bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Integrated single cell analysis of chromatin accessibility and cell surface markers

Elliott Swanson[1], Cara Lord[1,2], Julian Reading[1], Alexander T. Heubeck[1], Adam K. Savage[1], Richard Green[1], Troy R. Torgerson[1,3], Thomas F. Bumol[1], Lucas T. Graybuck[1], Peter J. Skene[1*]

1. Allen Institute for Immunology, Seattle, WA, USA.
2. Current affiliation: GlaxoSmithKline, Collegeville, PA, USA.
3. Department of Pediatrics, University of Washington, Seattle, WA, USA.

* Corresponding Author

Abstract

Profiling of open chromatin regions at the single cell level using ATAC-seq (scATAC-seq) has been instrumental in understanding the heterogeneous usage of transcription factors that drive differentiation, cellular responses to extracellular signals, and human disease states. The large size of the human genome and processing artefacts resulting in DNA damage are an inherent source of background in scATAC-seq. Furthermore, the downstream analysis of scATAC-seq to derive meaningful biological information is complicated by the lack of clear phenotypic information on each analyzed cell to allow an association between chromatin state and cell type. Using the heterogeneous mixture of cells in human peripheral blood as a test case, we developed a novel scATAC-seq workflow that increases signal-to-noise ratio and allows simultaneous measurement of cell surface markers: Integrated Cellular Indexing of Chromatin Landscape and Epitopes (ICICLE-seq). Combining cell surface marker barcoding with high quality scATAC-seq offers a novel tool to identify type-specific regulatory regions based on phenotypically defined cell types.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Main

Peripheral blood mononuclear cells (PBMCs) purified using gradient centrifugation are a major source of clinically relevant cells for the study of human immune health and disease[1]. Like most other human tissues, PBMCs are a complex, heterogeneous mixture of cell types derived from common stem cell progenitors[2]. Despite the genome being mostly invariant between different PBMC cell types, each immune cell type performs an important and distinct function. Understanding the genomic regulatory landscape that controls lineage specification, cellular maturation, activation state, and functional diversity in response to intra- and extracellular signals is key to understanding the immune system in both health and disease[3–5].

Recent improvements in single-cell genomic methods have enabled profiling of the regulatory chromatin landscape of complex cell type mixtures. In particular, droplet-based single-cell assays for transposase-accessible chromatin (scATAC-seq and dscATAC-seq) allow profiling of open chromatin at single-cell resolution[6,7]. Promising new methods have combined scATAC-seq with simultaneous measurement of nuclear mRNAs (e.g. sci-CAR[8], SNARE-seq[9], SHARE-seq[10]). However, identification of highly specified functional immune cell types is hampered by current computational labeling and label transfer methods, in part due to complexity in linking regulatory sites to gene expression. To overcome these limitations, we systematically tested whole cell and nuclear purification and preparation methods for PBMCs. We found that intact, permeabilized cells perform extremely well for scATAC-seq, exceeding conventional scATAC-seq on nuclei by some measures (**Fig. 1b**). This insight enables a new protocol analogous to Cellular Indexing of Transcriptomes and Epitopes (CITE-seq[11]) to measure both surface protein abundance and chromatin accessibility: Integrated Cellular Indexing of Chromatin Landscape and Epitopes (ICICLE-seq, **Fig. 1a and 3**).

Our initial scATAC-seq experiments followed the protocol described by 10x Genomics, which largely adhered to the Omni-ATAC workflow[12]. This protocol utilizes a combination of hypotonic lysis, detergents, and a saponin to isolate nuclei without releasing mitochondrial DNA. After performing this assay, sequencing, and tabulating data quality metrics (**Methods**), we identified two major populations of cell barcodes (**Fig. 1b; left panel**): (1) A large number of barcodes, shown in gray, that have a low number of unique fragments and a low Fraction of Reads in Peaks (FRIP). These barcodes contain little useful information but consume 80% of total sequenced reads (**Fig. 1e**, non-cell barcodes) at a sequencing depth of 200 million reads per library (20,000 reads per expected barcode); (2) Barcodes with higher quality as measured by FRIP (red points) that contain enough information to attempt downstream analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

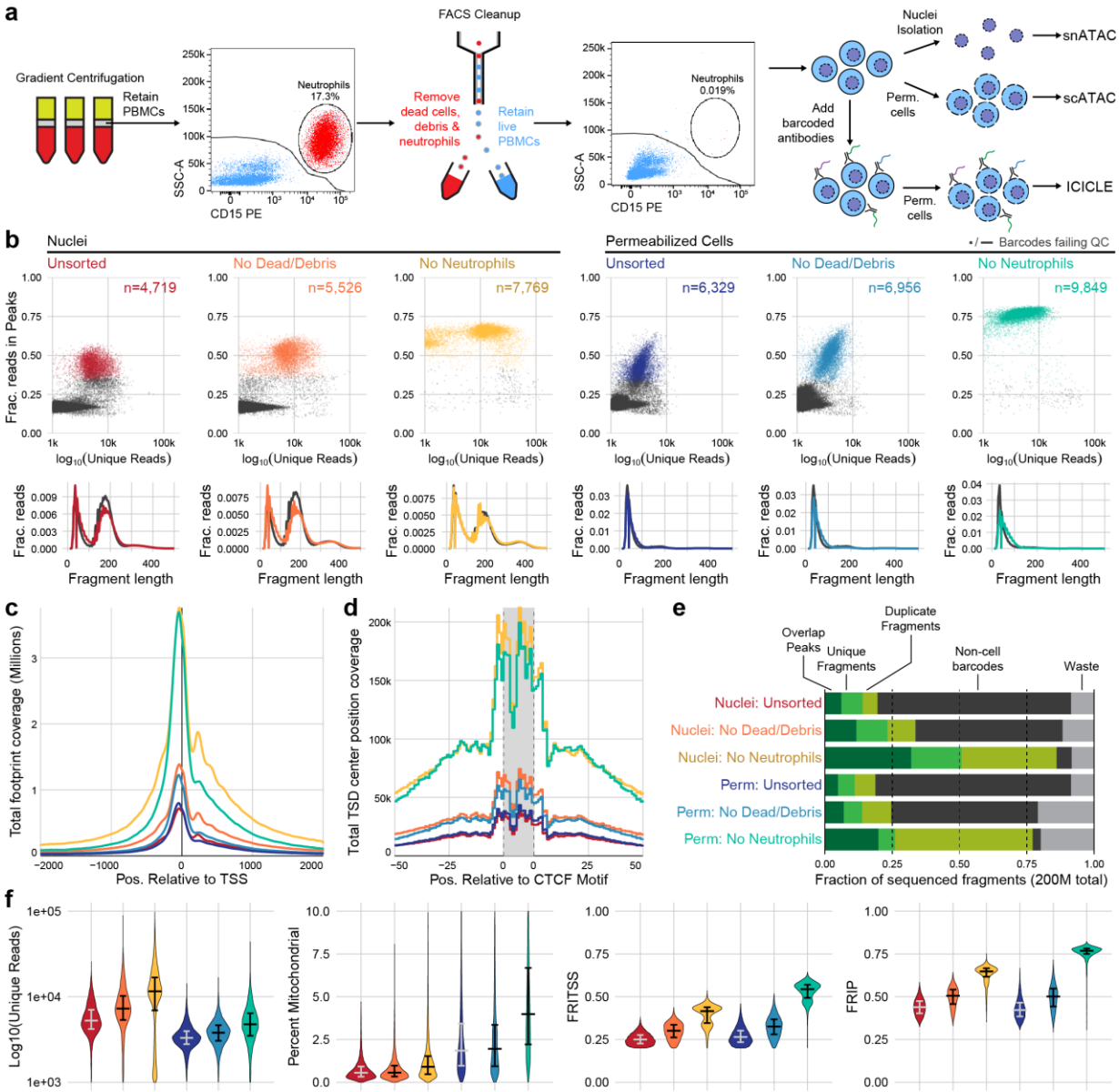


Figure 1. a, Schematic overview of major steps in snATAC, scATAC, and ICICLE-seq methods. **b**, Comparison of quality control characteristics of ATAC-seq libraries generated from nuclei isolation and permeabilized cells, with and without FACS. Top panels show signal-to-noise as assessed by fraction of reads in peaks (FRIP) on the y-axis, and quantity of unique fragments per cell barcode on the x-axis. Lower panels display fragment length distributions obtained from paired-end sequencing of ATAC libraries. Colored lines represent barcodes that pass QC filters; gray lines represent barcodes failing QC (non-cell barcodes). All libraries were equally downsampled to 200M total sequenced reads for comparison. Colors in b are re-used in remaining panels. **c**, Total coverage of Tn5 footprints summed across all transcription start sites (TSS). Tn5 footprints are 29 bp regions comprising the 9 bp target-site duplication (TSD) and 10 bp on either side, which represent accessible chromatin for each transposition event. **d**, Total coverage of TSD centers summed over a set of 100,000 genomic CTCF motifs found in previously published DNase hypersensitive sites[13]. TSD centers are obtained by shifting +4 and -5 bp from the 5' and 3' ends of uniquely aligned fragments, respectively. **e**, Barplot representations of the fraction of total aligned reads in various QC categories. Fragments overlapping a previously published peak set for PBMC dscATAC-seq[7] are in the 'Overlap Peaks' category. Unique fragments are the remaining uniquely aligned fragments that do not overlap peak regions. 'Waste' reads were not aligned, or were assigned to cell barcodes with fewer than 1,000 total reads. **f**, Violin plots showing distributions of QC metrics. Median (wide bar) and 25th and 75th quantiles (whiskers and narrow bars) are overlaid on violin plots. Median values are also in **Table 1**. Note that the y-axis of the first panel is on a logarithmic scale; remaining panels are linear.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The loss of 80% of sequenced reads to non-cell barcodes is costly. Previous studies of scRNA-seq data have shown that cellular lysis can release ambient RNA that increases abundance of low-quality barcodes and contaminates droplets, yielding barcodes with both cellular and ambient RNAs that reduces accuracy of the transcriptional readout[14]. We reasoned that nuclear isolation protocols may cause the release of ambient DNA, causing a similar effect in scATAC-seq datasets. Optimization of nuclear lysis protocols, especially changing to less stringent detergents, provided increased FRIP and decreased non-cell barcodes (**Supplementary Fig. 1, 2a**). Hypotonic lysis conditions used in these protocols may also be a biophysical stressor to the native chromatin state, as previously observed[15]. To reduce perturbation of chromatin, we performed cell membrane permeabilization under isotonic conditions to allow access to the nuclear DNA without isolating nuclei through hypotonic lysis. The saponin digitonin was used to cause concentration-dependent selective permeabilization of cholesterol-containing membranes while leaving inner mitochondrial membranes intact to and prevent high levels of Tn5 transposition in mitochondrial DNA[16,17]. Digitonin has previously been used for ATAC-seq assays under hypotonic conditions in Fast-ATAC[18] and plate-based scATAC-seq[19] protocols. Permeabilization of intact cells under isotonic conditions greatly reduced the amount of non-cell barcodes and their contribution to sequencing libraries (**Supplementary Fig. 2b**).

We also observed that PBMCs purified by leukapheresis rather than Ficoll gradient centrifugation had consistently higher FRIP scores and fewer non-cell barcodes (**Supplementary Fig. 2c**). A major difference between our Ficoll-purified PBMCs and these leukapheresis-purified PBMCs was the presence of residual neutrophils in our Ficoll-purified samples. We tested removal of dead cells and debris with and without removal of neutrophils using Fluorescent Activated Cell Sorting (FACS) from PBMC samples with high neutrophil content (**Fig. 1a, Supplementary Fig. 3**). When applied to either nuclei (**Fig. 1b**, left panels) or permeabilized cells (**Fig. 1b**, right panels), there was a large increase in FRIP and reduction in non-cell barcodes in our scATAC-seq libraries (**Fig. 1e**). Removal of neutrophils did not have an adverse effect on leukapheresis-purified PBMCs (**Supplementary Fig. 1c**, right panel), and depletion using anti-CD15 magnetic beads also improved data quality (**Supplementary Fig. 2d and Supplementary Fig. 4**), though not to the same extent as FACS-based depletion.

We assessed the quantitative and qualitative differences between nuclei and permeabilized cell protocols with and without sorting by performing both protocols on a single set of input cells. Permeabilized cells yielded many more high-quality cell barcodes than nuclear preps using equal loading of cells or nuclei (15,000 loaded, expected 10,000 captured, **Table 1**). scATAC-seq libraries obtained from nuclei had many more reads with fragments originating from nucleosomal DNA fragments (**Fig. 1b**, lower panels), and non-cell barcodes from nuclei (gray lines) contained more of these fragments than cell barcodes. Thus, an overabundance of mononucleosomal fragments may indicate non-cell fragment contamination.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Libraries from permeabilized cells consisted almost entirely of short fragments, suggesting that permeabilization under isotonic conditions did not loosen or release native chromatin structure at the time of tagmentation (**Fig. 1b**, lower panels). Previous bulk ATAC-seq studies have shown that differing nuclear isolation protocols lead to varying amounts of mononucleosomal fragments[20]. In agreement with *in vitro* experiments studying the effects of low salt on nucleosomal arrays[21], this further suggests that hypotonic lysis leads to alteration of chromatin structure, raising the possibility of artifactual measurements of accessibility in nuclei-based ATAC-seq. To assess the effect that this difference has on the data obtained by each method, we overlaid Tn5 footprints near transcription start sites (TSS, **Fig. 1c**) and CTCF transcription factor binding sites (TFBS, **Fig. 1d**). The signal at TSS was retained in permeabilized cells, but positions flanking the TSS (occupied by neighboring nucleosomes) had reduced signal compared to isolated nuclei (examined in detail in **Supplementary Fig. 5**). At CTCF motifs, we observed nearly identical patterns of accessibility in both nuclei and permeabilized cells, suggesting that scATAC-seq signal at regulatory TFBS is retained in permeabilized cells. Overall, permeabilized intact cells obtained by FACS had the highest FRIP score, fewest non-cell barcodes, and greatest cell capture efficiency (**Table 1, Fig. 1e**).

Source Type	FACS Depletion	N Nuclei/Cells Pass QC	Median N Fragments	Median % Mitochondrial	Median N Unique	Median FRITSS	Median FRIP
Nuclei	Unsorted	4719	7344	0.6%	5247	0.249	0.439
Nuclei	Dead/Debris	5526	10526	0.6%	7284	0.301	0.505
Nuclei	Dead/Debris/Neut.	7769	19972	0.9%	11528	0.415	0.648
Perm. Cells	Unsorted	6329	5541	1.9%	3308	0.264	0.422
Perm. Cells	Dead/Debris	6956	6733.5	2.0%	3795.5	0.326	0.502
Perm. Cells	Dead/Debris/Neut.	9849	14069	4.0%	4756	0.544	0.768

Table 1. QC metrics summary for experiments displayed in **Fig. 1**. Median % Mitochondrial was calculated as a fraction of total fragments; FRITSS and FRIP were calculated as a fraction of unique fragments. Neut., neutrophils; FRITSS, fraction of reads in transcription start sites; FRIP, fraction of reads in peaks.

