation test implemented through the 'cor' function from R. We obtained the motif footprints for accessibility data using the “getFootprints” function from the ArchR package using the default parameters. To plot accessibility footprints, we used “plotFootprints” from ArchR with normMethod ="divide" and used the default parameters otherwise. For DNA methylation footprints, we used the ggplot2 package to plot motif footprints we obtained from the methylTFR package.

Resource Availability

Processed data is available on the Gene Expression Omnibus (GEO) under accession number GSE306525. All code to analyze the datasets in this manuscript can be found on Github (https://github.com/EpigenomeInformatics/exposure_atlas_manuscript). The R package to compute TF activity from DNA methylation can be found on Github (https://github.com/EpigenomeInformatics/methylTFR). Fully processed scRNA data from the COVID-19 cohort were downloaded from https://www.covid19cellatlas.org/index.patient.html.

Author Contributions

W.J.G. and F.M. designed the study. W.J.G., J.R.E., and F.M. provided supervision. B.W. and D.C.C. conducted the experiments and generated the scATAC-seq dataset. A.B., C.B., R.C., and J.N. performed snmC-seq2, with A.B., C.B., and R.C. contributing to the development of the PBMC gating strategy and FACS. I.B.G. and B.W. carried out the data analysis and generated the figures. M.H. and W.W. are involved in data analysis. I.B.G., B.W., F.M., and W.J.G. wrote the manuscript. All authors read the manuscript and provided comments.

Competing Interests Statement

W.J.G. is named as an inventor on patents describing ATAC-seq methods. 10XGenomics has licensed intellectual property on which W.J.G. is listed as an inventor. W.J.G. is a consultant for Ultima Genomics and Guardant Health. W.J.G. is a scientific co-founder of Protillion Biosciences. J.R.E. serves as a consultant for Zymo Research, Inc.

Acknowledgments

We thank all members of the Ecker, Müller and Greenleaf labs for helpful discussions and advice. F.M. was supported by a NanoBioMed Young Investigator Grant at Saarland University. Figure 1A was created using BioRender. This work was supported by the Defense Advanced Research Projects Agency (DARPA) through the DARPA Epigenetic Characterization and Observation (ECHO) program for the project Single-cell Analysis for Forensic Epigenetics (SAFE), issued by the US Army Research Office under co-operative agreement W911NF-19-2-0185. J.R.E. is an Investigator of the Howard Hughes Medical Institute. We would like to thank former DARPA Biological Technologies Offices Program Manager Eric Van Gieson, as well as the current leadership team including Jean-Paul Chretien and Thomas Thomou, for their guidance and insightful comments. W.J.G. acknowledges funding support from National Institutes of Health (NIH) grants P50-HG007735 and UM1-HG009442, UM1-HG009436. V.G.F acknowledges funding from NIH grant 1R01AI165671. T.E. acknowledges Armen Donabedian. We would like to thank all the unidentified donors who contributed biological samples for this project through our collaborators.

Figure Legends

Figure 1: The chromatin accessibility landscape of human immune cells across viral and chemical exposures

A) Overview of sample cohorts and single-cell chromatin accessibility profiling B) UMAP dimensionality reduction of the 271,299 cells from the scATAC dataset colored by immune cell population C) UMAP dimensionality reduction colored by the relative gene activity scores (left) and transcription factor (TF) motif activity (ChromVAR, right) of IRF8, GATA3, TBX21, and PAX5 D) Genome accessibility tracks by cell type with a focus on genes (IRF8 and GATA3) and, their respective motif occurrences E) Heatmap of the 57 top variable TF motif cluster [45]

deviation Z-scores (ChromVAR) across cell types F) Bar plot of cell type composition across cohorts and exposure groups

Figure 2. Chromatin accessibility dynamics in T cell populations upon HIV exposure

A) UMAP projection of the T-cell subpopulation from the HIV-subsetted scATAC dataset colored by subjects (left) and colored by Louvain clusters (right) B) Side-by-side UMAP projections of the HIV-subsetted scATAC dataset colored by gene activity scores of T-cell exhaustion markers CTLA4 and HAVCR2 C) Heatmap representation of differentially accessible marker genes in differet clusters in CD8+ T cell population based on gene matrix scores D) Cell type proportion relative to the pre-stage over time for exhausted CD4+T cells (upper) and exhausted CD8+T cells (lower) per subject E) Viral load per subject over the disease progress of HIV F) Fold-change versus mean count (Volcano) plot of exhausted (cluster 2) and non-exhausted T-cells in the HIV cohort, hyper-accessible differential peaks highlighted in red, hypo-accessible peaks highlighted in blue. Differential accessibility was defined using an FDR ≤ 0.01 and $|\log_2$ fold change$| \geq 1.25$. G) TF motif enrichment plots in hyper-accessible (left) and hypo-accessible (right) H) Genome accessibility tracks per CD8 T subpopulation with a focus on T-cell exhaustion markers CTLA4 and HAVCR2 along with occurrences of the FOXP2 motif

Figure 3. Impact of COVID-19 on chromatin accessibility profiles of CD14+ classical monocytes

A) Fold-change versus mean count (MA) plots comparing moderate (left) and severe (right) cases in the COVID-19 cohort to healthy controls, hyper- and hypo-accessible DARs (p.adj < 0.05, log2fold-change > 0.5) are highlighted in green and blue, respectively. Point density is visualized as a 2D gradient; low-density outliers (< 0.05, $|\log_2$(fold-change) >= 0.5|) in CD14+ classical monocytes in mild, moderate, and severe cases of COVID-19 along with control samples. Hierarchical clustering with Euclidean distance (complete linkage) was applied to both rows and columns within the heatmap function. D) Enrichment analysis of TF motif clusters in

DARs (Fisher's Exact Test using LOLA) upon severe COVID-19 exposure compared to control [45]. The hyperaccessible cCREs are highlighted in red, while the hypoaccessible cCREs are highlighted in blue. E) Scatterplot comparing the fold-change in differential gene activity versus the fold-change obtained from differential gene expression analysis for protein-coding genes, with genes from Wimmers 2023 and Wilk 2021 highlighted [26,5]. F) Genome accessibility track visualization of the differential cCREs at the IL1B and IL6 gene loci

Figure 4. Integration of chromatin accessibility and DNA methylation, epigenetic TF activity across cell states

A) UMAP projections of 110,376 cells from scATAC with cluster cell labels (top-left) and mapped snmC labels (bottom-left), UMAP projection of 22,488 cells from snmC with FACS sorted cell labels (top-right) and, mapped ATAC labels (bottom-right) B) Heatmap of cell-type correspondence between scATAC-seq and snmC-seq. Upper triangle: projection of scATAC onto the snmC manifold; lower triangle: snmC onto scATAC. Values indicate the proportion of mapped cells per input cell type (row-normalized). Columns represent pseudobulk cell types. C) Heatmap of motifs with variable (0.01 =< z-score =< 5) chromVAR deviation Z-scores (left) and, their corresponding methylTFR deviation Z-scores (right) along with row-wise pseudobulk Pearson correlation coefficients and methyl-SELEX [29] annotations. D) TF motif activity Z-score difference (CD4T-Memory Z-score subtracted from T-naive Z-score) between chromatin accessibility and DNA methylation-based TF activities. The overall Pearson correlation is -0.28, and differentially active motifs identified from a two-sided Wilcoxon test (padj < 0.05) are highlighted. E) Side-by-side accessibility and methylation footprint plots of FOS and F) IRF4 motifs

Figure 5. Impact of COVID-19 on DNA methylation profiles of cCREs in CD14+ classical Monocytes

A) Density scatter plots comparing aggregated DNA methylation levels in peak regions between commercial controls and mild (left, n = 87) or severe (right, n = 2073) COVID-19

samples. Differentially methylated peaks (adj. $p < 0.05$) are highlighted in blue for control group and, in orange for the severe group. Point density is visualized using 2D kernel density estimation; outliers in low-density regions (<0.1% of all points) are shown individually (sparse.points = 0.001). B) Scatter plot of mean DNA methylation level difference between COVID-19 Severe and control samples versus fold-change obtained from differential accessibility between COVID-19 Severe and control samples, including a regression line based on all cCREs C) Enrichment analysis of TF motif clusters (LOLA) in hypermethylated and hypomethylated DMRs upon severe COVID-19 compared to control [45]. The gain in methylation enrichment is highlighted in orange, while the loss in methylation enrichment is highlighted in blue. D) Heatmap showing row-wise Z-scores of transcription factor (TF) activity: DNA methylationbased TF activity Z-scores (left) and chromatin accessibility-based differential TF activity Z-scores (right). Rows are clustered based on methylation profiles. E) Sideby-side accessibility and methylation footprint plots of motifs with differential TF activity (SPIB, FOSL2::JUNB(var.2), CEBPD, RELA, respectively) (SPIB, FOSL2::JUNB(var.2), CEBPD, RELA, respectively)

Supplementary Figure 1: Quality control of single-cell ATAC dataset

A) Violin plot of transcription start site (TSS) enrichment (top) and the number of unique fragments (bottom) across the exposure conditions B) UMAP projections of full scATAC dataset, colored by Louvain clusters (left) and colored by exposure cohort (right) C) Bar plot of the number of peaks by the number of unique fragments of each cluster, colored by peak annotation D) Correspondence between scATAC clusters and scRNA labels colored by Jaccard Index E) UMAP projections of the full scATAC dataset colored by relative gene activity scores of canonical markers (CD3D, CD14, CD34, and MS4A1) F) Genome accessibility track visualization of the full scATAC dataset with a focus on canonical surface markers (TBX21, PAX5, and IRF4) and, their respective motif occurrences G) Box plots depicting cell-type proportions across the cohort. Statistical significance was assessed using a two-sided Wilcoxon

rank-sum test with Bonferroni's correction for multiple hypothesis testing. Significance levels are denoted as follows: * padj < 0.05, ** padj < 0.01, and *** padj < 0.001.