We next examined the effect of methodological differences on downstream biological analyses (**Fig. 2**). Removal of neutrophils greatly improved the ability to separate various cell types in UMAP projections of both nuclei and cells (**Fig. 2a-b**). To provide ground truth for label transfer, we performed flow cytometry on the same cells used for scATAC-seq, above. A panel of 25 antibodies was used to determine the proportion of each of the 12 cell types used to label the scATAC-seq cells in the PBMC sample (**Methods and Supplementary Fig. 6**). Label transfer was enabled by the ArchR package[22] to generate gene scores and perform transfer from a reference scRNA-seq dataset using the label transfer method provided in the Seurat package[23] (**Methods**). Using these tools, removal of neutrophils improved label

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

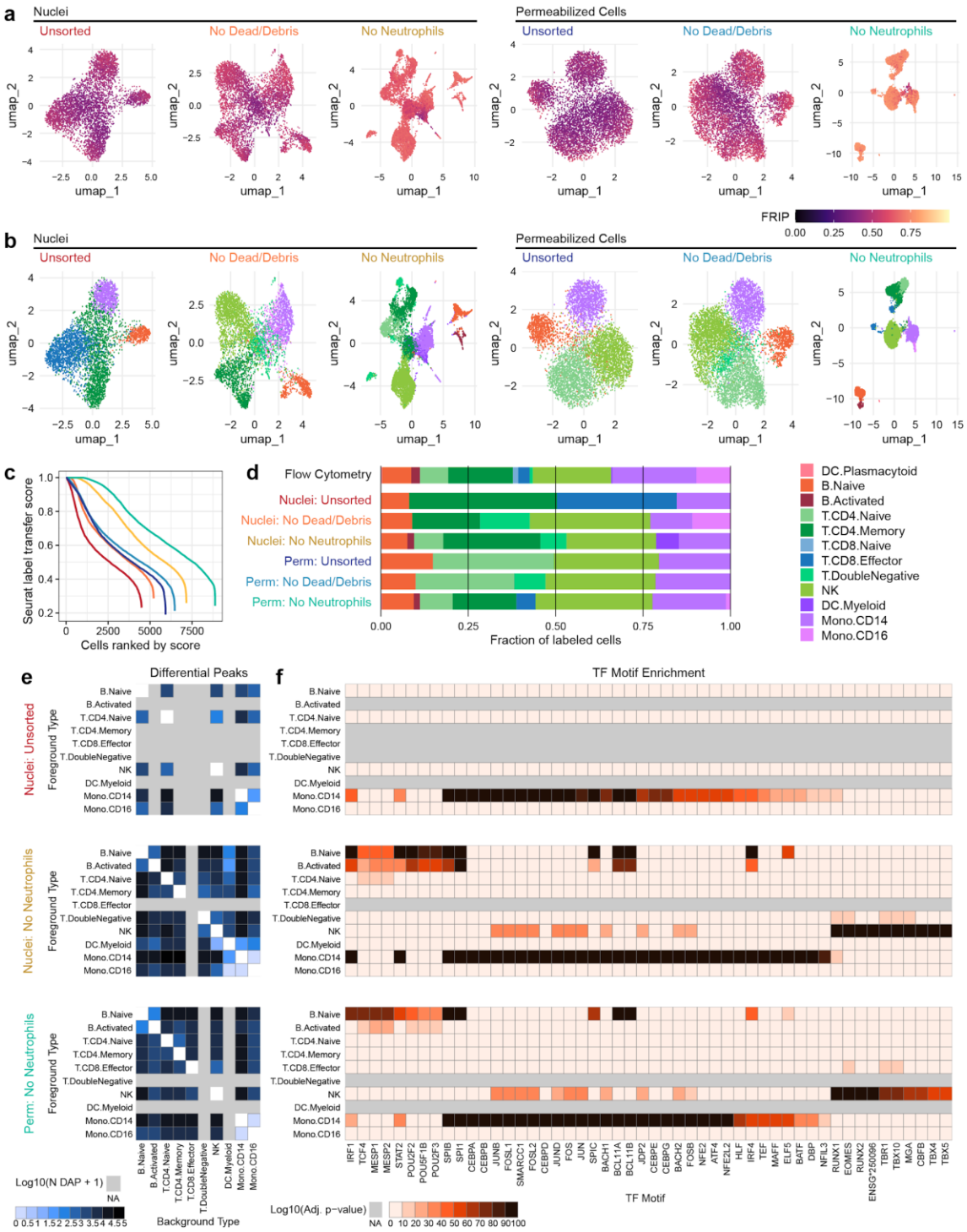


Figure 2. a, UMAP projection plots for corresponding datasets in Figure 1. Points are colored based on a common scale of fraction of reads in peaks (FRIP), bottom right. The number of cells in each panel are displayed in Fig. 1b. **b**, UMAP projection plots colored based on cell type obtained by label transfer from scRNA-seq (**Methods**). Colors for cell types are below, to the right. **c**, To visualize the number and quality of transferred labels, we ranked all cells based on the Seurat label transfer score obtained from label transfer results, and plotted lines through the score (y-axis) vs rank (x-axis) values. **d**, Bar plot showing the fraction of cells in each dataset that were assigned each cell type label. The top row shows cell type proportions for the same PBMC sample obtained by 25-color immunotyping flow cytometry (**Methods, Supplementary Table 1, Supplementary Figure 6**). Colors for cell types are to the right of the barplot. **e**, Heatmap plots of pairwise differentially accessible peaks (DAPs) between each pair of types, as computed with Wilcoxon tests using the ArchR package[22] (**Methods**). Values shown are the number of peaks more highly accessible in the foreground type (y-axis) compared to the background type (x-axis). Colors represent Log_{10}-scaled and binned counts of DAPs, as shown in the scale to the bottom-left. Gray regions represent cell types that were not

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

observed in a dataset. **f**, Heatmap plots of enriched transcription factor binding site (TFBS) motifs. Colors represent -1 x Log_{10}-scaled and binned adjusted p-values for tests of enrichment (**Methods**). Values > 100 are colored black.

transfer scores, and permeabilized cells yielded more cells with high label transfer scores than nuclei-based approaches (**Fig. 2b-c**). In addition, permeabilized cells provided labels most similar to the cell type proportions identified by flow cytometry (**Fig. 2d**), with identification of CD8 effector cells only observed in scATAC-seq with permeabilized cells. All methods yielded fewer CD16+ monocytes than observed by flow cytometry, suggesting that CD16+ monocytes may be lost during scATAC-seq using either nuclei or permeabilized cells, or that label transfer methods were not conducive to identifying this cell type (**Fig. 2d**). After labeling cell types, we used ArchR to call peaks for each cell type and perform pairwise tests of differential accessibility between each pair of cell types (**Fig. 2e**, left panels). We found many more differentially accessible sites in both cells and nuclei after removal of neutrophils. Differential accessibility was also used to identify differentially enriched TFBS motifs in each cell type (**Fig. 2e**, right panels). Without neutrophil removal (Nuclei Unsorted, top panel), we were unable to identify significantly enriched motifs in B cells and NK cells that were readily apparent in data from clean nuclei or permeabilized cells (bottom two panels). Together, these results demonstrate that neutrophil removal and the use of permeabilized cells allow for identification of specific cell types and TFBS motifs that are involved in regulation of gene expression.

Under standard scATAC-seq protocols, removal of the cell membrane severs the connection between the cell surface and the chromatin state of cells. By retaining the cell surface on permeabilized cells, we were able to extend scATAC-seq to simultaneously profile cell surface proteins and chromatin accessibility, which we term Integrated Cellular Indexing of Chromatin Landscapes and Epitopes (ICICLE-seq, **Fig. 3** and **Methods**). The ICICLE-seq protocol utilizes a custom Tn5 transposome complex with capture sequences compatible with the 10x Genomics 3' scRNA-seq gel bead capture reaction for simultaneous capture of ATAC fragments and polyadenylated antibody barcode sequences (**Supplementary Fig. 7**). Antibody-derived tags (ADTs) or ATAC-seq libraries could then be selectively amplified by PCR to generate separate libraries for sequencing (**Supplementary Fig. 7**). Due to the nature of fragment capture in this system, we obtain both a cell barcode and a single-end scATAC-seq read. We performed ICICLE-seq on a leukapheresis-purified PBMC sample, and were able to obtain 10,227 single cells with both scATAC-seq and ADT data from 3 capture wells that passed adjusted QC criteria: > 500 unique ATAC fragments (median = 761), FRIP > 0.65 (median = 0.725). Cells passing ATAC QC had a median of 3,871 ADT UMIs per cell (**Supplemental Fig. 8b**). UMAP projection and ATAC label transfer on ICICLE-seq data had resolution similar to scATAC-seq on intact permeabilized cells after dead cell and debris removal (**Fig. 3b**). We were able to leverage the additional ADT data to cluster and identify cell types based on their cell surface antigens at a much higher resolution (**Fig. 3d-f**). UMAP based on ADT data and Jaccard-Louvain clustering allowed identification of cell type-specific

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

clusters (**Fig. 3d**) based on clear association of cell type-specific markers with clusters (**Fig. 3e and Supplemental Fig. 8**). Once identified, we could leverage these cell type labels to identify differentially accessible peaks (DAPs) in the scATAC-seq data (**Fig. 3g**), even for types that were not separated based on label transfer from scRNA-seq (e.g. exhausted T cell subtypes). Thus, ICICLE-seq provides a novel platform for the identification of cell types in scATAC-seq data based on well-established cell surface markers.

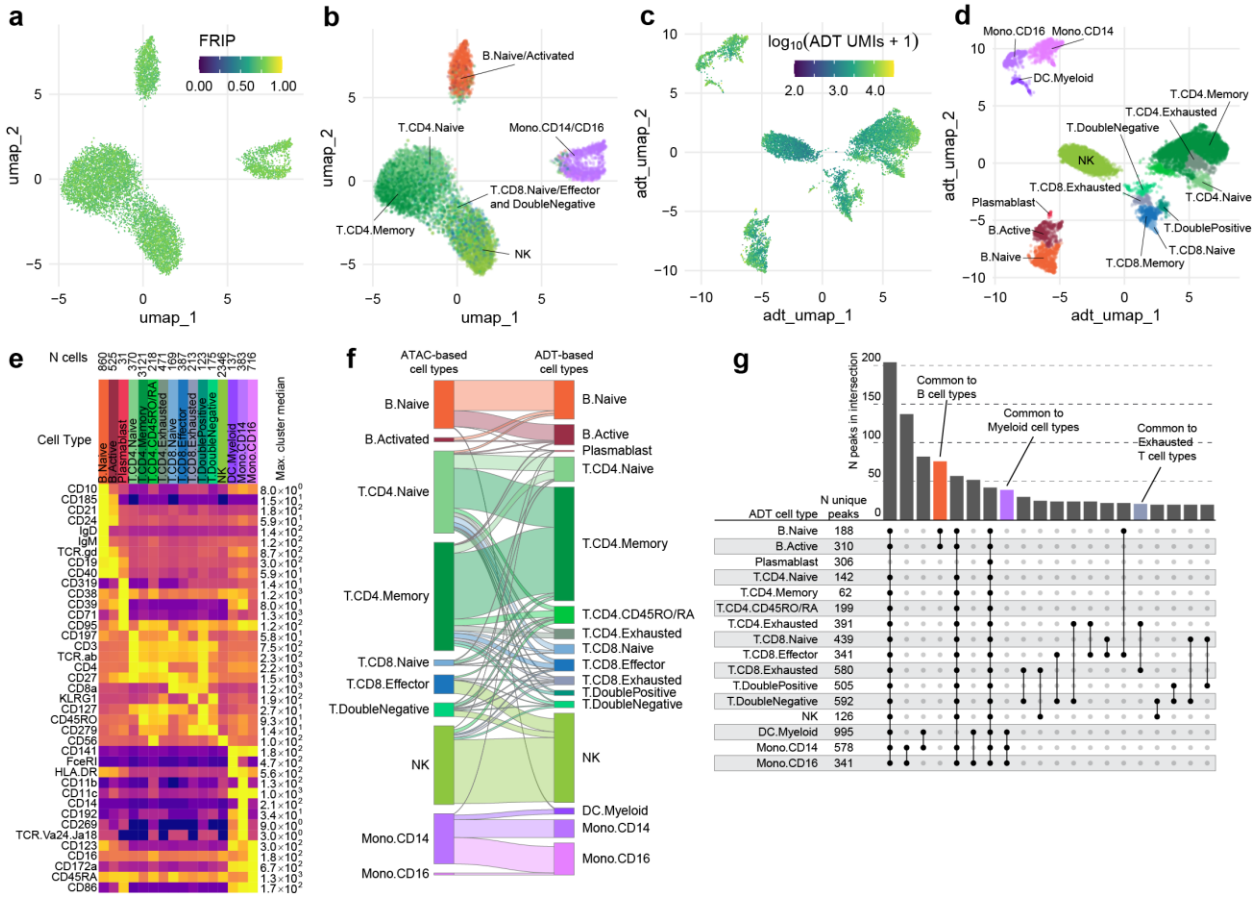


Figure 3. a, UMAP projection plot of ICICLE-seq cells based on scATAC-seq data. Cells are colored based on fraction of reads in peaks (FRIP). **b**, UMAP projection of scATAC-seq data, as in **a**. Cells are colored based on cell type labels obtained by ArchR label transfer (**Methods**). **c**, UMAP projection plot of ICICLE-seq cells based on antibody-derived tag (ADT) data. Cells are colored according to the total number of unique molecule indexes (UMIs) across all markers. **d**, UMAP projection based on ADT data, as in **c**, colored according to cell type labels derived from marker expression (**Methods**). **e**, Heatmap of median ADT count values for each marker in each cell type labeled in panel d. Values are separately scaled in each row between zero and the maximum value (right column) for each marker. **f**, Alluvial plot showing the relationship between ATAC-based cell type labels derived from label transfer in **b** and cell type labels derived from ADT-based markers in panel **d**. **g**, Peaks were called on aggregated single-cell data for each ADT-based cell type. For each cell type, the top 2,500 peaks were compared (except for plasmablasts, for which 592 peaks were identified). The number of unique peaks are displayed to the right of each cell type label, and peaks found in 20 combinations of cell types are displayed to the right. Closed/black points indicate cell types that are members of each intersection set, and bars above these points show the number of peaks in each intersection.