Supplementary Figure 2: Differential chromatin accessibility analysis

A) MA plots per cell type and exposure condition comparisons; Comparisons with insufficient sample size for differential analysis are marked "Comparison is not available." Differentials are highlighted with red color (p.adj < 0.5 and |log2(fold-change) >= 0.5)|. Point density is visualized using 2D kernel density estimation; outlier points from low-density regions (<0.1% of all points) are plotted individually (sparse.points = 0.001).

Supplementary Figure 3: Differential accessibility analysis in the COVID-19 cohort

A) MA plots per cell type and COVID-19 condition comparisons; Comparisons with insufficient sample size for differential analysis are marked "Comparison is not available". Differential peaks (adj.p < 0.05 and $|\log_2 FC| \geq 0.5$) are shown in red. Point density is visualized using 2D kernel density estimation; outlier points from low-density regions (<0.1% of of all points) are plotted individually (sparse.points = 0.001). B) Genome accessibility track visualization of differential peaks upon SARS-CoV-2 exposure with highlighted genes, including HLA Class II families, and CD4. C) GO enrichment analysis results for genes with increased expression/accessibility (top) and genes with decreased expression/accessibility (bottom), with previously reported genes highlighted [53]

Supplementary Figure 4: Multimodal TF activity analysis

A) Schematic of motif footprint-based TF deviation score computation with methylTFR. B) PCA plots of chromVAR deviations (left) with snmC-mapped cell types and methylTFR deviations (right) with sorted cell types. C) Boxplots of Spearman correlation coefficients between chromVAR and methylTFR TF activities across methyl-SELEX labels, with the number of motifs per category annotated on top. D) Scatterplots comparing motif activity per cell type. Motif activity was defined as the z-score of methylTFR vs chromVAR. Differentially active motifs in

the two modalities (One vs other cell types, padj < 0.05) are highlighted as: red differential in both, green differential only in chromVAR, blue: only differential in methylTFR, gray: not-differential.

Supplementary Figure 5: TF signatures of memory formation in T-cells

A) Z-score difference (CD8T-Memory Z-score subtracted from T-naive Z-score) between DNA methylation-based TF activity and chromatin accessibility-based TF activity, with highlighted differential motifs. B) Side-by-side accessibility and methylation footprint plots of TF motifs with variable TF activity (CEBPE, BATF, RELA, RUNX3, FOSL2::JUND, EOMES, MAF::NFE2, FOXP1, and ETS1, respectively)

Supplementary Figure 6: Differential DNA methylation analysis in the COVID-19 cohort

A) Density scatter plots of COVID-19 condition compared with controls per cell type. Point density is visualized using 2D kernel density estimation; outlier points from low-density regions (<0.1% of of all points) are plotted individually (sparse.points = 0.001). B B) Side-byside accessibility and methylation footprint plots of differentially active TF motifs (CREB1, CEBPA, JUN, BATF, REL, BATF::JUN, SPIC and, FOSB::JUN respectively)

Supplementary Tables

Table S1: Sample metadata of the scATAC dataset

Table S2: Pairwise differential TF motif activity (Wilcoxon test of chromVAR deviation scores) across different cell types

Table S3: List of differentially accessible genes in different clusters within CD8+ T cells

Table S4: List of differentially accessible cCREs in different clusters within CD8+ T cells

Table S5: Differential gene activity and gene expression table of protein-coding genes for COVID-19 severe vs control in CD14+ monocytes

Table S6: Differentially accessible cCREs for COVID-19 severe vs control in CD14+ monocytes

Table S7: Pairwise Wilcoxon test between one vs other cell type manner for methylTFR and chromVAR z-scores

Table S8: Differentially methylated cCREs in CD14+ monocytes

References

1. Kazer, S.W., Aicher, T.P., Muema, D.M., Carroll, S.L., Ordovas-Montanes, J., Miao, V.N., Tu, A.A., Ziegler, C.G.K., Nyquist, S.K., Wong, E.B., et al. (2020). Integrated single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. Nat. Med. *26*, 511–518. https://doi.org/10.1038/s41591-020-0799-2.
2. Wimmers, F., Burrell, A.R., Feng, Y., Zheng, H., Arunachalam, P.S., Hu, M., Spranger, S., Nyhoff, L.E., Joshi, D., Trisal, M., et al. (2023). Multi-omics analysis of mucosal and systemic immunity to SARS-CoV-2 after birth. Cell *186*, 4632-4651.e23. https://doi.org/10.1016/j.cell.2023.08.044.
3. 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. https://doi.org/10.1038/s41591-020-0944-y.
4. Zhu, L., Yang, P., Zhao, Y., Zhuang, Z., Wang, Z., Song, R., Zhang, J., Liu, C., Gao, Q., Xu, Q., et al. (2020). Single-Cell Sequencing of Peripheral Mononuclear Cells Reveals Distinct Immune Response Landscapes of COVID-19 and Influenza Patients. Immunity *53*, 685-696.e3. https://doi.org/10.1016/j.immuni.2020.07.009.
5. 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. https://doi.org/10.1084/jem.20210582.
6. Liu, N., Jiang, C., Cai, P., Shen, Z., Sun, W., Xu, H., Fang, M., Yao, X., Zhu, L., Gao, X., et al. (2021). Single-cell analysis of COVID-19, sepsis, and HIV infection reveals

hyperinflammatory and immunosuppressive signatures in monocytes. Cell Rep. *37*, 109793. https://doi.org/10.1016/j.celrep.2021.109793.

7. Peter, J.V., Sudarsan, T.I., and Moran, J.L. (2014). Clinical features of organophosphate poisoning: A review of different classification systems and approaches. Indian J. Crit. Care Med.□: Peer-Rev., Off. Publ. Indian Soc. Crit. Care Med. *18*, 735–745. https://doi.org/10.4103/0972-5229.144017.
8. Martin, T.J., and Whalen, M.M. (2017). Exposures to the environmental toxicants pentachlorophenol (PCP) and dichlorodiphenyltrichloroethane (DDT) modify the secretion of interleukin 1-beta (IL-1β) from human immune cells. Arch. Toxicol. *91*, 1795–1808. https://doi.org/10.1007/s00204-016-1829-1.
9. Paul, K.C., Chuang, Y.-H., Cockburn, M., Bronstein, J.M., Horvath, S., and Ritz, B. (2018). Organophosphate pesticide exposure and differential genome-wide DNA methylation. Sci. Total Environ. *645*, 1135–1143. https://doi.org/10.1016/j.scitotenv.2018.07.143.
10. Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). 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.
11. Granja, J.M., Klemm, S., McGinnis, L.M., Kathiria, A.S., Mezger, A., Corces, M.R., Parks, B., Gars, E., Liedtke, M., Zheng, G.X.Y., et al. (2019). Single-cell multi-omic analysis identifies regulatory programs in mixed-phenotype acute leukemia. Nat. Biotechnol. *37*, 1458–1465. https://doi.org/10.1038/s41587-019-0332-7.
12. Trevino, A.E., Müller, F., Andersen, J., Sundaram, L., Kathiria, A., Shcherbina, A., Farh, K., Chang, H.Y., Pa□ca, A.M., Kundaje, A., et al. (2021). Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. Cell *184*, 5053-5069.e23. https://doi.org/10.1016/j.cell.2021.07.039.
13. Li, H., Tang, Y., Wang, Y., Li, Y., Yang, Y., Liao, K., Song, F., Deng, S., and Chen, Y. (2022). Single-cell sequencing resolves the landscape of immune cells and regulatory

mechanisms in HIV-infected immune non-responders. Cell Death Dis. *13*, 849. https://doi.org/10.1038/s41419-022-05225-6.

14. Wu, V.H., Nordin, J.M.L., Nguyen, S., Joy, J., Mampe, F., Estrada, P.M. del R., Torres-Ruiz, F., González-Navarro, M., Luna-Villalobos, Y.A., Ávila-Ríos, S., et al. (2023). Profound phenotypic and epigenetic heterogeneity of the HIV-1-infected CD4+ T cell reservoir. Nat. Immunol. *24*, 359–370. https://doi.org/10.1038/s41590-022-01371-3.
15. Wimmers, F., Donato, M., Kuo, A., Ashuach, T., Gupta, S., Li, C., Dvorak, M., Foecke, M.H., Chang, S.E., Hagan, T., et al. (2021). The single-cell epigenomic and transcriptional landscape of immunity to influenza vaccination. Cell *184*, 3915-3935.e21. https://doi.org/10.1016/j.cell.2021.05.039.
16. Wang, W., Hariharan, M., Bartlett, A., Barragan, C., Castanon, R., Rothenberg, V., Song, H., Nery, J., Aldridge, A., Altshul, J., et al. (2023). Human Immune Cell Epigenomic Signatures in Response to Infectious Diseases and Chemical Exposures. bioRxiv, 2023.06.29.546792. https://doi.org/10.1101/2023.06.29.546792.
17. Luo, C., Rivkin, A., Zhou, J., Sandoval, J.P., Kurihara, L., Lucero, J., Castanon, R., Nery, J.R., Pinto-Duarte, A., Bui, B., et al. (2018). Robust single-cell DNA methylome profiling with snmC-seq2. Nat. Commun. *9*, 3824. https://doi.org/10.1038/s41467-018-06355-2.
18. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.
19. Conde, C.D., Xu, C., Jarvis, L.B., Rainbow, D.B., Wells, S.B., Gomes, T., Howlett, S.K., Suchanek, O., Polanski, K., King, H.W., et al. (2022). Cross-tissue immune cell analysis reveals tissue-specific features in humans. Sci. (N. York, NY) *376*, eabl5197–eabl5197. https://doi.org/10.1126/science.abl5197.

20. Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E., et al. (2020). Global reference mapping of human transcription factor footprints. Nature *583*, 729–736. https://doi.org/10.1038/s41586-020-2528-x.
21. Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods *14*, 975–978. https://doi.org/10.1038/nmeth.4401.
22. Satpathy, A.T., Granja, J.M., Yost, K.E., Qi, Y., Meschi, F., McDermott, G.P., Olsen, B.N., Mumbach, M.R., Pierce, S.E., Corces, M.R., et al. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat. Biotechnol. *37*, 925–936. https://doi.org/10.1038/s41587-019-0206-z.
23. Shrivastava, A., Prasad, A., Kuzontkoski, P.M., Yu, J., and Groopman, J.E. (2015). Slit2N Inhibits Transmission of HIV-1 from Dendritic Cells to T-cells by Modulating Novel Cytoskeletal Elements. Sci. Rep. *5*, 16833. https://doi.org/10.1038/srep16833.
24. Anand, A.R., Nagaraja, T., and Ganju, R.K. (2011). A novel role for Slit2/Robo1 axis in modulating HIV-1 replication in T cells. AIDS *25*, 2105–2111. https://doi.org/10.1097/qad.0b013e32834bab86.
25. Lam, M.T.Y., Duttke, S.H., Odish, M.F., Le, H.D., Hansen, E.A., Nguyen, C.T., Trescott, S., Kim, R., Deota, S., Chang, M.W., et al. (2023). Dynamic activity in cis-regulatory elements of leukocytes identifies transcription factor activation and stratifies COVID-19 severity in ICU patients. Cell Rep. Med. *4*, 100935. https://doi.org/10.1016/j.xcrm.2023.100935.
26. Wimmers, F., Burrell, A.R., Feng, Y., Zheng, H., Arunachalam, P.S., Hu, M., Spranger, S., Nyhoff, L.E., Joshi, D., Trisal, M., et al. (2023). Multi-omics analysis of mucosal and systemic immunity to SARS-CoV-2 after birth. Cell *186*, 4632-4651.e23. https://doi.org/10.1016/j.cell.2023.08.044.

27. Zhen, A., Krutzik, S.R., Levin, B.R., Kasparian, S., Zack, J.A., and Kitchen, S.G. (2014). CD4 Ligation on Human Blood Monocytes Triggers Macrophage Differentiation and Enhances HIV Infection. J. Virol. *88*, 9934–9946. https://doi.org/10.1128/jvi.00616-14.
28. 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. https://doi.org/10.1016/j.cell.2020.08.001.
29. Héberlé, É., and Bardet, A.F. (2019). Sensitivity of transcription factors to DNA methylation. Essays Biochem. *63*, 727–741. https://doi.org/10.1042/ebc20190033.
30. Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science *356*. https://doi.org/10.1126/science.aaj2239.
31. Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., et al. (2019). JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *48*, D87–D92. https://doi.org/10.1093/nar/gkz1001.
32. Calderon, D., Nguyen, M.L.T., Mezger, A., Kathiria, A., Müller, F., Nguyen, V., Lescano, N., Wu, B., Trombetta, J., Ribado, J.V., et al. (2019). Landscape of stimulation-responsive chromatin across diverse human immune cells. Nat. Genet. *51*, 1494–1505. https://doi.org/10.1038/s41588-019-0505-9.
33. Durek, P., Nordström, K., Gasparoni, G., Salhab, A., Kressler, C., Almeida, M. de, Bassler, K., Ulas, T., Schmidt, F., Xiong, J., et al. (2016). Epigenomic Profiling of Human CD4+ T Cells Supports a Linear Differentiation Model and Highlights Molecular Regulators of Memory Development. Immunity *45*, 1148–1161. https://doi.org/10.1016/j.immuni.2016.10.022.
34. Abdelsamed, H.A., Zebley, C.C., Nguyen, H., Rutishauser, R.L., Fan, Y., Ghoneim, H.E., Crawford, J.C., Alfei, F., Alli, S., Ribeiro, S.P., et al. (2020). Beta cell-specific

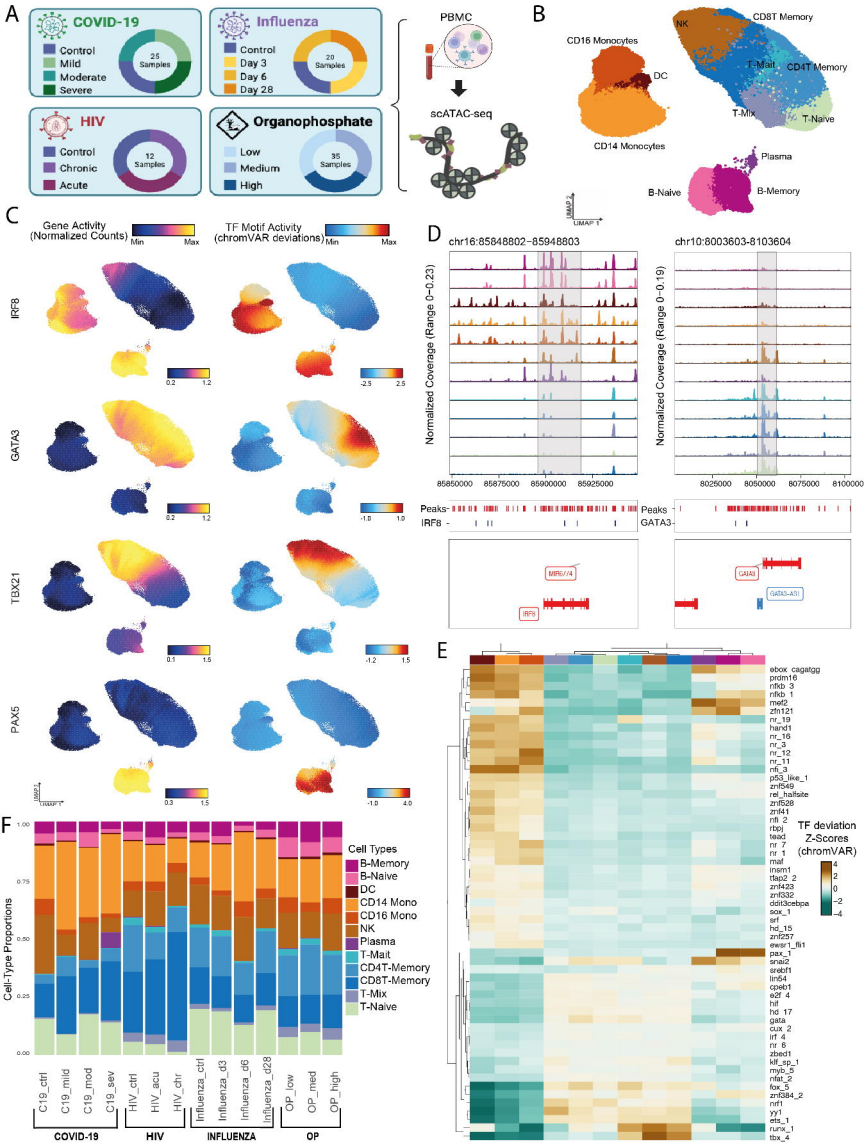
CD8+ T cells maintain stem cell memory-associated epigenetic programs during type 1 diabetes. Nat. Immunol. *21*, 578–587. https://doi.org/10.1038/s41590-020-0633-5.

35. Liu, M., Gao, W., Velkinburgh, J.C. van, Wu, Y., Ni, B., and Tian, Y. (2016). Role of Ets Proteins in Development, Differentiation, and Function of T-Cell Subsets. Med. Res. Rev. *36*, 193–220. https://doi.org/10.1002/med.21361.
36. Godoy-Tena, G., Barmada, A., Morante-Palacios, O., Calle-Fabregat, C. de la, Martins-Ferreira, R., Ferreté-Bonastre, A.G., Ciudad, L., Ruiz-Sanmartín, A., Martínez-Gallo, M., Ferrer, R., et al. (2022). Epigenetic and transcriptomic reprogramming in monocytes of severe COVID-19 patients reflects alterations in myeloid differentiation and the influence of inflammatory cytokines. Genome Med. *14*, 134. https://doi.org/10.1186/s13073-022-01137-4.
37. Wen, A.Y., Sakamoto, K.M., and Miller, L.S. (2010). The Role of the Transcription Factor CREB in Immune Function. J. Immunol. *185*, 6413–6419. https://doi.org/10.4049/jimmunol.1001829.
38. Holms, R.D. (2022). The COVID-19 Cell Signalling Problem: Spike, RAGE, PKC, p38, NFκB & IL-6 Hyper-Expression and the Human Ezrin Peptide, VIP, PKA-CREB Solution. Immuno *2*, 260–282. https://doi.org/10.3390/immuno2020017.
39. Rose, J.R., Akdogan-Ozdilek, B., Rahmberg, A.R., Powell, M.D., Hicks, S.L., Scharer, C.D., and Boss, J.M. (2023). Distinct transcriptomic and epigenomic modalities underpin human memory T cell subsets and their activation potential. Commun. Biol. *6*, 363. https://doi.org/10.1038/s42003-023-04747-9.
40. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). 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.
41. Conde, C.D., Xu, C., Jarvis, L.B., Rainbow, D.B., Wells, S.B., Gomes, T., Howlett, S.K., Suchanek, O., Polanski, K., King, H.W., et al. (2022). Cross-tissue immune cell analysis