Optimization of scATAC-seq data collection from PBMCs will be of use to many researchers in the immunology field and beyond who seek to get the most high quality data from precious clinical samples. We find that isotonic cell permeabilization generates scATAC-seq libraries with high quality as measured

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

by FRIP, with low nucleosomal content, suggesting that chromatin state in the nuclei is unperturbed (**Fig. 1 and 2**). While a previous study utilized FACS followed by scATAC on individually sorted cells (Pi-ATAC[24]), the use of permeabilized cells enables simultaneous interrogation of chromatin accessibility state in the nucleus and the functional state of cells based on their cell surface proteins at unprecedented scale (ICICLE-seq, **Fig. 3**). This novel method adapts existing reagents to perform a paired reading of scATAC-seq and cell surface antibody barcodes. Further optimization of these methods, perhaps with the use of a bespoke or modified capture sequence rather than poly-A, may result in increased depth of scATAC-seq data in future ICICLE-seq methods. We anticipate that permeabilized cells may provide a stronger link between high-quality cytoplasmic scRNA-seq data and scATAC-seq data using truly paired methods like SNARE-seq[9], SHARE-seq[10], and 10x Single Cell Multiome sequencing. These methods provide a viable path towards simultaneous measurement of 3 or more compartments of cells (e.g. mRNA, chromatin accessibility, and cell surface proteins), and together with scCUT&Tag methods[25] will allow interrogation of specific epigenetic modifications at high cell type resolution to expand our view of the full picture of immune cell state in health and disease.

Author contributions

P.J.S., L.T.G., and E.S. designed the study. E.S. and C.L. performed scATAC-seq experiments. L.T.G. and R.G. performed scATAC-seq data processing. L.T.G. performed scATAC-seq analysis. E.S. designed and performed ICICLE-seq experiments. E.S. and L.T.G. performed ICICLE-seq analysis. J.R. performed FACS experiments. A.T.H. and J.R. performed flow cytometry. A.K.S., A.T.H., and J.R. designed flow cytometry panels and gating strategies. T.F.B. provided oversight of the Allen Institute for Immunology. T.R.T. provided oversight of the AIFI Experimental Immunology group and edited the manuscript. L.T.G., E.S., and P.J.S. wrote the manuscript, with input from all coauthors.

Acknowledgments

The authors thank Nina Kondza, Tanja Smith, Leila Shiraiwa, and Ernie Coffey for operational support, and Morgan Weiss for assistance with preliminary experiments. The authors thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement, and support.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Sample collection and preparation

Sample Collection and Processing

Biological specimens were purchased from BioIVT as cryopreserved PBMCs and Bloodworks NW as freshly drawn whole blood. All sample collections were conducted by BioIVT and Bloodworks NW under IRB-approved protocols, and all donors sign informed consent forms. See **Supplementary Table 8** for a list of sources and samples used for data displayed in each figure.

PBMCs sourced from BioIVT were isolated using either Ficoll-Paque or leukapheresis. Following isolation, PBMCs were subjected to RBC lysis, washing, and counting. PBMC aliquots were cryopreserved in Cryostor CS10 (StemCell Technologies, 07930) and stored in vapor phase liquid nitrogen.

For fresh blood samples from Bloodworks NW, PBMC processing occurred in-house. Blood tubes were pooled, gently swirled until fully mixed, about 30 times, and diluted with an equivalent volume of room temperature PBS (Thermo Fisher Scientific, 14190235). PBMCs were isolated using one or more Leucosep tubes (Greiner Bio-One, 227290) loaded with 15 mL of Ficoll Premium (GE Healthcare, 17-5442-03) to which a 3 mL cushion of PBS had been slowly added on top of the Leucosep barrier. Diluted whole blood (24-30mL) was slowly added to each tube and spun at 1000×g for 10 minutes at 20°C with no brake (Beckman Coulter Avanti J-15RIVD with JS4.750 swinging bucket, B99516). PBMCs were recovered from the Leucosep tube by quickly pouring all volume above the barrier into a sterile 50 mL conical tube (Corning, 352098). 15 mL cold PBS+0.2% BSA (Sigma, A9576; “PBS+BSA”) was added and the cells were pelleted at 400×g for 5-10 minutes at 4-10°C. The supernatant was quickly decanted, the pellet dispersed by flicking the tube, and the cells washed with 25-50 mL cold PBS+BSA. Cell pellets were combined as needed, the cells were pelleted as before, supernatant quickly decanted, and residual volume was carefully aspirated. PBMCs were resuspended in 1 mL cold PBS+BSA per 15 mL whole blood processed and counted with a ViCell (Beckman Coulter) using VersaLyse reagent (Beckman Coulter, A09777) or with a Cellometer Spectrum Cell Counter (Nexcelom) using ViaStain Acridine Orange/Propidium Iodide solution (Nexcelom, C52-0106-5). PBMCs were cryopreserved in Cryostor10 (StemCell Technologies, 07930) or 90% FBS (Thermo Fisher Scientific, 10438026) / 10% DMSO (Fisher Scientific, D12345) at 5×10^6 cells/mL by slow freezing in a Coolcell LX (VWR, 75779-720) overnight in a -80°C freezer followed by transfer to liquid nitrogen.

Cell Thawing

Cryopreserved PBMCs were removed from liquid nitrogen storage and thawed in a 37°C water bath for 3-5 minutes until no ice was visible. Cells were diluted to 10 mL in 37°C AIM V medium (Gibco, 12055091) with the first 3 mL added dropwise. Cells were then washed once with 10 mL DPBS without calcium and magnesium (Corning, 21-031-CM) supplemented with 0.2% w/v BSA (Sigma-Aldrich,

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A2934). Cells were counted on a Cellometer Spectrum Cell Counter (Nexcelom) using ViaStain Acridine Orange/Propidium Iodide solution (Nexcelom, C52-0106-5) and stored on ice.

FACS Neutrophil Depletion

To remove dead cells, debris, and neutrophils, PBMC samples were sorted by fluorescence activated cell sorting (FACS) prior to nuclei isolation or cell permeabilization. Cells were incubated with Fixable Viability Stain 510 (BD, 564406) for 15 minutes at room temperature and washed with AIM V medium (Gibco, 12055091) plus 25mM HEPES before incubating with TruStain FcX (BioLegend, 422302) for 5 minutes on ice, followed by staining with anti-CD45 (BioLegend, 304038) and anti-CD15 (BD, 562371) antibodies for 20 minutes on ice. Cells were washed with AIM V medium plus 25mM HEPES and sorted on a BD FACSAria Fusion. A standard viable CD45+ cell gating scheme was employed; FSC-A v SSC-A (to exclude sub-cellular debris), two FSC-A doublet exclusion gates (FSC-W followed by FSC-H), dead cell exclusion gate (BV510 LIVE/DEAD negative) followed by CD45+ inclusion gate. Neutrophils (defined as SSC^{high}, CD15+) were then excluded in the final sort gate (**Supplementary Fig. 3**). An aliquot of each post-sort population was used to collect 50,000 events to assess post-sort purity.

Magnetic Bead Neutrophil Depletion

Bead-based neutrophil depletion was performed using a biotin conjugated monoclonal anti-CD15 antibody in combination with streptavidin coated magnetic beads. A high neutrophil content (approximately 1.1%) Ficoll isolated PBMC sample was processed to evaluate efficacy, and a low neutrophil leukapheresis isolated PBMC sample was processed to control for off-target effects. Briefly, 1×10^7 PBMCs were resuspended in 100 µl of chilled DPBS without calcium and magnesium (Corning, 21-031-CM) supplemented with 0.2% w/v BSA (Sigma-Aldrich, A2934). 10 µl TruStain FcX (BioLegend 422302) was added to the cell suspension, mixed by pipette, and incubated on ice for 10 minutes. Anti-CD15 antibody (BioLegend, 301913) was added to the cell suspension, mixed by pipette, and incubated on ice for 30 minutes. Following antibody binding, 25 µl of Dynabeads MyOne Streptavidin T1 magnetic beads (Invitrogen, 65601) was added to the cell suspension, mixed by pipette, and incubated at room temperature for 5 minutes. The cell suspension was then diluted with 900 µl of room temperature DPBS+0.2% w/v BSA and placed on an EasySep magnet (Stemcell Technologies, 18103) for 3 minutes. The supernatant (approximately 1 ml) was transferred to a new tube and stored on ice until further processing.

Non-depleted and neutrophil depleted PBMCs from each sample were analyzed by flow cytometry using an 8-color panel to assess the effects of the bead based depletion on major PBMC populations. For each sample and condition, 1×10^6 cells were centrifuged (750×g for 5 minutes at 4°C) using a swinging bucket rotor (Beckman Coulter Avanti J-15RIVD with JS4.750 swinging bucket, B99516), the supernatant was removed using a vacuum aspirator pipette, and the cell pellet was resuspended in 100 µl of DPBS without calcium and magnesium (Corning, 21-031-CM) supplemented with 0.2% w/v BSA (Sigma-Aldrich, A2934). Cells were incubated with Fixable Viability Stain 510 (BD, 564406) and TruStain FcX (BioLegend, 422302) for 30 minutes on ice, and washed in chilled FACS buffer (DPBS, 0.2% w/v BSA, 0.1% sodium azide (VWR, BDH7465-2)). Cells were stained with a cocktail of antibodies (**Supplementary Table 6**) including 10 µl of Brilliant Stain Buffer Plus (BD, 566385) at a staining volume of 100 µl for 30 minutes on ice, then washed twice with chilled FACS buffer. Cells were passed through 35 µm Falcon Cell Strainers (Corning, 352235) and analyzed on a BD FACS

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Symphony flow cytometer. Gating analysis was performed using FlowJo cytometry software (Version 10.7).

A sequential gating scheme was used to identify viable singlet CD45+/CD15+/CD16+ neutrophils: 1, Time vs. SSC-A gate (to confirm that no abnormalities occurred in the fluidics), 2. FSC-A vs SSC-A (to exclude sub-cellular debris), 3. two FSC-A doublet exclusion gates (FSC-W followed by FSC-H), 3. dead cell exclusion gate (BV510 LIVE/DEAD negative) followed by 4. CD45+ inclusion gate. Neutrophils were defined as either SSC-A^{high}/CD15+ or CD15+/CD16+. The neutrophil population defined by SSC-A^{high}/CD15+ was larger than that defined by CD15+/CD16+ due to the presence of some contaminating CD15^{low} monocytes. Therefore, we used the CD45+/CD15+/CD16+ gate for subsequent analysis including summary statistics. (**Supplementary Fig. 4 and Supplementary Table 5**).

Standard Nuclei Isolation

Isolation of nuclei suspensions was performed according to the Demonstrated Protocol: Nuclei Isolation for Single Cell ATAC Sequencing (10x Genomics, CG000169 Rev C). Briefly, 8×10^5 to 1×10^6 cells were added to a 1.5 mL low binding tube (Eppendorf, 022431021) and centrifuged (300×g for 5 minutes at 4℃) using a swinging bucket rotor (Beckman Coulter Avanti J-15RIVD with JS4.750 swinging bucket, B99516). The supernatant was removed using a vacuum aspirator pipette and the cell pellet was resuspended in 100 µl of chilled 10x Genomics Nuclei Isolation Buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 0.1 % NP-40 Substitute CAS 9016-45-9 (BioVision 2127-50), 0.01% Digitonin (MP Biomedicals 0215948082), 1% BSA) by pipette-mixing 10 times. Cells were incubated on ice for 3 minutes, followed by dilution with 1 mL of chilled 10x Wash Buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% Tween-20 (BioRad 1610781), 1% BSA) by pipette-mixing 5 times. Nuclei were centrifuged (500×g for 3 minutes at 4℃) and the supernatant was slowly removed using a vacuum aspirator pipette. Nuclei were resuspended in chilled 1x Nuclei Buffer (10x Genomics, 2000207) to a target concentration of 3,000 - 6,000 nuclei per µl. Nuclei suspensions were passed through 35 µm Falcon Cell Strainers (Corning, 352235) and counted on a Cellometer Spectrum Cell Counter (Nexcelom) using ViaStain Acridine Orange/Propidium Iodide Staining Solution (Nexcelom, C52-0106-5).

Nuclei Isolation Optimization

In addition to 10x Nuclei Isolation Buffer (10xNIB), we tested an alternative Nuclei Isolation Buffer (ANIB) as described previously[26] (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% Tween-20, 0.1 % IGEPAL CAS 9002-93-1 (Sigma, I8896), 1x Protease Inhibitor (Roche, 11836170001)). For each buffer, we generated a titration series of detergent concentrations relative to the concentrations described above, but did not alter the concentration of other buffer ingredients: 1x, 0.5x, 0.25x, 0.1x for 10xNIB, and 1x and 0.1x for ANIB. The resulting nuclei were imaged using an EVOS M5000 Imaging System (Thermo Fisher Scientific, AMF5000) in transmitted light mode at 40x magnification to visually evaluate nuclear integrity (**Supplementary Fig. 1**). The 1x 10xNIB, 0.25x 10xNIB, 0.1x 10xNIB, and 1x ANIB were used for 10X scATAC-seq (**Supplementary Fig. 2**).