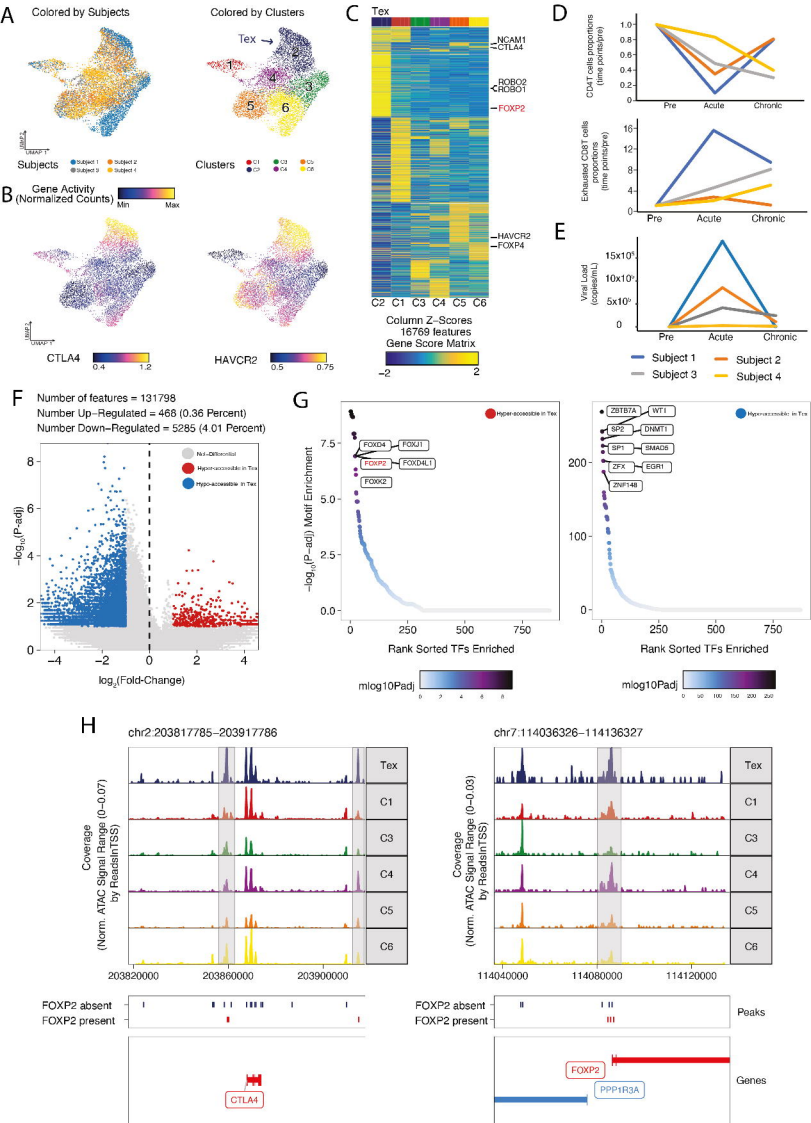
reveals tissue-specific features in humans. Sci. (N. York, NY) *376*, eabl5197–eabl5197. https://doi.org/10.1126/science.abl5197.

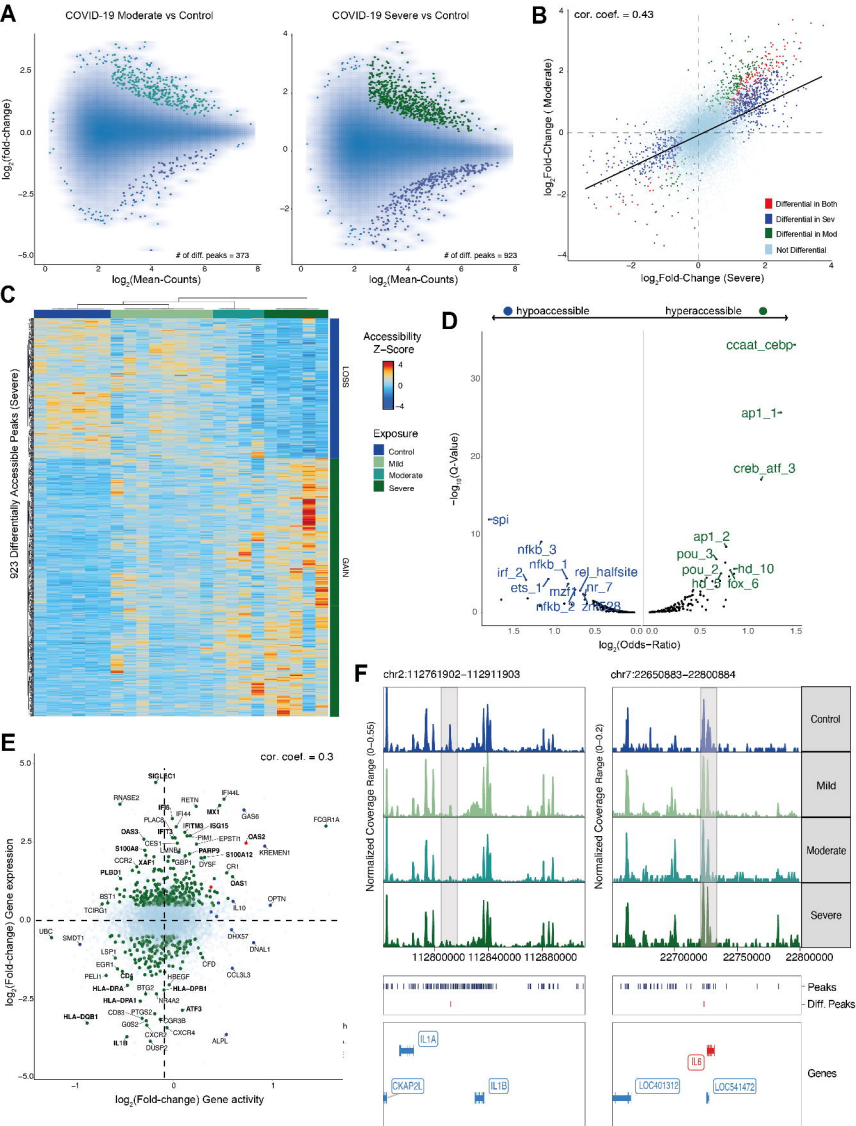
42. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
43. Kassambara, A. (2023). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.6.0. https://rpkgs.datanovia.com/ggpubr
44. Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. Bioinformatics *25*, 1841–1842. https://doi.org/10.1093/bioinformatics/btp328.
45. Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E., et al. (2020). Global reference mapping of human transcription factor footprints. Nature *583*, 729–736. https://doi.org/10.1038/s41586-020-2528-x.
46. Mueller, F. (2023). ChrAccR: Analyzing Chromatin Accessibility Data in R. R package version 0.9.21.
47. Melville, J. (2023). uwot: The Uniform Manifold Approximation and Projection (UMAP) Method for Dimensionality Reduction. R package version 0.1.16. https://CRAN.R-project.org/package=uwot
48. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., et al. (2023). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat. Biotechnol., 1–12. https://doi.org/10.1038/s41587-023-01767-y.
49. Baglama, J., Hunt, B., & Roy, A. (2024). irlba: Fast Truncated Singular Value Decomposition and Principal Components Analysis for Large Dense and Sparse Matrices. R package version 2.3.5. https://CRAN.R-project.org/package=irlba
50. Müller, F., Scherer, M., Assenov, Y., Lutsik, P., Walter, J., Lengauer, T., and Bock, C. (2019). RnBeads 2.0: comprehensive analysis of DNA methylation data. Genome Biol. *20*, 55. https://doi.org/10.1186/s13059-019-1664-9.

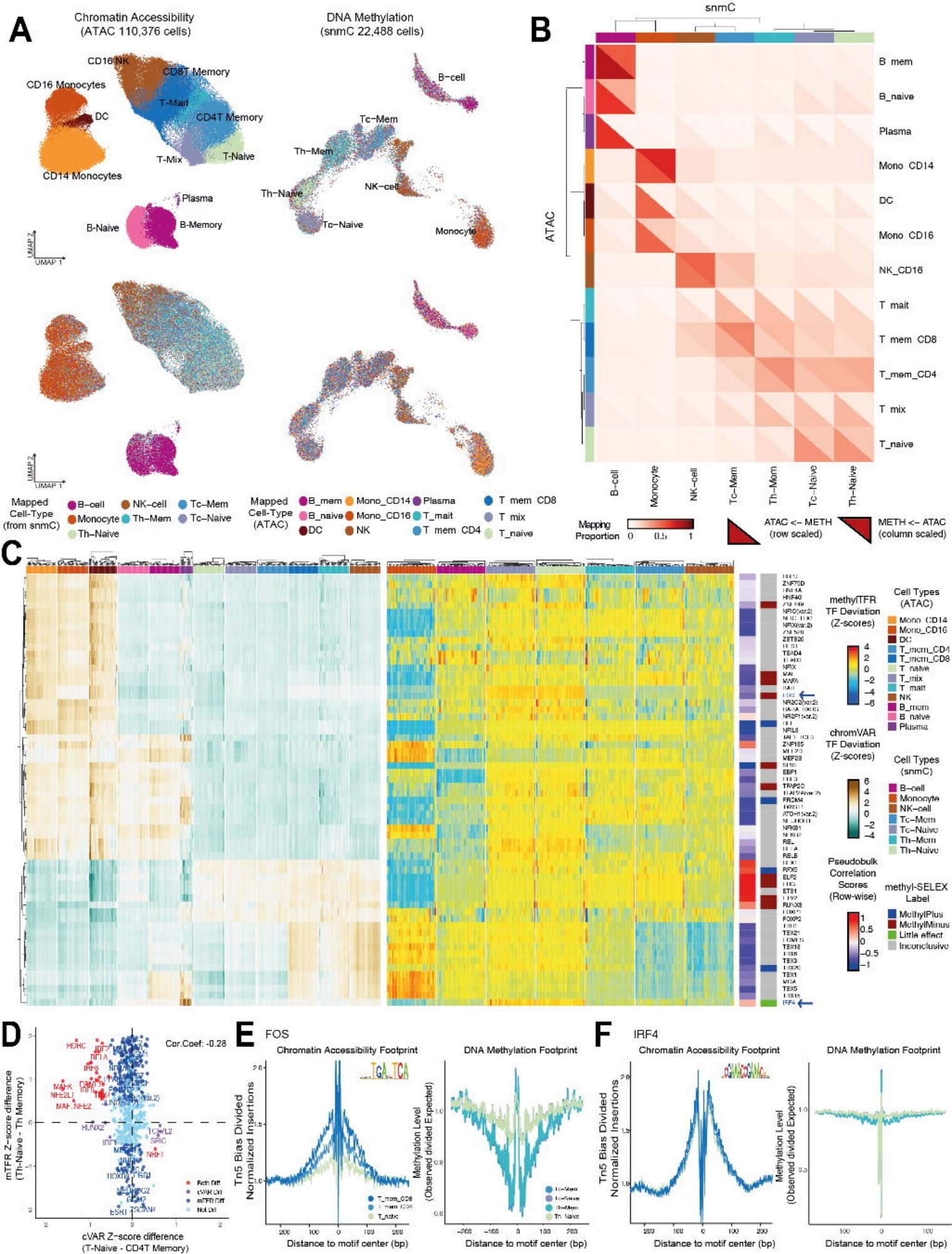
51. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47–e47. https://doi.org/10.1093/nar/gkv007.
52. Zerbino, D.R., Wilder, S.P., Johnson, N., Juettemann, T., and Flicek, P.R. (2015). The Ensembl Regulatory Build. Genome Biol. *16*, 56. https://doi.org/10.1186/s13059-015-0621-5.
53. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation, 2(3), 100141. https://doi.org/10.1016/j.xinn.2021.100141
54. Wu, Y., Borde, M., Heissmeyer, V., Feuerer, M., Lapan, A.D., Stroud, J.C., Bates, D.L., Guo, L., Han, A., Ziegler, S.F., et al. (2006). FOXP3 Controls Regulatory T Cell Function through Cooperation with NFAT. Cell *126*, 375–387. https://doi.org/10.1016/j.cell.2006.05.042.
55. Zheng, Y., Josefowicz, S.Z., Kas, A., Chu, T.-T., Gavin, M.A., and Rudensky, A.Y. (2007). Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells. Nature *445*, 936–940. https://doi.org/10.1038/nature05563.

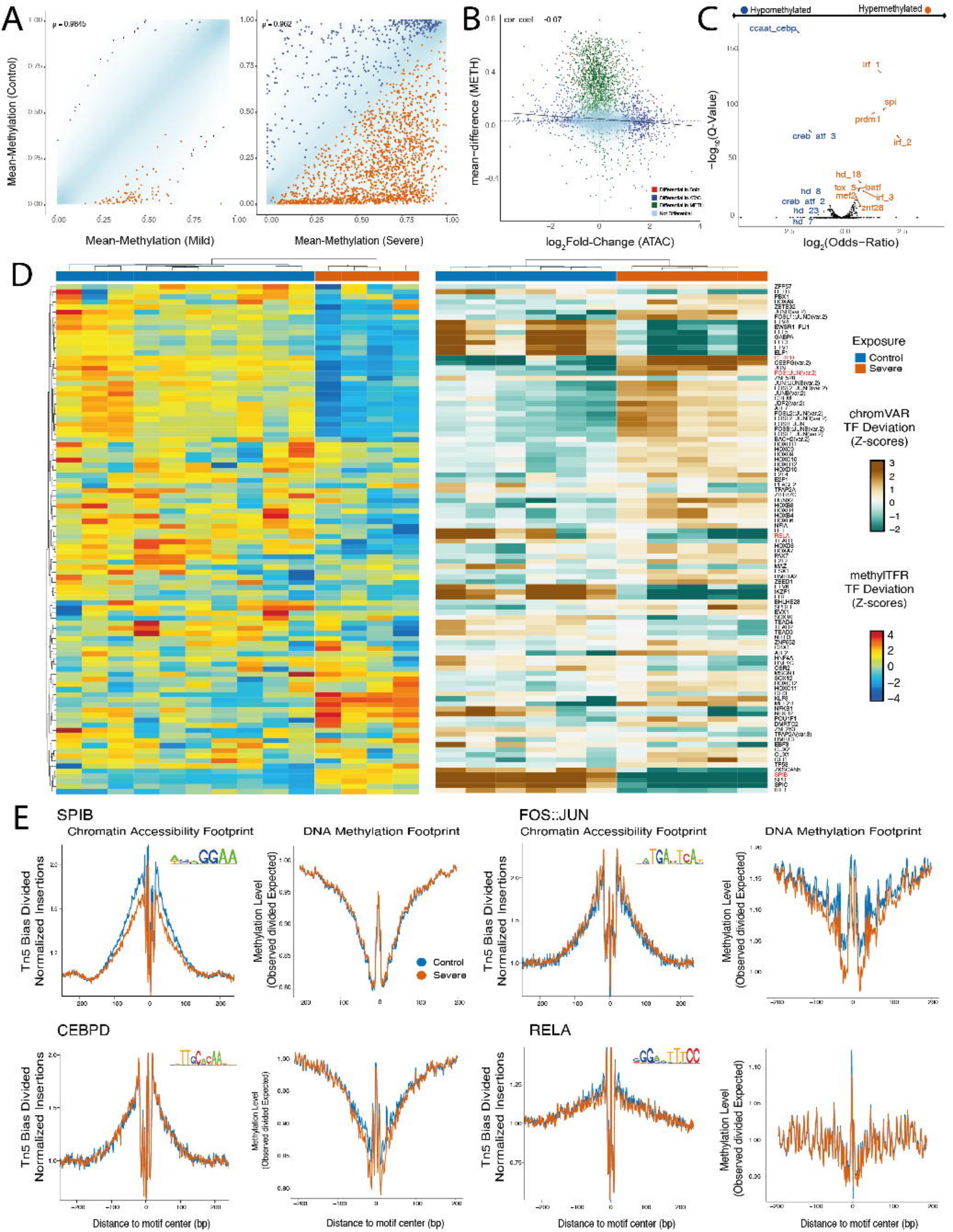


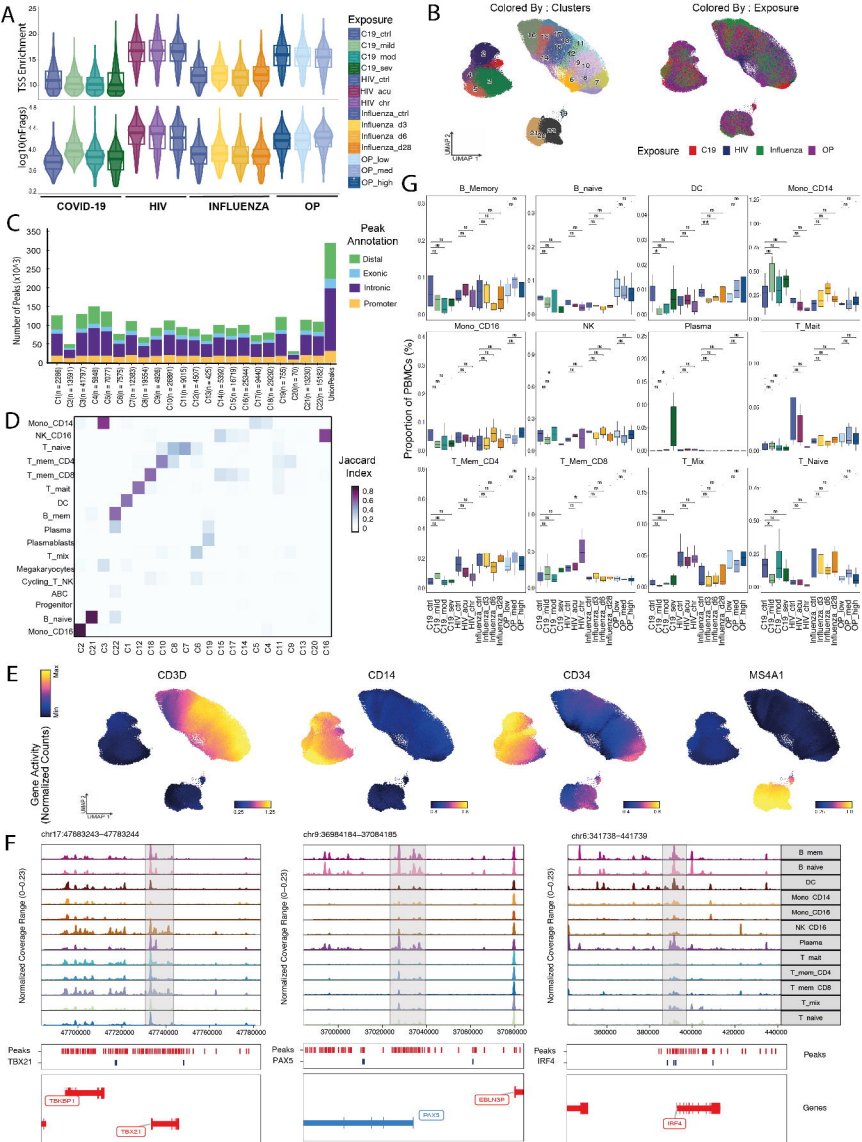
A
COVID-19
Control
Mild
Moderate
Severe
25 Samples
Influenza
Control
Day 3
Day 6
Day 28
20 Samples
HIV
Control
Chronic
Acute
12 Samples
Organophosphate
Low
Medium
High
35 Samples
PBMC
scATAC-seq
B
CD16 Monocytes
DC
CD14 Monocytes
NK
CD8T Memory
T-Mait
CD4T Memory
T-Mix
T-Naive
Plasma
B-Naive
B-Memory
UMAP 2
UMAP 1
C
Gene Activity (Normalized Counts)
Min
Max
TF Motif Activity (chromVAR deviations)
Min
Max
IRF8
GATA3
TBX21
PAX5
0.2
1.2
-2.5
2.5
-1.0
1.0
0.1
1.5
-1.2
1.5
0.3
1.5
-1.0
4.0
D
chr16:85848802–85948803
chr10:8003603-8103604
Normalized Coverage (Range 0–0.23)
Normalized Coverage (Range 0–0.19)
85850000
85875000
85900000
85925000
8025000
8050000
8075000
8100000
Peaks
IRF8
GATA3
MIR6774
GATA3
GATA3-AS1
E
ebox cagatgg
prdm16
nfkb 3
nfkb 1
mef2
zfn121
nr_19
hand1
nr_16
nr_3
nr_12
nr_11
nfi_3
p53_like_1
znf549
rel_halfsite
znf528
znf41
nfi 2
rbpj
tead
nr 7
nr 1
maf
insm1
tfap2 2
znf423
znf332
ddit3cebpa
sox_1
srf
hd_15
znf257
ewsr1_fli1
pax_1
snai2
srebf1
lin54
cpeb1
e2f 4
hif
hd 17
gata
cux 2
irf 4
nr 6
zbed1
klf_sp_1
myb_5
nfat_2
fox_5
znf384_2
nrf1
yy1
ets_1
runx_1
tbx_4
TF deviation Z-Scores (chromVAR)
4
2
0
−2
−4
F
Cell-Type Proportions
1.00
0.75
0.50
0.25
0.00
Cell Types
B-Memory
B-Naive
DC
CD14 Mono
CD16 Mono
NK
Plasma
T-Mait
CD4T-Memory
CD8T-Memory
T-Mix
T-Naive
C19_ctrl
C19_mild
C19_mod
C19_sev
HIV_ctrl
HIV_acu
HIV_chr
Influenza_ctrl
Influenza_d3
Influenza_d6
Influenza_d28
OP_low
OP_med
OP_high
COVID-19
HIV
INFLUENZA
OP