Cell Permeabilization

We prepared a 5% w/v digitonin stock by diluting powdered Digitonin (MP Biomedicals, 0215948082) with 100% DMSO (Fisher Scientific, D12345) and creating 20 µl aliquots which were stored at -20°C.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To permeabilize, 1,000,000 cells were added to a 1.5 mL low binding tube (Eppendorf, 022431021) and centrifuged (400×g for 5 minutes at 4°C) using a swinging bucket rotor (Beckman Coulter Avanti J-15RIVD with JS4.750 swinging bucket, B99516). The supernatant was removed using a vacuum aspirator pipette and the cell pellet was resuspended in 100 µl of chilled isotonic Perm Buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 3 mM MgCl2, 0.01% Digitonin) by pipette-mixing ten times. Cells were incubated on ice for 5 minutes, after which they were diluted with 1 mL of isotonic Wash Buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 3 mM MgCl2) by pipette-mixing 5 times. Cells were centrifuged (400×g for 5 minutes at 4°C) using a swinging bucket rotor and the supernatant was slowly removed using a vacuum aspirator pipette. The cell pellet was resuspended in chilled TD1 buffer (Illumina, 15027866) by pipette-mixing to a target concentration of 2,300 - 10,000 cells per µl. Cells were passed through 35 µm Falcon Cell Strainers (Corning, 352235) and counted on a Cellometer Spectrum Cell Counter (Nexcelom) using ViaStain Acridine Orange/Propidium Iodide solution (Nexcelom, C52-0106-5). For optimization, we used varying final digitonin concentrations in the Perm Buffer: 0.01% w/v, 0.05% w/v, 0.1% w/v, and 0.2% w/v. The optimal concentration observed was 0.01% w/v.

snATAC-seq and scATAC-seq

10X ATAC-seq Library Preparation

scATAC-seq libraries were prepared according to the Chromium Single Cell ATAC v1.1 Reagent Kits User Guide (CG000209 Rev B) with several modifications. 15,000 cells or nuclei were loaded into each tagmentation reaction. Nuclei were brought up to a volume of 5 µl in 1x Nuclei Buffer (10x Genomics, 2000207), mixed with 10 µl of a transposition master mix consisting of ATAC Buffer B (10x Genomics, 2000193) and ATAC Enzyme (Tn5 transposase; 10x Genomics, 2000123). Permeabilized cells were brought up to a volume of 9 µl in TD1 buffer (Illumina, 15027866) and mixed with 6 µl of Illumina TDE1 Tn5 transposase (Illumina, 15027916). Transposition was performed by incubating the prepared reactions on a C1000 Touch thermal cycler with 96–Deep Well Reaction Module (Bio-Rad, 1851197) at 37°C for 60 minutes, followed by a brief hold at 4°C. A Chromium NextGEM Chip H (10x Genomics, 2000180) was placed in a Chromium Next GEM Secondary Holder (10x Genomics, 3000332) and 50% Glycerol (Teknova, G1798) was dispensed into all unused wells. A master mix composed of Barcoding Reagent B (10x Genomics, 2000194), Reducing Agent B (10x Genomics, 2000087), and Barcoding Enzyme (10x Genomics, 2000125) was then added to each sample well, pipette-mixed, and loaded into row 1 of the chip. Chromium Single Cell ATAC Gel Beads v1.1 (10x Genomics, 2000210) were vortexed for 30 seconds and loaded into row 2 of the chip, along with Partitioning Oil (10x Genomics, 2000190) in row 3. A 10x Gasket (10x Genomics, 370017) was placed over the chip and attached to the Secondary Holder. The chip was loaded into a Chromium Single Cell Controller instrument (10x Genomics, 120270) for GEM generation. At the completion of the run, GEMs were collected and linear amplification was performed on a C1000 Touch thermal cycler with 96–Deep Well Reaction Module: 72°C for 5 min, 98°C for 30 sec, 12 cycles of: 98°C for 10 sec, 59°C for 30 sec and 72°C for 1 min.

GEMs were separated into a biphasic mixture through addition of Recovery Agent (10x Genomics, 220016), the aqueous phase was retained and removed of barcoding reagents using Dynabead MyOne SILANE (10x Genomics, 2000048) and SPRIselect reagent (Beckman Coulter, B23318) bead clean-ups. Sequencing libraries were constructed by amplifying the barcoded ATAC fragments in a sample indexing PCR consisting of SI-PCR Primer B (10x Genomics, 2000128), Amp Mix (10x Genomics, 2000047) and Chromium i7 Sample Index Plate N, Set A (10x Genomics, 3000262) as described in the

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10x scATAC User Guide. Amplification was performed in a C1000 Touch thermal cycler with 96–Deep Well Reaction Module: 98°C for 45 sec, for 9 to 11 cycles of: 98°C for 20 sec, 67°C for 30 sec, 72°C for 20 sec, with a final extension of 72°C for 1 min. Final libraries were prepared using a dual-sided SPRIselect size-selection cleanup. SPRIselect beads were mixed with completed PCR reactions at a ratio of 0.4x bead:sample and incubated at room temperature to bind large DNA fragments. Reactions were incubated on a magnet, the supernatant was transferred and mixed with additional SPRIselect reagent to a final ratio of 1.2x bead:sample (ratio includes first SPRI addition) and incubated at room temperature to bind ATAC fragments. Reactions were incubated on a magnet, the supernatant containing unbound PCR primers and reagents was discarded, and DNA bound SPRI beads were washed twice with 80% v/v ethanol. SPRI beads were resuspended in Buffer EB (Qiagen, 1014609), incubated on a magnet, and the supernatant was transferred resulting in final, sequencing-ready libraries.

Sequencing

Final libraries were quantified using a Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific, P7589) on a SpectraMax iD3 (Molecular Devices). Library quality and average fragment size was assessed using a Bioanalyzer (Agilent, G2939A) High Sensitivity DNA chip (Agilent, 5067-4626). Libraries were sequenced on the Illumina NovaSeq platform with the following read lengths: 51nt read 1, 8nt i7 index, 16nt i5 index, 51nt read 2.

10X scATAC-seq Data Processing

Demultiplexing of raw base call files into FASTQ files was performed using 10x cellranger-atac mkfastq (10x Genomics v.1.1.0). To assess samples at an equal sequencing depth, FASTQ files were downsampled to a uniform total raw read count among compared samples: 2×10^8 fragments for comparison of nuclei and cells across FACS conditions (**Fig. 1 and 2**); 1.25×10^8 fragments for optimization experiments (**Supplementary Fig. 2**) due to lower available total read depth after sequencing. 10x cellranger-atac count was used to process sequencing reads by performing adapter trimming and sequence alignment to the GRCh38 (hg38) reference genome (refdata-cellranger-atac-GRCh38-1.1.0). The output files fragments.tsv.gz and singlecell.csv were utilized for downstream processing and quality control analysis.

To evaluate quality control metrics across all scATAC-seq datasets, we utilized bedtools (v2.29.1) and GNU parallel[27] v20161222 to generate overlap counts and feature count matrices for a panel of reference genomic regions: 518,766 peaks from a previous study of PBMCs by scATAC-seq[7] (supplementary file GSE123577_pbmc_peaks.bed.gz from GEO accession GSE123577) were converted from hg19 to hg38 coordinates using the UCSC liftOver tool[28] (kent source v402) and used to compute a standardized fraction of reads in peaks score for cells in each dataset (FRIP); 33,496 transcription start site regions (TSS ± 2kb) from Hg38 ENSEMBL release 93[29] were filtered to select genes used in the 10x Genomics cellranger GRCh38 reference for scRNA-seq (refdata-cellranger-GRCh38-3.0.0) and used to compute the fraction of reads in TSS (FRITSS); and a set of 3,591,898 reference DNase hypersensitive sites from ENCODE[13] (ENCODE File ID ENCFF503GCK) were used to assess distal regulatory element accessibility. In addition, we generated tiled window counts across the genome in 5k, 20k, 100k bins.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

scATAC-seq Quality Control

Custom R scripts were used to assess and filter preprocessed scATAC-seq data along a variety of quality metrics. Cells or nuclei with > 1,000 uniquely aligned fragments, FRIP > 0.2, FRITSS > 0.2, and fraction of fragments overlapping ENCODE reference regions > 0.5 were retained for downstream analysis. Cells or nuclei that passed these QC cutoffs were used to generate sparse count matrices and filtered fragments.tsv.gz files for downstream analysis.

To examine aggregate TSS accessibility, we selected fragments from fragments.tsv.gz that overlapped TSS regions describe above (TSS ± 2kb). For plotting, fragments were separated using cell barcodes (and fragment length in the case of **Supplementary Fig. 5**) into separate groups. Fragment positions were converted to positions relative to TSS (sensitive to transcript strand orientation), and the number of fragments overlapping each position were calculated.

To examine CTCF motif accessibility, CTCF motif locations were obtained from genome-wide motif scans of non-redundant TF motifs[30] (https://resources.altius.org/~jvierstra/projects/motif-clustering/releases/v1.0). Motifs were filtered to select CTCF motifs that overlapped ENCODE reference regions[13] (ENCODE File ID ENCFF503GCK). Selected were ranked by their MOODS match, and the top 100,000 motifs were selected for analysis. Motif locations were expanded to a total of 4kb centered on the middle of each CTCF motif (using the resize function from GenomicRanges in R). Fragments from cells that passed QC filtering were converted to target site duplication (TSD) center positions (+5 bp from the 5' end and -4 bp from the 3' end of each fragment). All TSD centers that overlapped expanded CTCF motif regions were selected, and the number of TSD centers that overlapped each position relative to the CTCF motif were calculated (sensitive to CTCF motif strand orientation).

scATAC-seq Dimensionality Reduction

For 2D projections of scATAC-seq data, we used binarized sparse matrices of 20kb window accessibility across the hg38 genome (excluding mitochondrial regions, chrM). Independently for each dataset, we selected features found in > 3% of cells/nuclei, weighted features using term frequency - inverse document frequency, log-transformed the resulting weights, and performed PCA using singular value decomposition to generate 50 reduced dimensions as described previously[31,32]. We then removed the first PC, which was strongly correlated with the number of available fragments and retained the remaining PCs up to PC 30. For display, we further reduced the dimensionality of selected PCs using UMAP[33,34] (R package uwot, v0.1.8, parameters: scale = TRUE, min_dist = 0.2).

scATAC-seq Cell Type Labeling

Labeling of scATAC-seq datasets was performed using the ArchR package[22] v0.9.4. In brief, filtered fragments.tsv.gz files after quality control were used to generate an ArchR GeneScore matrix and a tiled genome feature matrix for each dataset. Cells were grouped by performing iterative latent semantic indexing (LSI) on the tile matrix, followed by the shared nearest neighbor clustering approach implemented in Seurat[23] v3.1.5. GeneScore data was then used to compare scATAC-seq clusters to a labeled reference scRNA-seq dataset consisting of 9,380 PBMCs generated by 10x Genomics, with labels provided by the Satija lab (https://www.dropbox.com/s/zn6khirjafoyyxl/pbmc_10k_v3.rds?dl=0) using ArchR's implementation of the FindTransferAnchors method from Seurat. The best-scoring labels

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for each scATAC-seq cluster were used for downstream analysis and display (**Fig. 2**), and label transfer scores for individual cells were used to compare label transfer between methods (**Fig. 2c**).

scATAC-seq Peak Analysis

After labeling cell types in each dataset, peaks for each cell type were generated using the ArchR functions addGroupCoverages and addReproduciblePeakSet. Within each dataset, peaks scores from each pair of cell types were compared using getMarkerFeatures performed in each direction separately by swapping foreground and background cell types. Differentially accessible peaks (DAPs) from each comparison were selected with filter string "FDR <= 0.05 & Log2FC >= log2(1.5)".

To identify enriched TFBS motifs, CIS-BP motif annotations[35] were attached to each peak identified by ArchR using the addMotifAnnotations. Marker peaks for each cell type were identified using getMarkerFeatures without specifying foreground and background groups. Enriched motifs were identified using the ArchR function peakAnnoEnrichment (parameters: peakAnnotation = "Motif", cutOff = "FDR <= 0.01 & Log2FC >= log2(1.5)"). Up to the top 10 enriched motifs for each cell type were plotted using the ArchR function plotEnrichHeatmap.

Cell Type Flow Cytometry

To assess cell type proportions, PBMCs were analyzed with a 25-color immunophenotyping flow cytometry panel. 1×10^6 thawed PBMCs were centrifuged (750×g for 5 minutes at 4°C) using a swinging bucket rotor (Beckman Coulter Avanti J-15RIVD with JS4.750 swinging bucket, B99516), the supernatant was removed using a vacuum aspirator pipette, and the cell pellet was resuspended in DPBS without calcium and magnesium (Corning 21-031-CM). Cells were incubated with Fixable Viability Stain 510 (BD, 564406) and TruStain FcX (BioLegend, 422302) for 30 minutes at 4°C, then washed with chilled PBS+0.2% BSA (Sigma, A9576; “PBS+BSA”). Cells were stained with a cocktail of antibodies (**Supplementary Table 3**) at a staining volume of 100 µl for 30 minutes at 4°C, then washed with PBS+0.2% BSA. Fixation was performed by resuspending cells in 100 µl of 4% Paraformaldehyde (Electron Microscopy Sciences, 15713) and incubating for 15 minutes at 25°C, protected from light. Following fixation, cells were washed twice with PBS+0.2% BSA and resuspended in 100 µl PBS (without BSA). Stained cells were analyzed on a 5 laser Cytek Aurora spectral flow cytometer. Spectral unmixing was calculated with pre-recorded reference controls using Cytek SpectroFlo software (Version 2.0.2). Cell types were quantified by traditional bivariate gating analysis performed with FlowJo cytometry software (Version 10.7, **Supplementary Figure 6**).

ICICLE-seq

Tn5 Complexing

The assembly of Tn5 transposomes was performed as previously described[26]. DNA complexes containing mosaic-end sequences with either a poly-T or Nextera R2N 5' overhang (Poly-T Top-L/MOSAIC_Bot, Tn5ME-s7_Top/MOSAIC_Bot) were created by annealing equimolar amounts of top and bottom oligos (**Supplementary Fig. 7 and Supplementary Table 2**) on a C1000 Touch thermal cycler with 96-Deep Well Reaction Module (Bio-Rad, 1851197) at 95°C for 5 minutes followed by 5°C decreases every 2 minutes until the temperature reached 20°C. Oligos were annealed at a concentration of 16 µM in 2x Dialysis Buffer (100 mM HEPES-KOH pH 7.5 (Teknova, 550000-016), 200

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mM NaCl, 0.2 mM EDTA, 2 mM DTT (IBI Scientific, 21040), 0.2% Triton X-100 (Sigma-Aldrich, T8787), 20% Glycerol (Teknova, G1798)). Annealed complexes were mixed together 1:1 for a final concentration of 8 nM. Tn5 transposase (Beta Lifescience, TN5-BL01) was supplemented with 5M NaCl at a final volume ratio of 1:8 NaCl to Tn5. The resulting NaCl/Tn5 mixture was mixed with the annealed complexes at a volume ratio of 1.2:1 ratio of DNA complexes to Tn5 and incubated at 25°C for 60 minutes to form final, reaction ready Tn5 complexes, which were stored at -20°C until use.

Antibody Staining

PBMCs were depleted of neutrophils, dead cells, and debris through FACS as described above. 2×10^6 sorted PBMCs were centrifuged (400×g for 5 minutes at 4°C) using a swinging bucket rotor (Beckman Coulter Avanti J-15RIVD with JS4.750 swinging bucket, B99516), the supernatant was removed using a vacuum aspirator pipette, and the cell pellet was resuspended in 100 µl of DPBS without calcium and magnesium (Corning 21-031-CM) supplemented with 0.2% w/v BSA (Sigma-Aldrich A2934). 10 µl TruStain FcX (BioLegend, 422302) was added and cells were incubated on ice for 10 minutes. A panel of 46 barcoded oligo-conjugated antibodies (BioLegend TotalSeq-A) including a mouse IgG1κ isotype negative control (**Supplementary Table 1**) was added and incubated on ice for 30 minutes. Cells were washed three times in 4 mL of DPBS plus 2% BSA to remove unbound antibodies and used as input into cell permeabilization with 0.01% digitonin as described above.