A
Colored by Subjects
Colored by Clusters
Tex
Subjects
Subject 1
Subject 2
Subject 3
Subject 4
Clusters
C1
C2
C3
C4
C5
C6
B
Gene Activity (Normalized Counts)
Min
Max
CTLA4
HAVCR2
C
Tex
NCAM1
CTLA4
ROBO2
ROBO1
FOXP2
HAVCR2
FOXP4
Column Z-Scores
16769 features
Gene Score Matrix
D
CD4T cells proportions (time points/pre)
Exhausted CD8T cells proportions (time points/pre)
Pre
Acute
Chronic
E
Viral Load (copies/mL)
Subject 1
Subject 2
Subject 3
Subject 4
F
Number of features = 131798
Number Up-Regulated = 468 (0.36 Percent)
Number Down-Regulated = 5285 (4.01 Percent)
Not-Differential
Hyper-accessible in Tex
Hypo-accessible in Tex
−log10(P-adj)
log2(Fold-Change)
G
Hyper-accessible in Tex
FOXD4
FOXJ1
FOXP2
FOXD4L1
FOXK2
ZBTB7A
WT1
SP2
DNMT1
SP1
SMAD5
ZFX
EGR1
ZNF148
−log10(P-adj) Motif Enrichment
Rank Sorted TFs Enriched
mlog10Padj
H
chr2:203817785-203917786
chr7:114036326-114136327
Coverage (Norm. ATAC Signal Range (0-0.07) by ReadsinTSS)
Coverage (Norm. ATAC Signal Range (0-0.03) by ReadsinTSS)
Tex
C1
C3
C4
C5
C6
FOXP2 absent
FOXP2 present
Peaks
Genes
CTLA4
FOXP2
PPP1R3A

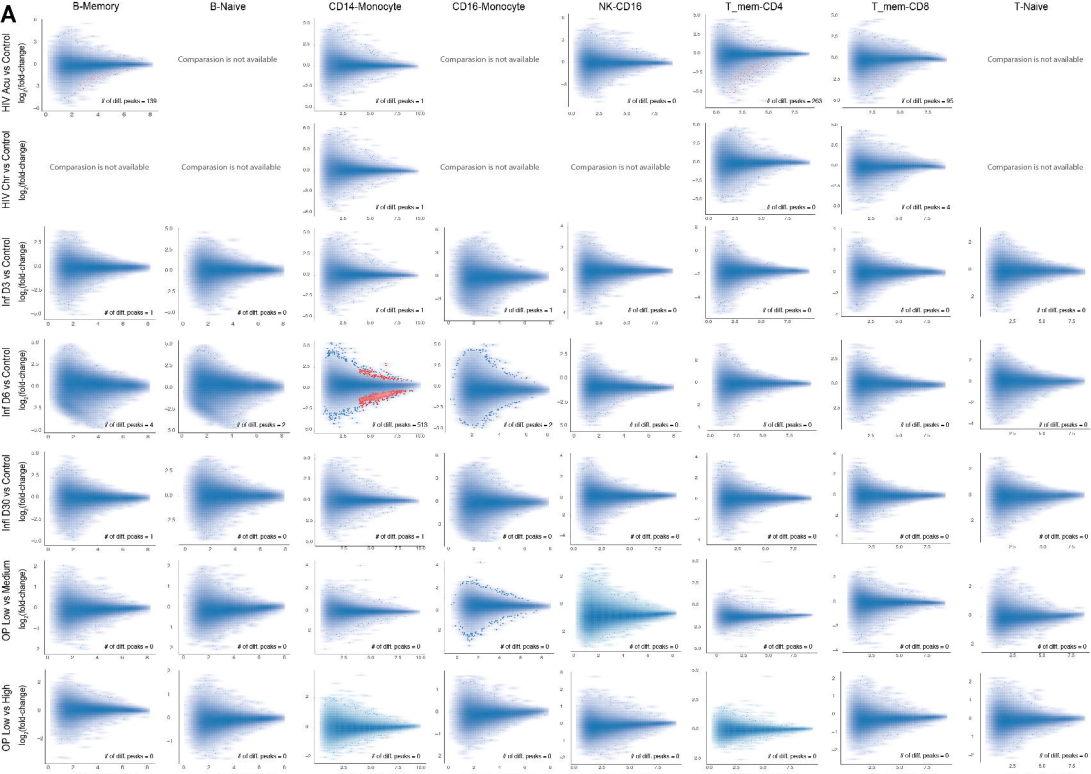

A
COVID-19 Moderate vs Control
COVID-19 Severe vs Control
log2(fold-change)
log2(Mean-Counts)
# of diff. peaks = 373
# of diff. peaks = 923
B
cor. coef. = 0.43
log2Fold-Change (Moderate)
log2Fold-Change (Severe)
Differential in Both
Differential in Sev
Differential in Mod
Not Differential
C
923 Differentiality Accessible Peaks (Severe)
LOSS
GAIN
Accessibility Z-Score
Exposure
Control
Mild
Moderate
Severe
D
hypoaccessible
hyperaccessible
−log10(Q-Value)
log2(Odds-Ratio)
ccaat_cebp
ap1_1
creb_atf_3
spi
nfkb_3
nfkb_1
irf_2
rel_halfsite
ets_1
mzf1
nr_7
nfkb_2
ap1_2
pou_3
pou_2
hd_10
hd_3
fox_6
E
cor. coef. = 0.3
log2(Fold-change) Gene expression
log2(Fold-change) Gene activity
F
chr2:112761902−112911903
chr7:22650883−22800884
Normalized Coverage Range (0–0.55)
Normalized Coverage Range (0–0.2)
Control
Mild
Moderate
Severe
112800000 112840000 112880000
22700000 22750000 22800000
Peaks
Diff. Peaks
Genes
IL1A
CKAP2L
IL1B
IL8
LOC401312
LOC541472

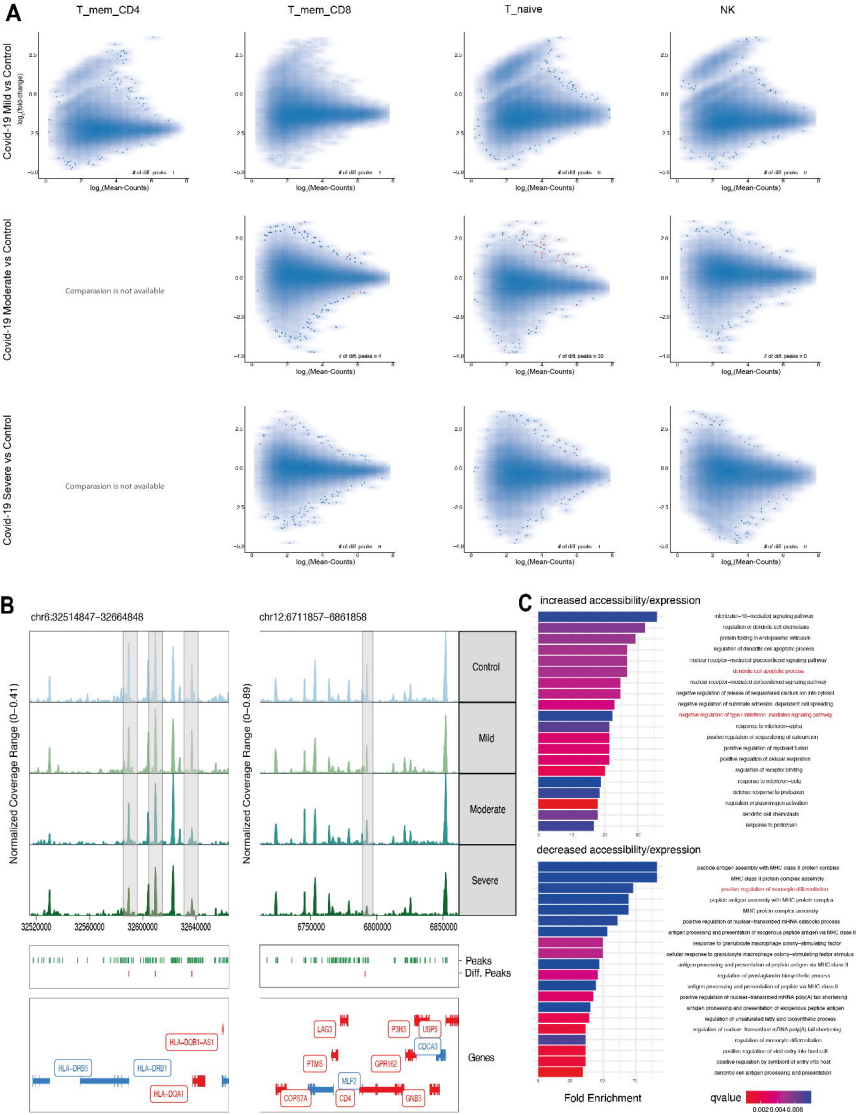

A
Chromatin Accessibility
(ATAC 110,376 cells)
DNA Methylation
(snmC 22,488 cells)
CD16 NK
CD8T Memory
CD16 Monocytes
DC
T-Mait
CD4T Memory
T-Mix
T-Naive
CD14 Monocytes
Plasma
B-Naive
B-Memory
B-cell
Tc-Mem
Th-Mem
Th-Naive
NK-cell
Tc-Naive
Monocyte
UMAP 2
UMAP 1
Mapped Cell-Type (from snmC)
B-cell
NK-cell
Tc-Mem
Monocyte
Th-Mem
Tc-Naive
Th-Naive
Mapped Cell-Type (ATAC)
B_mem
Mono_CD14
Plasma
T mem CD8
B_naive
Mono_CD16
T_mait
T mix
DC
NK
T mem CD4
T_naive
B
snmC
ATAC
B mem
B_naive
Plasma
Mono CD14
DC
Mono CD16
NK_CD16
T mait
T mem CD8
T_mem_CD4
T mix
T_naive
B-cell
Monocyte
NK-cell
Tc-Mem
Th-Mem
Tc-Naive
Th-Naive
Mapping Proportion
0
0.5
1
ATAC <- METH (row scaled)
METH <- ATAC (column scaled)
C
methylTFR TF Deviation (Z-scores)
Cell Types (ATAC)
Mono_CD14
Mono_CD16
DC
T_mem_CD4
T_mem_CD8
T_naive
T_mix
T_mait
NK
B_mem
B_naive
Plasma
chromVAR TF Deviation (Z-scores)
Cell Types (snmC)
B-cell
Monocyte
NK-cell
Tc-Mem
Tc-Naive
Th-Mem
Th-Naive
Pseudobulk Correlation Scores (Row-wise)
methyl-SELEX Label
MethylPlus
MethylMinus
Little effect
Inconclusive
D
Cor.Coef: -0.28
mTFR Z-score difference (Th-Naive - Th Memory)
cVAR Z-score difference (T-Naive - CD4T Memory)
Both Diff
cVAR Diff
mTFR Diff
Not Diff
E
FOS
Chromatin Accessibility Footprint
DNA Methylation Footprint
Tn5 Bias Divided Normalized Insertions
Methylation Level (Observed divided Expected)
Distance to motif center (bp)
T_mem_CD8
T_mem_CD4
T_naive
Tc-Mem
Tc-Naive
Th-Mem
Th-Naive
F
IRF4
Chromatin Accessibility Footprint
DNA Methylation Footprint
Tn5 Bias Divided Normalized Insertions
Methylation Level (Observed divided Expected)
Distance to motif center (bp)

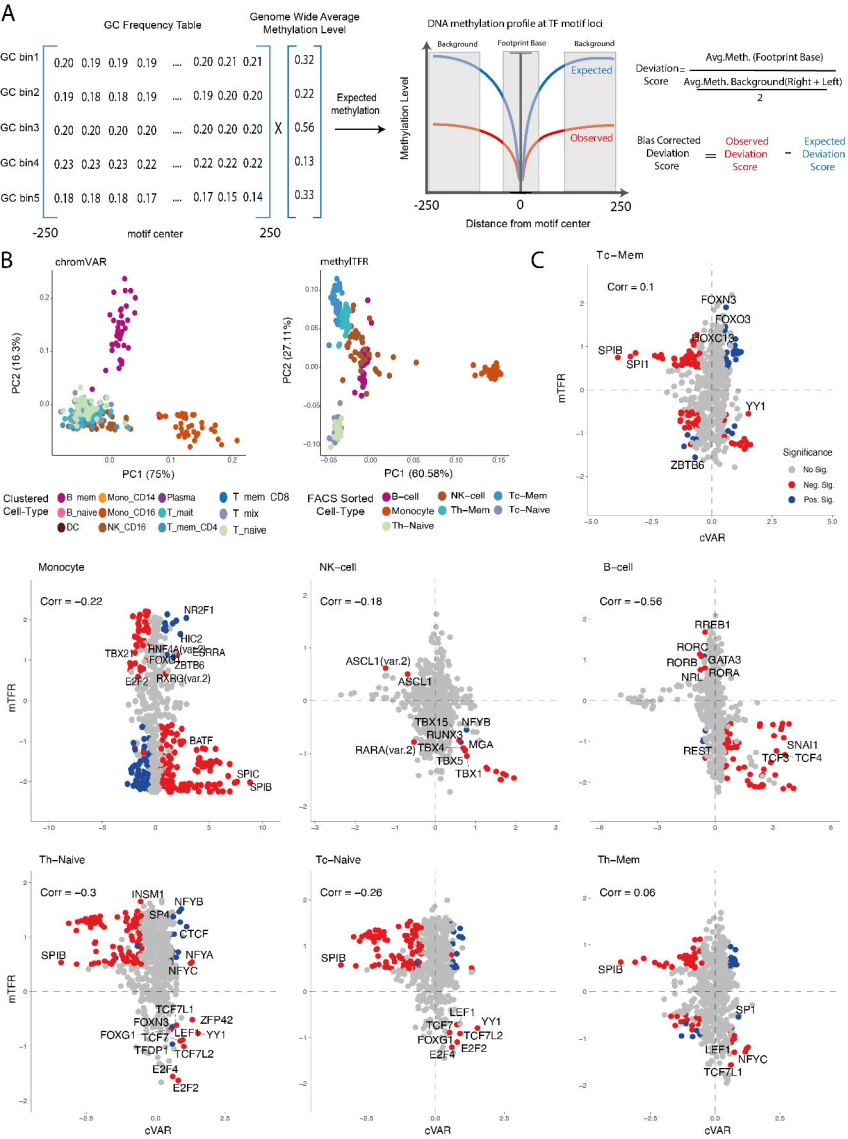

A
Mean-Methylation (Control)
Mean-Methylation (Mild)
Mean-Methylation (Severe)
B
mean−difference (METH)
log2Fold-Change (ATAC)
C
Hypomethylated
Hypermethylated
−log10(Q-Value)
log2(Odds−Ratio)
ccaat_cebp
irf_1
spi
prdm1
creb_atf_3
irf_2
hd_18
fox_5
batf
mef2
irf_3
znf28
hd_8
creb_atf_2
hd_23
hd_7
D
Exposure
Control
Severe
chromVAR
TF Deviation
(Z-scores)
methylTFR
TF Deviation
(Z-scores)
E
SPIB
Chromatin Accessibility Footprint
DNA Methylation Footprint
FOS::JUN
Chromatin Accessibility Footprint
DNA Methylation Footprint
Tn5 Bias Divided
Normalized Insertions
Methylation Level
(Observed divided Expected)
Control
Severe
CEBPD
RELA
Distance to motif center (bp)