ICICLE-seq Library Preparation

Transposition was performed by aliquoting 20,000 permeabilized cells in TD1 buffer (Illumina, 15027866), bringing the volume up to 9 µl in TD1 buffer, and mixing with 6 µl of Poly-T overhang Tn5 complexes. Reactions were incubated on a C1000 Touch thermal cycler with 96-Deep Well Reaction Module (Bio-Rad, 1851197) at 37°C for 120 minutes, followed by a brief hold at 4°C. Cell barcodes were then added to ATAC and antibody derived tags (ADTs) via GEM generation using 10x Genomics 3' RNA beads and subsequent amplification. Briefly, a Chromium Next GEM Chip G (10x Genomics, 2000177) was placed in a Chromium Next GEM Secondary Holder (10x Genomics, 3000332) and 50% Glycerol (Teknova, G1798) was dispensed into all unused wells. A barcoding master mix was prepared which consisted of NEBNext Ultra II Q5 Master Mix (New England Biolabs, M0544), Reducing Agent B (10x Genomics, 2000087), F BC Primer (0.2 µM **Supplementary Table 2**), and ADT-Rev-AMP (0.2 µM **Supplementary Table 2**). The master mix was added to each sample well, pipette-mixed, and loaded into row 1 of the chip. Chromium Single Cell 3' v3.1 Gel Beads (10x Genomics, 2000164) were vortexed for 30 seconds and loaded into row 2 of the chip, along with Partitioning Oil (10x Genomics, 2000190) in row 3. A 10x Gasket (10x Genomics, 370017) was placed over the chip and attached to the Secondary Holder. The chip was loaded into a Chromium Single Cell Controller instrument (10x Genomics, 120270) for GEM generation. At the completion of the run, GEMs were collected and amplification was performed on a C1000 Touch thermal cycler with 96-Deep Well Reaction Module: 72°C for 5 min, 98°C for 30 sec, 12 cycles of: 98°C for 10 sec, 42°C for 30 sec and 65°C for 30 sec, followed by a final extension of 65°C for 1 min.

GEMs were separated into a biphasic mixture through addition of Recovery Agent (10x Genomics, 220016), the aqueous phase was retained and removed of barcoding reagents using Dynabead MyOne SILANE (10x Genomics, 2000048) beads. Next, a dual-sided 0.6x/2.0x bead:sample SPRIselect reagent (Beckman Coulter, B23318) size-selection clean-up was performed to remove large DNA fragments and unused primers. Libraries were split into two reactions in a 3:1 ATAC:ADT ratio and

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

amplified separately using different indexed P7 primers. ATAC fragments were amplified in a 100 µl reaction consisting of Buffer EB (Qiagen, 1014609), Amp Mix (10x Genomics, 2000047), SI-P5-22 primer (20 µM **Supplementary Table 2**), and Chromium i7 Multiplex Kit N Set A (10x Genomics, 3000262). ATAC PCR was performed in a C1000 Touch thermal cycler with 96-Deep Well Reaction Module: 98°C for 45 sec, 7 cycles of: 98°C for 20 sec, 54°C for 30 sec, 72°C for 20 sec, followed by a final extension of 72°C for 1 min. ADT fragments were amplified in a 100 µl reaction consisting of Buffer EB (Qiagen, 1014609), KAPA HiFi HotStart ReadyMix (KAPA Biosystems, KM2602), SI-P5-22 primer (10 µM **Supplementary Table 2**), and ADT i7 primer (10 µM **Supplementary Table 2**). ADT PCR was performed in a C1000 Touch thermal cycler with 96-Deep Well Reaction Module: 95°C for 3 min, 15 cycles of: 95°C for 20 sec, 60°C for 30 sec, 72°C for 20 sec, followed by a final extension of 72°C for 5 min. SPRIselect reagent cleanups were performed with a 1.2x bead:sample ratio for ADT libraries and a dual-sided size-selection of 0.4x/1.2x bead:sample ratio for ATAC libraries.

ICICLE-seq Sequencing

Final libraries were quantified using qPCR (KAPA Biosystems Library Quantification Kit for Illumina, KK4844) on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, 1855195). Library quality and average fragment size were assessed using a Bioanalyzer (Agilent, G2939A) High Sensitivity DNA chip (Agilent, 5067-4626). Libraries were sequenced on the Illumina NovaSeq platform with the following read lengths: 28 bp read 1 (Cell barcode and UMI), 8 bp i7 index, 100 bp read 2 (ATAC-seq sequence or ADT barcode). A Truseq read 1 primer (0.3 µM **Supplementary Table 2**) was included as a Custom Read 1 primer to mitigate the risk of off-target priming of the standard Illumina Nextera read 1 primer on the partial Nextera R1N sequence included in the mosaic end portion of the Poly-T Tn5 insertion.

ICICLE-seq Data Preprocessing

Demultiplexing of raw base call files into FASTQ files was performed using bcl2fastq2 (Illumina v2.20.0.422). Read 2 was trimmed of adapter sequences, low quality bases and reads, and polyA tailing using fastp[36] v0.21.0 (parameters: --adapter_sequence=CTGTCTCTTATACACATCT --cut_tail --trim_poly_x) and the resulting read 2 sequences were aligned to the GRCh38 (hg38) reference genome (Illumina iGenomes, https://support.illumina.com/sequencing/sequencing_software/igenome.html) using Bowtie 2[37] (v2.3.0, parameters: --local, --sensitive, --no-unal, --phred33). Aligned reads in SAM format were filtered by alignment score (greater than or equal to 30) then tagged with cell barcode and UMI sequence and quality scores using custom python code (python3 v3.7.3). Barcode sequences were compared against the 10x Genomics v3 3' GEX barcode whitelist (3M-february-2018.txt.gz). Sequences not included in the whitelist were corrected to a valid whitelist barcode by allowing a single base mismatch (Hamming distance of 1). Sequences with more than one possible match were corrected at the position with the lowest sequencing quality score. Reads with barcodes that could not be corrected were excluded from further analysis. Filtered and tagged SAM files were converted to sorted, indexed BAM files using GATK[38] (Broad Institute v4.1.4.0). Genomic coordinates were converted to BED format using bedtools[39] (v2.26.0). Custom python code was used to collapse aligned fragments into a list of fragments with unique cell barcode and genomic coordinate combinations. These fragments were then written as a fragments.tsv.gz file in the format: chr, start position, end position, cell barcode, UMI count, and strand (+/-).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ADT Data Preprocessing

Methods for ADT counting were developed in-house and were implemented as an optimized, highly efficient C program named BarCounter. BarCounter was used for single-cell ADT counting as follows: firstly, barcode sequences were compared against the 10x Genomics v3 3' GEX barcode whitelist (3M-february-2018.txt.gz). Sequences not included in the whitelist were corrected to a valid whitelist barcode by allowing a single base mismatch (hamming distance of 1) at a low quality basecall (sequencing quality score < 20). Reads with barcodes that could not be corrected were excluded from further analysis. Next, ADT barcode sequences were compared against a CSV taglist containing ADT barcode / antibody associations. Antibody barcodes in the current TotalSeq-A catalog (BioLegend) have a Hamming distance from all other barcodes of at least 3. Therefore, a single base mismatch (Hamming distance of 1) was allowed. Reads containing ADT sequences that could not be assigned to an antibody in the taglist were excluded. Finally, UMI sequences that were unique within their assigned ADT for their assigned cell barcode were counted. Final ADT UMI counts were written by cell barcode to a CSV file for use in downstream analysis.

ADT features were filtered by comparison to the mouse IgG1κ isotype control, which should not bind to human cell surface proteins. The distribution of counts for each antibody was compared to the control using a Mann-Whitney test (R function wilcox.test with parameter alternative = "greater"). Any features for which the test returned a p-value > 1×10^{-9} were considered similar to the control and were removed from downstream analysis (**Supplementary Table 1**).

ICICLE-seq Analysis

Aligned ICICLE-seq chromatin accessibility fragments.tsv.gz files were preprocessed as for 10x scATAC-seq samples, above. QC filtering was performed as described, with modified cutoffs: > 500 uniquely aligned reads, FRIP > 0.65, FRITSS > 0.2, and fraction of fragments overlapping ENCODE reference regions > 0.5 were retained for downstream analysis. For 2D projections of the scATAC-seq data, we used binarized sparse matrices of 20kb window accessibility across the hg38 genome, selected features found in > 0.5% of cells, weighted features using LogTF-IDF, and performed PCA as described above. We then removed the first PC and retained the remaining PCs up to PC 20. UMAP was performed with adjusted parameters (scale = FALSE, min_dist = 0.2). To assign cell type labels, filtered fragments.tsv.gz files were used as input to ArchR. ArchR functions addIterativeLSI and addGeneIntegrationMatrix (parameters transferParams = list(dims = 1:10, k.weight = 20) and nGenes = 4000) were used to transfer labels from the scRNA-seq PBMC reference described above (scATAC Cell Type Labeling).

Count matrices for ADT data were scaled for each cell by dividing by the thousands of total ADT UMIs per cell, then transformed using Log_{10}(scaled count + 1). Normalized features were used for PCA using the R function prcomp with default parameters. Filtered and normalized features were used as direct input to UMAP (R package uwot with parameter min_dist = 0.2) with the first 2 PCs from PCA used as initial coordinates to aid reproducibility of UMAP projection. Cells were clustered using a Jaccard-Louvain method (parameters k = 15, radius = 1) using UMAP coordinates. Clusters with high signal from the mouse IgG1κ isotype control antibody were removed from subsequent analysis (1 cluster, n = 32 cells). The remaining clusters were manually labeled by examination of cell type marker enrichment. The R package scratch.vis[40] (https://github.com/alleninstitute/scrattch.vis) was used to generate the

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cluster median heatmap plot and a river/alluvial plot comparing the cell type labels obtained from ATAC-seq and ADT-based analyses.

To compare peaks between cell types, all filtered fragments for cells in each cell type were aggregated and used as input to the MACS2 peak caller[41] (parameters -f BED, -g hs, --no-model). The top 2,500 peaks from each cell type were selected for comparison (except for Plasmablasts, for which all 592 peaks were used). A master set of peaks across all types was constructed by combining all narrowPeak results files and combining the outer coordinates of overlapping peaks (GenomicRanges function reduce). A binary matrix of peak overlaps for each cell type was generated and used to construct the peak comparison figure inspired by UpSet plots[42].

Data analysis and visualization software

Post-processing analysis of summary statistics and visualization of snATAC-seq, scATAC-seq, and ICICLE-seq was performed using R v.3.6.3 and greater[43] in the Rstudio IDE[44] (Integrated Development Environment for R) or using the Rstudio Server Open Source Edition as well as the following packages: for scATAC-seq specific analyses and comparisons to scRNA-seq data, ArchR[22] and Seurat[23]; for general data analysis and manipulation, data.table[45], dplyr, Matrix[46], matrixStats[47], purrr[48], and reshape2[49]; for data visualization, ggplot2[50], and cowplot[51]; for dimensionality reduction and clustering, igraph[52], RANN, and the R uwot implementation[53] of UMAP[33,34]; for manipulation of genomic region data, bedtools2[39] and GenomicRanges[54].

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

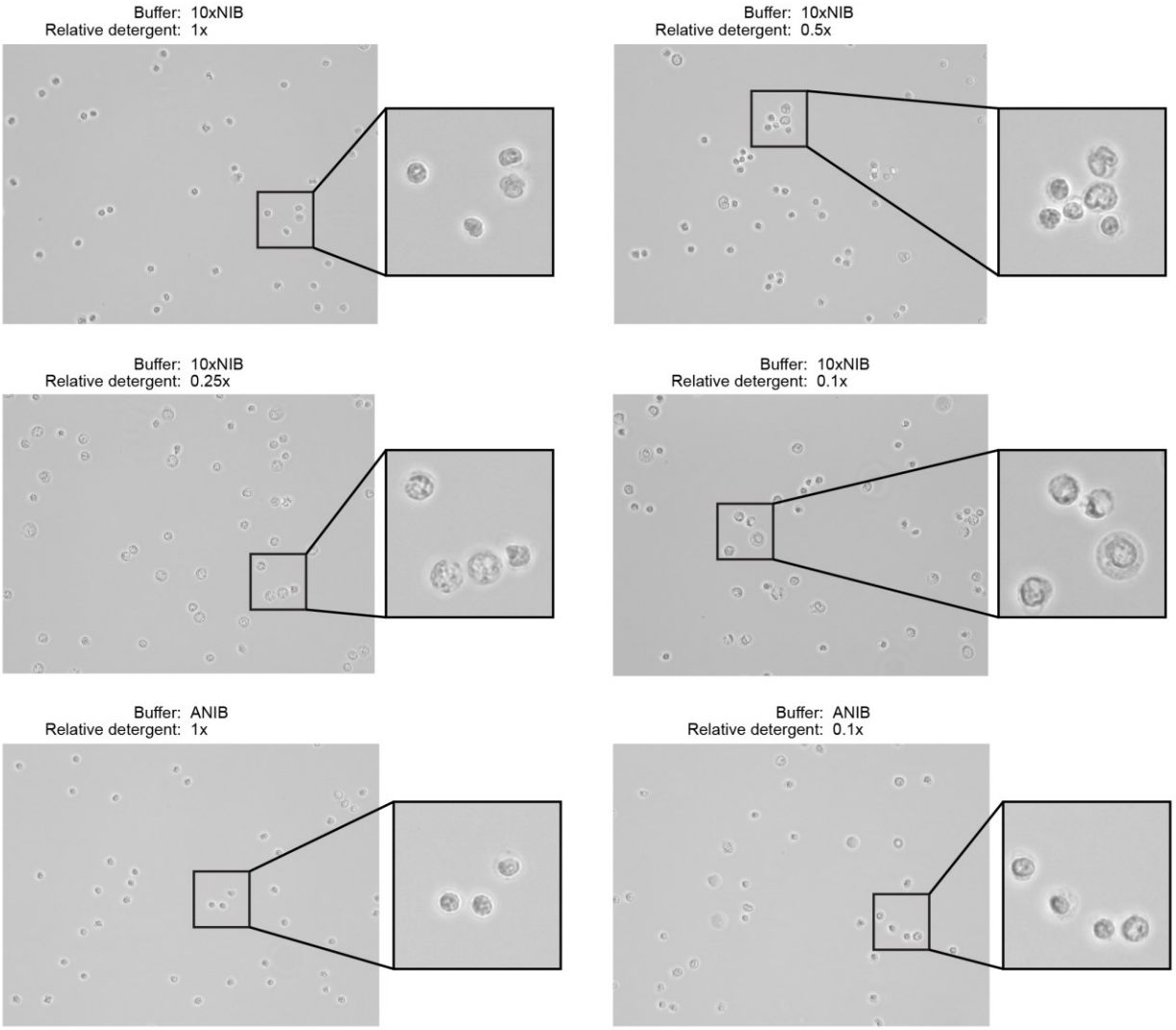
1. Böyum, A. Isolation of mononuclear cells and granulocytes from human blood. *Scand. J. Clin. Lab. Invest. Suppl.* **97**, 77–89 (1968).
2. Laurenti, E. & Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* (2018). doi:10.1038/nature25022
3. Wang, W. *et al.* Joint profiling of chromatin accessibility and CAR-T integration site analysis at population and single-cell levels. *Proc. Natl. Acad. Sci. U. S. A.* (2020). doi:10.1073/pnas.1919259117
4. Satpathy, A. T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* (2019). doi:10.1038/s41587-019-0206-z
5. Zheng, Y. *et al.* A human circulating immune cell landscape in aging and COVID-19. *Protein Cell* (2020). doi:10.1007/s13238-020-00762-2
6. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–90 (2015).
7. Lareau, C. A. *et al.* Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat. Biotechnol.* (2019). doi:10.1038/s41587-019-0147-6
8. Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science (80-.).* (2018). doi:10.1126/science.aau0730
9. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* (2019). doi:10.1038/s41587-019-0290-0
10. Ma, S. *et al.* Chromatin potential identified by shared single cell profiling of RNA and chromatin. *bioRxiv* (2020). doi:10.1101/2020.06.17.156943
11. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* (2017). doi:10.1038/nmeth.4380
12. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* (2017). doi:10.1038/nmeth.4396
13. Meuleman, W. *et al.* Index and biological spectrum of human DNase I hypersensitive sites. *Nature* (2020). doi:10.1038/s41586-020-2559-3
14. Marquina-Sanchez, B. *et al.* Single-cell RNA-seq with spike-in cells enables accurate quantification of cell-specific drug effects in pancreatic islets. *Genome Biol.* (2020). doi:10.1186/s13059-020-02006-2
15. Lima, A. F. *et al.* Osmotic modulation of chromatin impacts on efficiency and kinetics of cell fate modulation. *Sci. Rep.* (2018). doi:10.1038/s41598-018-25517-2
16. Colbeau, A., Nachbaur, J. & Vignais, P. M. Enzymac characterization and lipid composition of rat liver subcellular membranes. *BBA - Biomembr.* (1971). doi:10.1016/0005-2736(71)90123-4
17. Adam, S. A., Marr, R. S. & Gerace, L. Nuclear protein import in permeabilized mammalian cells requires soluble cytoplasmic factors. *J. Cell Biol.* (1990). doi:10.1083/jcb.111.3.807
18. Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* (2016). doi:10.1038/ng.3646
19. Chen, X., Miragaia, R. J., Natarajan, K. N. & Teichmann, S. A. A rapid and robust method for single cell chromatin accessibility profiling. *Nat. Commun.* (2018). doi:10.1038/s41467-018-07771-0
20. Li, Z. *et al.* Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* (2019). doi:10.1186/s13059-019-1642-2
21. Allahverdi, A., Chen, Q., Korolev, N. & Nordenskiöld, L. Chromatin compaction under mixed salt conditions: Opposite effects of sodium and potassium ions on nucleosome array folding. *Sci. Rep.* (2015). doi:10.1038/srep08512
22. Granja, J. M. *et al.* ArchR: An integrative and scalable software package for single-cell chromatin accessibility analysis. *bioRxiv* (2020). doi:10.1101/2020.04.28.066498
23. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* (2019). doi:10.1016/j.cell.2019.05.031

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

24. Chen, X. *et al.* Joint single-cell DNA accessibility and protein epitope profiling reveals environmental regulation of epigenomic heterogeneity. *Nat. Commun.* (2018). doi:10.1038/s41467-018-07115-y
25. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell profiling of histone modifications in the mouse brain. *bioRxiv* 2020.09.02.279703 (2020). doi:10.1101/2020.09.02.279703
26. Mulqueen, R. *et al.* Improved single-cell ATAC-seq reveals chromatin dynamics of in vitro corticogenesis. *bioRxiv* (2019). doi:10.1101/637256
27. Tange, O. GNU Parallel: the command-line power tool. ;*login USENIX Mag.* (2011). doi:10.5281/zenodo.16303
28. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* (2006). doi:10.1093/nar/gkj144
29. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res.* (2020). doi:10.1093/nar/gkz966
30. Vierstra, J. Non-redundant TF motif matches genome-wide. (2020).
31. Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* (2018). doi:10.1016/j.cell.2018.06.052
32. Hill, A. J. Dimensionality Reduction for scATAC Data. (2019). Available at: http://andrewjohnhill.com/blog/2019/05/06/dimensionality-reduction-for-scatac-data/.
33. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* (2018). doi:10.21105/joss.00861
34. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* (2019). doi:10.1038/nbt.4314
35. Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* (2014). doi:10.1016/j.cell.2014.08.009
36. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. in *Bioinformatics* (2018). doi:10.1093/bioinformatics/bty560
37. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* (2012). doi:10.1038/nmeth.1923
38. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* (2010). doi:10.1101/gr.107524.110
39. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btq033
40. Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* (2018). doi:10.1038/s41586-018-0654-5
41. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* (2008). doi:10.1186/gb-2008-9-9-r137
42. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992 (2014).
43. R Core Team. R: A Language and Environment for Statistical Computing.
44. RStudio Team. RStudio: Integrated Development Environment for R. (2020).
45. Dowle, M. & Srinivasan, A. data.table: Extension of `data.frame`. (2019).
46. Bates, D. & Maechler, M. Matrix: Sparse and Dense Matrix Classes and Methods. (2018).
47. Bengtsson, H. matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors). (2018).
48. Henry, L. & Wickham, H. purrr: Functional Programming Tools. (2019).
49. Wickham, H. Reshaping Data with the {reshape} Package. *J. Stat. Softw.* **21**, 1–20 (2007).
50. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
51. Wilke, C. O. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. (2018).
52. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Syst.* (2006).
53. Melville, J. uwot: The Uniform Manifold Approximation and Projection (UMAP) Method for Dimensionality Reduction. (2020).
54. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).

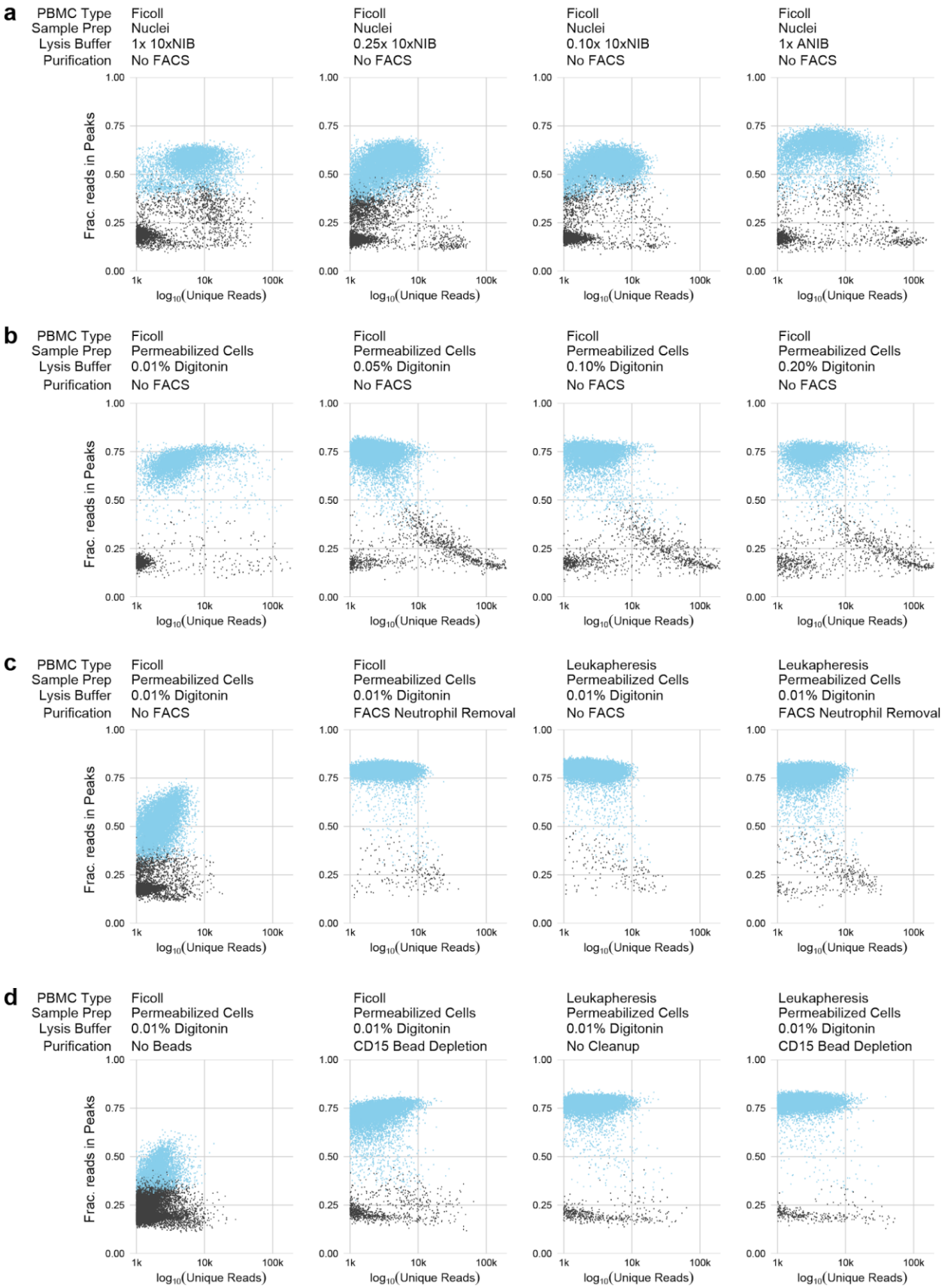
bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplemental Figures



Supplementary Figure 1. Nuclei were isolated from PBMCs using two buffer compositions and varying detergent concentrations. The two buffers, 10xNIB and ANIB, are defined in **Methods**, and the concentration of the detergent components of the buffers was varied for each sample. After isolation, nuclei were imaged using an EVOS M5000 Imaging System in transmitted light mode at 40x magnification to assess completeness of isolation and the condition of nuclei. Ideally, nuclei will not retain cell membranes and will have round, well-defined edges with minimal blebbing.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



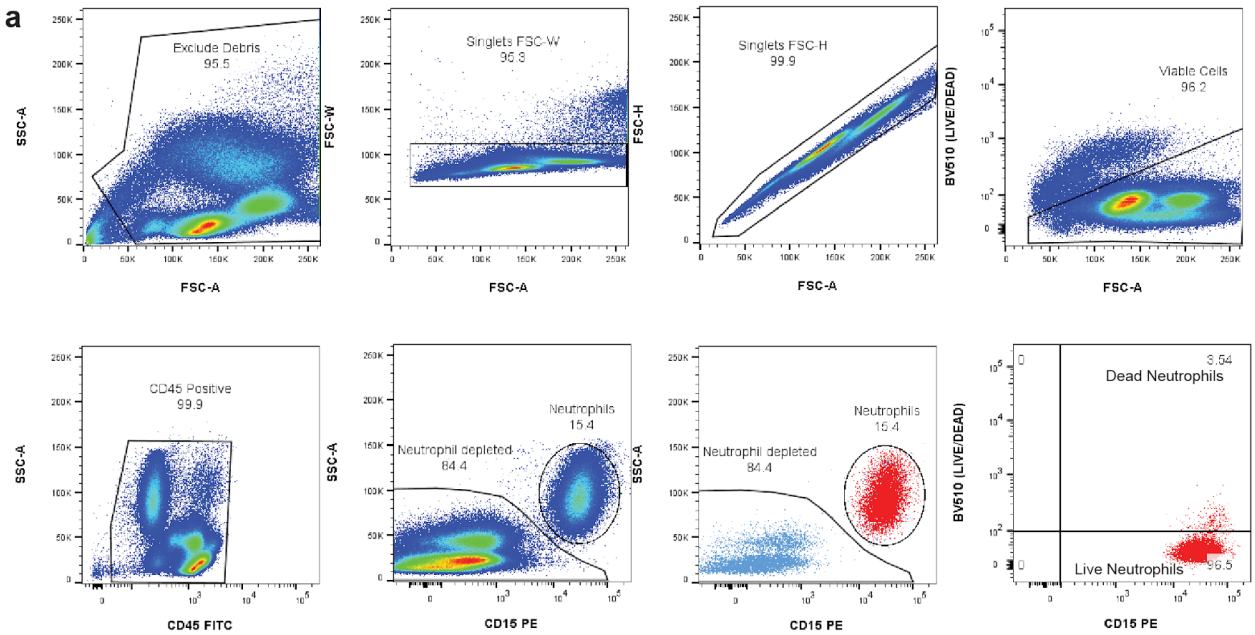
Supplementary Figure 2. Quality control plots for snATAC-seq and scATAC-sec experimental conditions. Each plot compares the quantity of data obtained from each cell (x-axis, log-scaled unique

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

fragment counts) against signal-to-noise as measured by fraction of reads in peaks (y-axis, FRIP). Each point represents a single 10x cell barcode. Blue points are barcodes that pass QC cutoffs (> 1,000 unique fragments, FRIP > 0.2, FRITSS > 0.2, and fraction of reads in ENCODE index peaks > 0.5); Gray points are barcodes that fail QC. **a**, Comparisons of nuclei isolation buffer compositions. Components of 10xNIB and ANIB buffers are defined in **Methods**. **b**, Comparisons of cell permeabilization conditions by varying the concentration of digitonin. Additional buffer components are defined in **Methods**. **c**, Comparisons of neutrophil removal by FACS on Ficoll-purified PBMCs (left 2 panels) and leukapheresis-purified PBMCs (right 2 panels). Neutrophil removal was performed using the gating scheme shown in **Supplementary Fig. 3**. **d**, Comparisons of neutrophil removal by anti-CD15 magnetic bead depletion on Ficoll-purified PBMCs (left 2 panels) and leukapheresis-purified PBMCs (right 2 panels).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