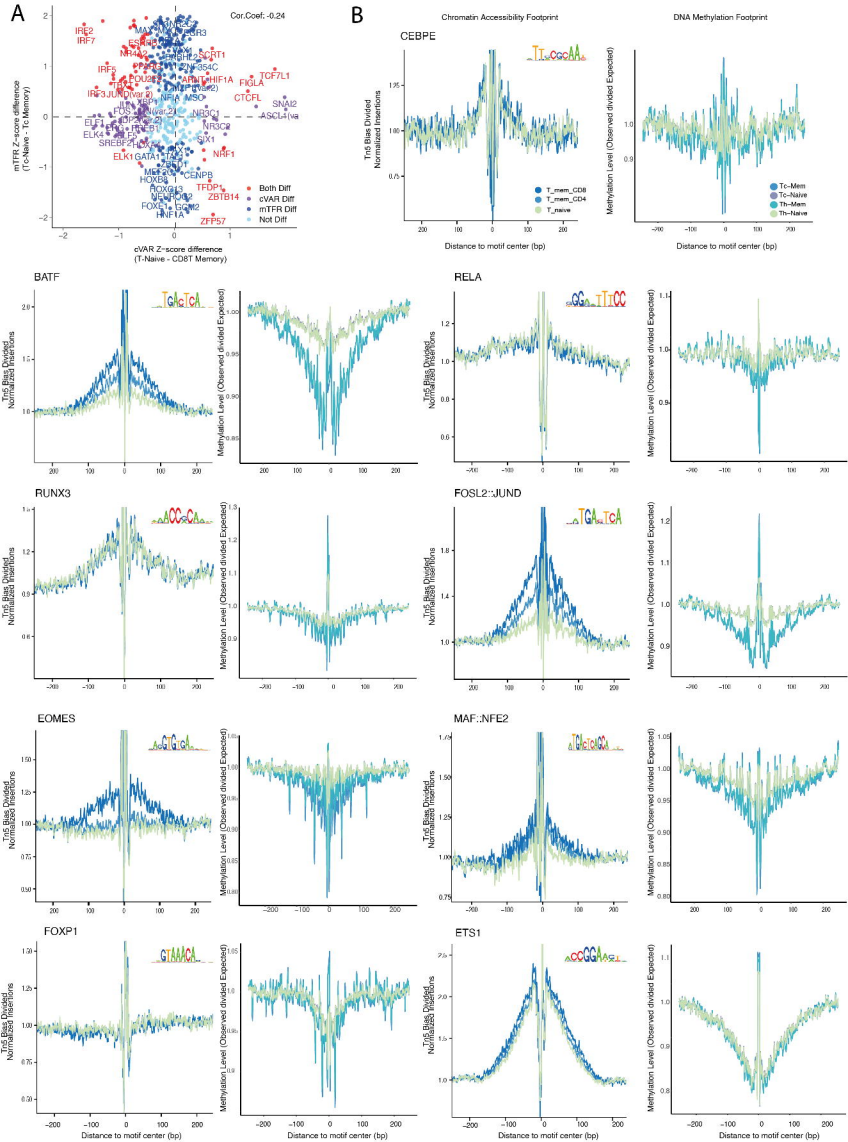


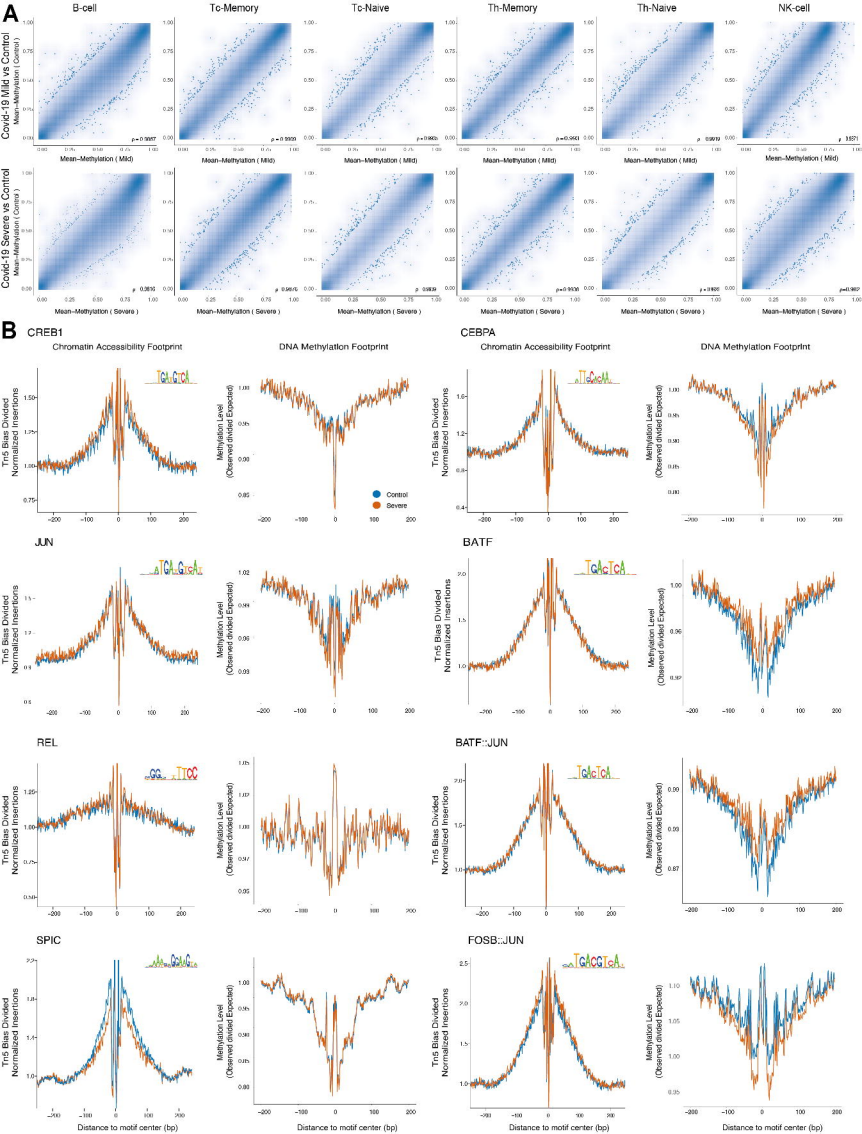
A
TSS Enrichment
log10(nFrags)
Exposure
C19_ctrl
C19_mild
C19_mod
C19_sev
HIV_ctrl
HIV_acu
HIV_chr
Influenza_ctrl
Influenza_d3
Influenza_d6
Influenza_d28
OP_low
OP_med
OP_high
COVID-19
HIV
INFLUENZA
OP
B
Colored By : Clusters
Colored By : Exposure
Exposure
C19
HIV
Influenza
OP
C
Number of Peaks (x10^3)
Peak Annotation
Distal
Exonic
Intronic
Promoter
D
Jaccard Index
Mono_CD14
NK_CD16
T_naive
T_mem_CD4
T_mem_CD8
T_mait
DC
B_mem
Plasma
Plasmablasts
T_mix
Megakaryocytes
Cycling_T_NK
ABC
Progenitor
B_naive
Mono_CD16
G
B_Memory
B_naive
DC
Mono_CD14
Mono_CD16
NK
Plasma
T_Mait
T_Mem_CD4
T_Mem_CD8
T_Mix
T_Naive
Proportion of PBMCs (%)
E
CD3D
CD14
CD34
MS4A1
Gene Activity (Normalized Counts)
F
chr17:47683243–47783244
chr9:36984184–37084185
chr6:341738–441739
Normalized Coverage Range (0–0.23)
B mem
B naive
DC
Mono_CD14
Mono_CD16
NK_CD16
Plasma
T mait
T_mem_CD4
T mem CD8
T_mix
T naive
Peaks
TBX21
PAX5
IRF4
Genes


A
B-Memory
B-Naive
CD14-Monocyte
CD16-Monocyte
NK-CD16
T_mem-CD4
T_mem-CD8
T-Naive
HIV Acu vs Control
HIV Chr vs Control
Inf D3 vs Control
Inf D6 vs Control
Inf D30 vs Control
OP Low vs Medium
OP Low vs High
log2(fold-change)
log2(Mean-Counts)
Comparasion is not available
# of diff. peaks = 139
# of diff. peaks = 1
# of diff. peaks = 0
# of diff. peaks = 263
# of diff. peaks = 95
# of diff. peaks = 4
# of diff. peaks = 2
# of diff. peaks = 513


A
T_mem_CD4
T_mem_CD8
T_naive
NK
Covid-19 Mild vs Control
Covid-19 Moderate vs Control
Covid-19 Severe vs Control
log2(fold-change)
log2(Mean-Counts)
Comparasion is not available
Comparasion is not available
B
chr6:32514847–32664848
chr12:6711857–6861858
Normalized Coverage Range (0–0.41)
Normalized Coverage Range (0–0.89)
Control
Mild
Moderate
Severe
32520000
32560000
32800000
32840000
6750000
6800000
6850000
Peaks
Diff. Peaks
HLA-DRB5
HLA-DRB1
HLA-DQA1
HLA-DQB1-AS1
LAG3
PTMS
COPS7A
CD4
MLF2
GPRH2
GNB3
CDCA3
USP5
Genes
C
increased accessibility/expression
decreased accessibility/expression
Fold Enrichment
qvalue
0.0020.0040.006


A
GC Frequency Table
Genome Wide Average Methylation Level
GC bin1 0.20 0.19 0.19 0.19 .... 0.20 0.21 0.21 0.32
GC bin2 0.19 0.18 0.18 0.19 .... 0.19 0.20 0.20 0.22
GC bin3 0.20 0.20 0.20 0.20 .... 0.20 0.20 0.20 0.56
GC bin4 0.23 0.23 0.23 0.22 .... 0.22 0.22 0.22 0.13
GC bin5 0.18 0.18 0.18 0.17 .... 0.17 0.15 0.14 0.33
-250
motif center
250
X
Expected methylation
DNA methylation profile at TF motif loci
Background
Footprint Base
Background
Expected
Observed
Methylation Level
-250
0
250
Distance from motif center
Deviation Score = Avg.Meth. (Footprint Base) / Avg.Meth. Background(Right + Left) / 2
Bias Corrected Deviation Score = Observed Deviation Score - Expected Deviation Score
B
chromVAR
PC2 (16.3%)
PC1 (75%)
methylTFR
PC2 (27.11%)
PC1 (60.58%)
Clustered Cell-Type
B mem
B_naive
DC
Mono_CD14
Mono_CD16
NK_CD16
Plasma
T_mait
T_mem_CD4
T mem CD8
T mix
T_naive
FACS Sorted Cell-Type
B-cell
Monocyte
Th-Naive
NK-cell
Th-Mem
Tc-Mem
Tc-Naive
C
Tc-Mem
Corr = 0.1
FOXN3
FOXO3
HOXC13
SPIB
SPI1
YY1
ZBTB6
mTFR
cVAR
Significance
No Sig.
Neg. Sig.
Pos. Sig.
Monocyte
Corr = −0.22
NR2F1
HIC2
TBX21
HNF4A(var.2)
ESRRA
FOXG1
ZBTB6
E2F2
RXRG(var.2)
BATF
SPIC
SPIB
NK-cell
Corr = −0.18
ASCL1(var.2)
ASCL1
TBX15
NFYB
RUNX3
RARA(var.2)
TBX4
MGA
TBX5
TBX1
B-cell
Corr = −0.56
RREB1
RORC
RORB
GATA3
NRL
RORA
REST
SNAI1
TCF3
TCF4
Th-Naive
Corr = −0.3
INSM1
NFYB
SP4
CTCF
SPIB
NFYA
NFYC
TCF7L1
FOXN3
ZFP42
FOXG1
TCF7
LEF1
YY1
TFDP1
TCF7L2
E2F4
E2F2
Tc-Naive
Corr = −0.26
SPIB
LEF1
TCF7
YY1
FOXG1
TCF7L2
E2F4
E2F2
Th-Mem
Corr = 0.06
SPIB
SP1
LEF1
NFYC
TCF7L1


A
Cor.Coef: -0.24
mTFR Z-score difference
(T-Naive - Th-Memory)
cVAR Z-score difference
(T-Naive - CD8T Memory)
Both Diff
cVAR Diff
mTFR Diff
Not Diff
B
Chromatin Accessibility Footprint
DNA Methylation Footprint
CEBPE
T_mem_CD8
T_mem_CD4
T_naive
Tc-Mem
Tc-Naive
Th-Mem
Th-Naive
Tn5 Bias Divided Normalized Insertions
Methylation Level (Observed divided Expected)
Distance to motif center (bp)
BATF
RELA
RUNX3
FOSL2::JUND
EOMES
MAF::NFE2
FOXP1
ETS1


A
B-cell
Tc-Memory
Tc-Naive
Th-Memory
Th-Naive
NK-cell
Covid-19 Mild vs Control
Covid-19 Severe vs Control
Mean-Methylation (Control)
Mean-Methylation (Mild)
Mean-Methylation (Severe)
B
CREB1
CEBPA
JUN
BATF
REL
BATF::JUN
SPIC
FOSB::JUN
Chromatin Accessibility Footprint
DNA Methylation Footprint
Tn5 Bias Divided Normalized Insertions
Methylation Level (Observed divided Expected)
Control
Severe
Distance to motif center (bp)