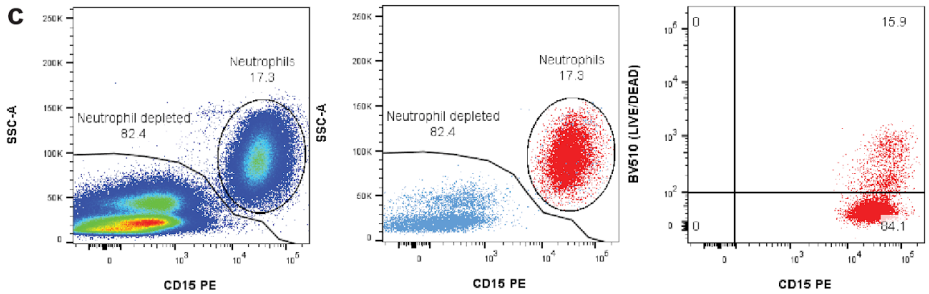
a



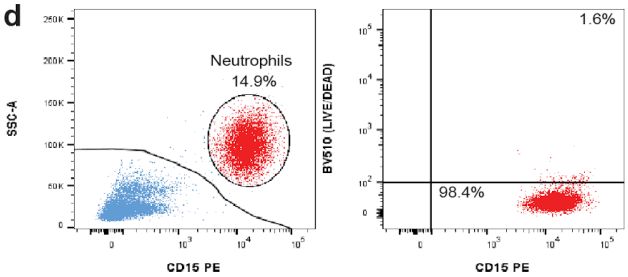
b

	% of parent	% of total
All events	---	100.0
Exclude Debris	95.5	95.5
Singlets FSC-W	95.3	91.0
Singlets FSC-H	99.9	90.9
Viable Cells	96.2	87.5
CD45 Positive	99.9	87.4
Neutrophils	15.4	13.5
Live Neutrophils	96.5	13.0
Dead Neutrophils	3.54	0.48
Neutrophil depleted	84.4	73.7

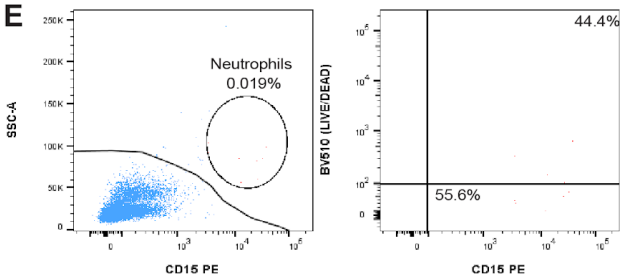
c



d



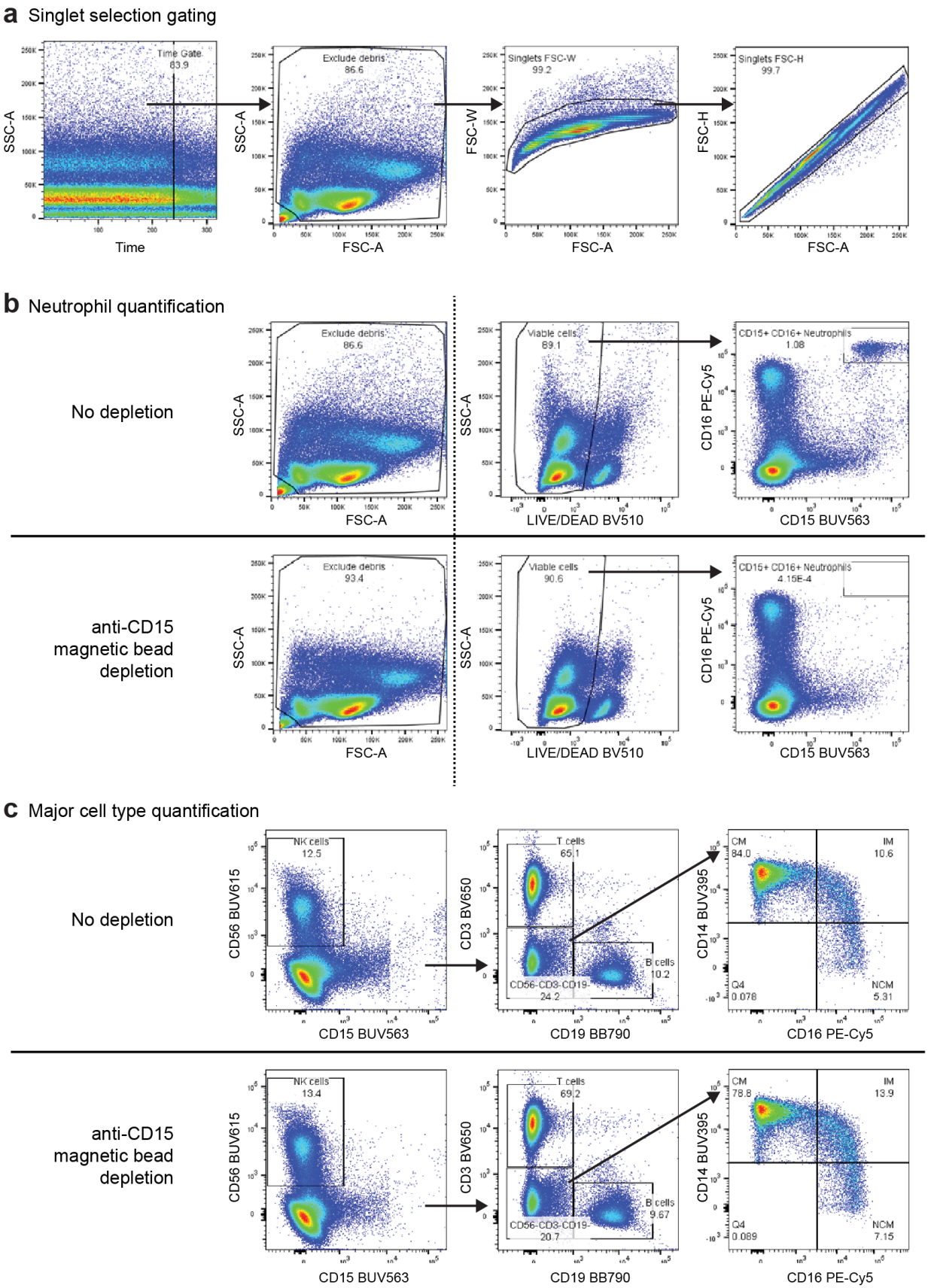
E



bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 3. a, Gating strategy for FACS removal of debris, doublets, dead cells, and neutrophils. **b**, Gating hierarchy for FACS neutrophil depletion shown in **a**, with demonstration data used to provide percentages of parent gates and total cells. **c**, Pre-sort collection sequential gating shows the live/dead proportion of total neutrophils. During FACS, dead neutrophils are removed using the live/dead gate and CD15/SSC gate leading to an underestimation of the total neutrophil content in the sample. **d**, Post-sort analysis of neutrophil content after debris, doublet, and dead cell removal. Left panel displays the percentage of viable singlets in the neutrophil gate (CD45+ gate in **b**). The right panel displays the percentage neutrophils that are live (bottom-right) and dead (top-right). **e**, Post-sort analysis of neutrophil content and viability after neutrophil exclusion sort (Neutrophil depleted gate in **b**).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

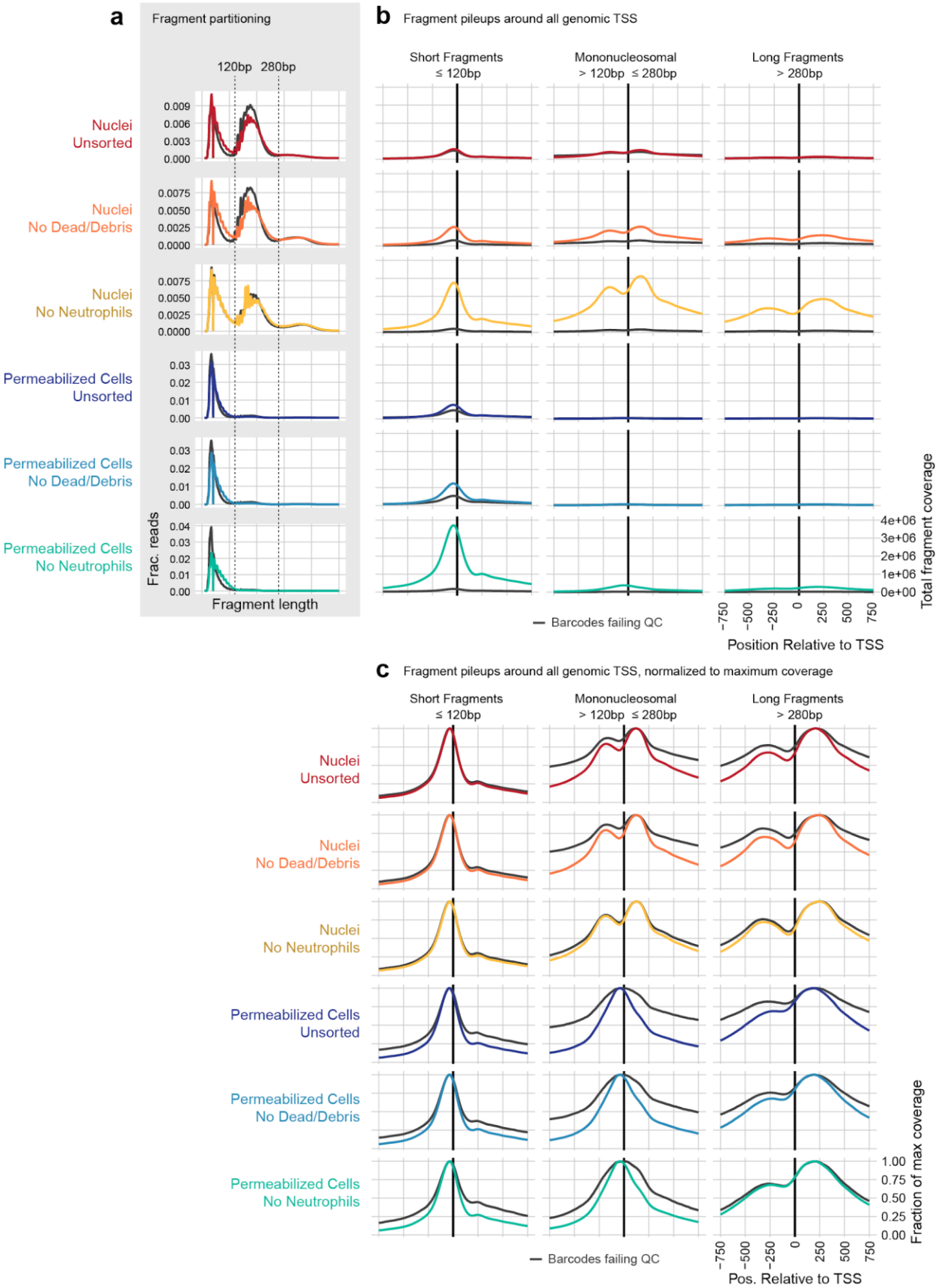


Supplementary Figure 4. Gating strategy for the 8-color flow cytometry panel used to evaluate anti-CD15 bead-based neutrophil removal. **a**, Preparatory gating used to select high-quality events (plot 1),

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

exclude debris (plot 2), and select singlet events (plots 3 and 4) **b**, Comparison of select gated populations of the PBMC sample before (top) and after depletion (bottom). The dashed line indicates that events from the left panel were not the parent gates for the two plots to the right, which were filtered first by the singlet gates described in **a**. Dead cells and debris were slightly reduced following depletion (Left-most plots). CD15+/CD16+ neutrophils were depleted from 1.080% to < 0.001% following depletion (right-most plots). **c**, Gating strategy for identification of major cells types in the PBMC sample before (top) and after depletion (bottom). Following time, debris, singlet, live/dead, and CD45+ gating, NK cells were defined as CD15-/CD56+, T cells were defined as CD3+/CD19-, and B cells were defined as CD3-/CD19+. Cells in the CD3-/CD19-/CD56- gate were used to identify monocyte subsets. Classical monocytes were defined as CD14+/CD16-, intermediate monocytes were defined as CD14+/CD16+, and non-classical monocytes were defined as CD14-/CD16+. We observed modest depletion of some classical monocytes (CM, from 84.0% to 78.8% of monocytes) with corresponding increases in the proportion of Intermediate (IM) and Non-classical (NCM) monocytes after bead depletion.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

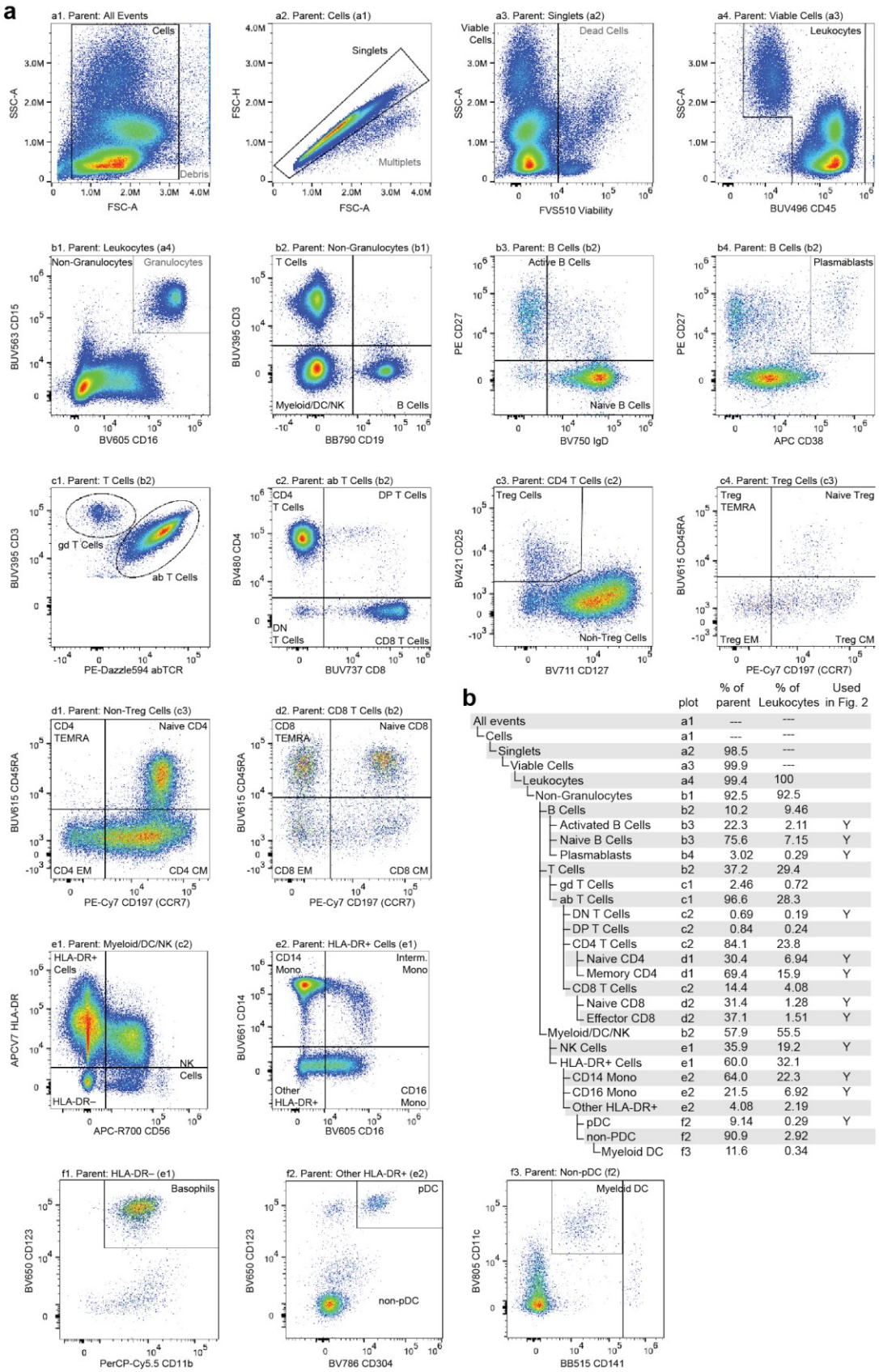


Supplemental Figure 5. a, To compare the contribution of nucleosomal fragments to signal near genomic transcription start sites, we partitioned unique, aligned fragments into 3 categories: Short

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

subnucleosomal fragments (≤120bp), Mononucleosomal fragments (> 120bp and ≤ 280 bp), and longer fragments including di- and tri-nucleosomal fragments (> 280 bp). The distribution of fragment lengths for each sample is displayed as a line plot, where colored lines show the distribution of cell barcode fragments, and gray lines show the distribution of fragments from non-cell barcodes. **b**, All fragments that overlapped a window around all genomic TSS ± 2,000bp were assembled, and the overlaps of fragments in each of the 3 size categories defined in (a) are plotted relative to the TSS ± 750bp. Axis labels shown on the bottom-right panel apply to all plots in (b). Nucleosomal fragments flanking the TSS are clearly visible for the Nuclei: No Neutrophils sample (third row, gold). These are less frequent in all Permeabilized Cell samples (rows 4-6, blue colors). **c**, as for (b), but each line has been scaled by dividing by the maximum coverage value for that line to make the shape of each distribution visible. Here, we can see nucleosomal positioning in each of the Nuclei samples (top 3 rows, warm colors), and a lack of mononucleosmal positioning in Permeabilized Cell samples (bottom 3 rows, blue colors). However, longer flanking fragments (third column) have a similar qualitative shape in both Nuclei and Permeabilized cells.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



b

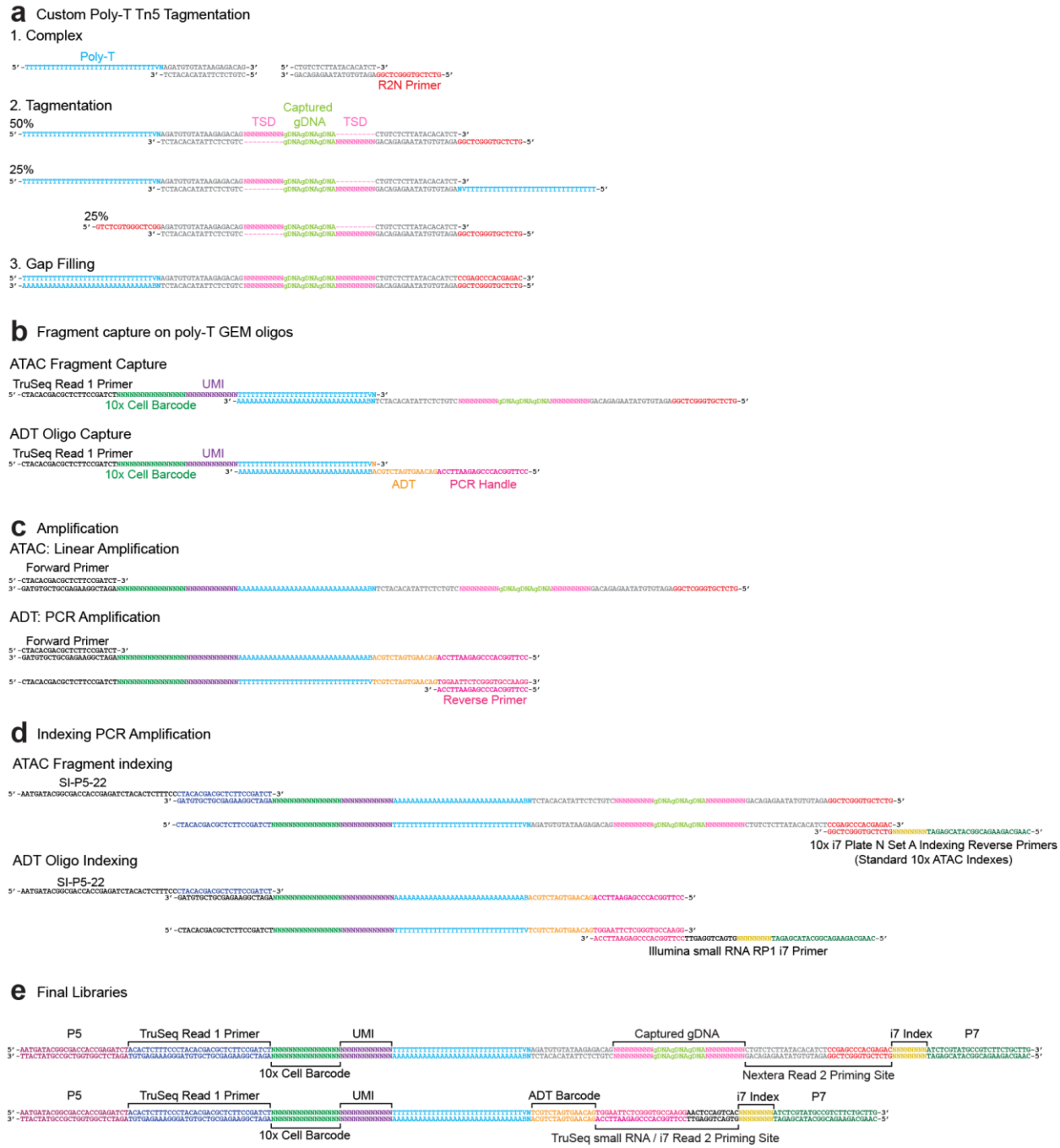
	plot	% of parent	% of Leukocytes	Used in Fig. 2
All events	a1	---	---	
Cells	a1	---	---	
Singlets	a2	98.5	---	
Viable Cells	a3	99.9	---	
Leukocytes	a4	99.4	100	
Non-Granulocytes	b1	92.5	92.5	
B Cells	b2	10.2	9.46	
Activated B Cells	b3	22.3	2.11	Y
Naive B Cells	b3	75.6	7.15	Y
Plasmablasts	b4	3.02	0.29	Y
T Cells	b2	37.2	29.4	
gd T Cells	c1	2.46	0.72	
ab T Cells	c1	96.6	28.3	
DN T Cells	c2	0.69	0.19	Y
DP T Cells	c2	0.84	0.24	
CD4 T Cells	c2	84.1	23.8	
Naive CD4	d1	30.4	6.94	Y
Memory CD4	d1	69.4	15.9	Y
CD8 T Cells	c2	14.4	4.08	
Naive CD8	d2	31.4	1.28	Y
Effector CD8	d2	37.1	1.51	Y
Myeloid/DC/NK	b2	57.9	55.5	
NK Cells	e1	35.9	19.2	Y
HLA-DR+ Cells	e1	60.0	32.1	
CD14 Mono	e2	64.0	22.3	Y
CD16 Mono	e2	21.5	6.92	Y
Other HLA-DR+	e2	4.08	2.19	
pDC	f2	9.14	0.29	Y
non-PDC	f2	90.9	2.92	
Myeloid DC	f3	11.6	0.34	

Supplemental Figure 6. a, Flow cytometry gating used to classify and quantitate cell type abundance. Cells were stained with a panel of 25 fluorescently labeled antibodies in total, fluorescence was measured using a Cytek Aurora spectral flow cytometer, and events were manually gated to assess cell

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

type abundance. Pairwise feature plots show the gates used for cell type assessment. Plots are arranged in rows (labeled a-f) and columns (labeled 1-4). The plot and label of the parent gate is at the top of each plot, and gates derived from each set of markers are labeled within the plots. **b**, Gating hierarchy used to quantify cell types, with references to the plot in which each gate is defined. Cell type labels used to generate reference proportions in **Fig. 2** are specified in the last column.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

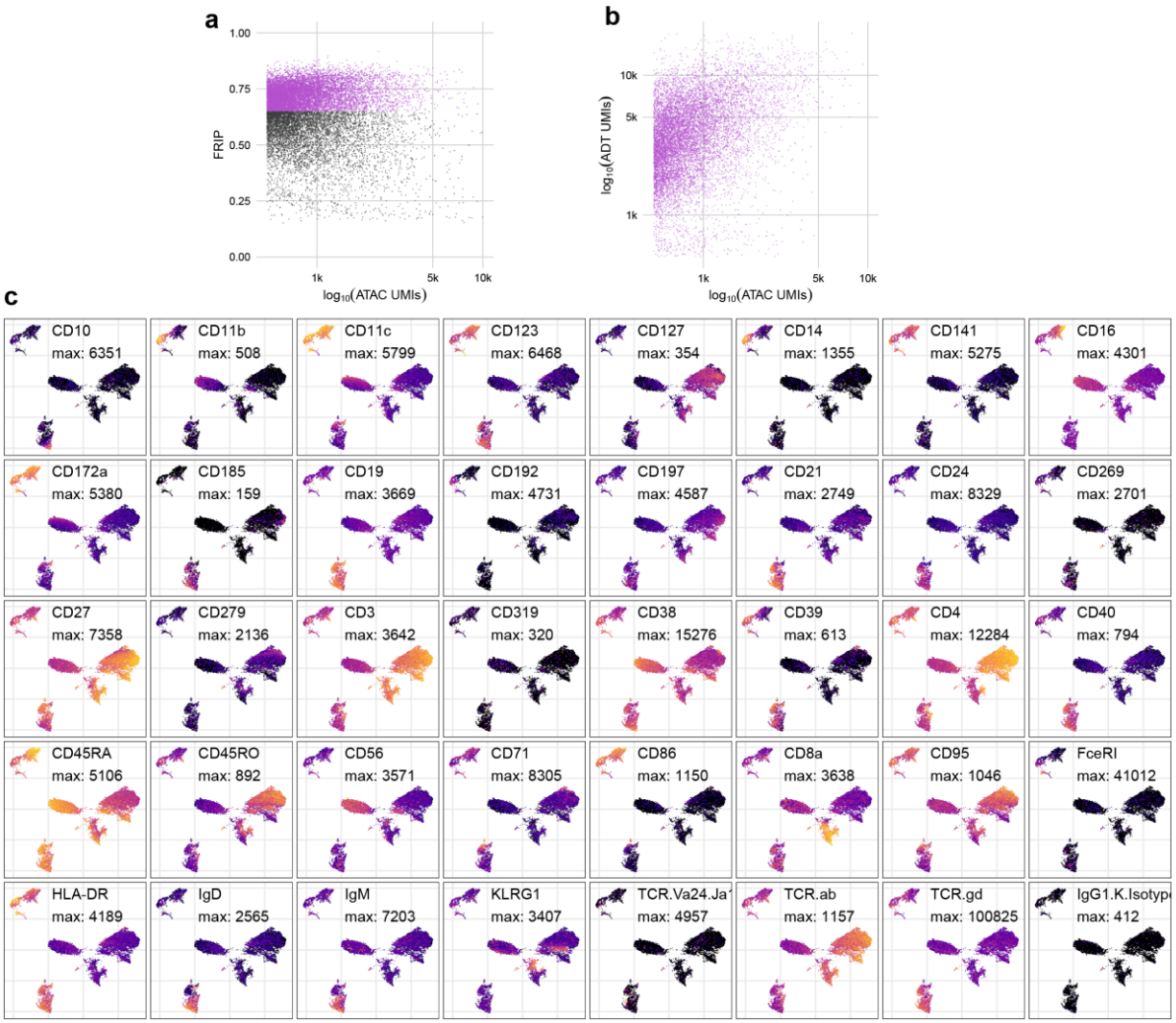


Supplemental Figure 7. Sequence-level ICICLE-seq design and workflow. **a**, Tn5 transposase complexes are formed by annealing Poly-T and Nextera R2N top oligos to MOSAIC-end bottom oligos to form double stranded DNA complexes. Complexes are bound to Tn5 transposase in the mosaic-end regions and transposed into target gDNA via a tagmentation reaction. Approximately half of the resulting ATAC fragments have both a Poly-T and R2N overhang and are capable of being barcoded and indexed in subsequent amplification reactions. **b**, Following GEM generation and gap-filling, DNA fragments are denatured to allow annealing to 10x Genomics 3' bead oligos and extension by a DNA polymerase, resulting in the addition of cell barcode and UMI sequences. **c**, Barcoded ATAC fragments are amplified linearly using a forward primer, while barcoded ADT fragments are amplified exponentially

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

by PCR through the addition of an ADT specific reverse primer. **d**, Intermediate libraries are amplified and Illumina P5 and indexed P7 adapter sequences are added by PCR. Prior to amplification, libraries are divided into separate ATAC and ADT reactions and amplified using library-specific P7 primers. **e**, Final ATAC and ADT library structure includes all components required for sequencing. The TruSeq Read 1 Primer reads the 10x Cell Barcode and UMI for both libraries as Read 1, followed by readout of the i7 Index sequence with the Nextera Read 2 Priming Site (ATAC) or TruSeq small RNA /i7 Read 2 Priming site (ADT). After strand switching, these two latter binding sites are used to read out the captured gDNA sequence (ATAC) or ADT Barcode as Read 2.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283887; this version posted September 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Supplementary Figure 8. a, QC metrics plot for the ATAC-seq component of ICICLE-seq. Cell barcodes with FRIP > 0.65 (purple points) were selected for downstream analysis. **b**, Comparison of the number of ATAC UMIs (x-axis) to ADT UMIs (y-axis) for each cell barcode that passed the QC cutoff in (**a**). **c**, UMAP plots based on ADT UMI counts. Each panel shows log-transformed expression of a different marker, and values scaled between 0 (black) and the maximum observed value (yellow).