nature biotechnology

Check for updates

Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells

Eleni P. Mimitou[1,15], Caleb A. Lareau [2,3,4,15], Kelvin Y. Chen [5,6,15], Andre L. Zorzetto-Fernandes[7], Yuhan Hao[8,9], Yusuke Takeshima [5], Wendy Luo [3,4], Tse-Shun Huang [7], Bertrand Z. Yeung[7], Efthymia Papalexi[8,9], Pratiksha I. Thakore[3], Tatsuya Kibayashi[5], James Badger Wing [5,10], Mayu Hata[5], Rahul Satija [8,9], Kristopher L. Nazor[7], Shimon Sakaguchi[5,6,16], Leif S. Ludwig [3,4,11,16], Vijay G. Sankaran [3,4,16], Aviv Regev [3,9,12,13,14,16] and Peter Smibert [1,16] ✉

Recent technological advances have enabled massively parallel chromatin profiling with scATAC-seq (single-cell assay for transposase accessible chromatin by sequencing). Here we present ATAC with select antigen profiling by sequencing (ASAP-seq), a tool to simultaneously profile accessible chromatin and protein levels. Our approach pairs sparse scATAC-seq data with robust detection of hundreds of cell surface and intracellular protein markers and optional capture of mitochondrial DNA for clonal tracking, capturing three distinct modalities in single cells. ASAP-seq uses a bridging approach that repurposes antibody:oligonucleotide conjugates designed for existing technologies that pair protein measurements with single-cell RNA sequencing. Together with DOGMA-seq, an adaptation of CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) for measuring gene activity across the central dogma of gene regulation, we demonstrate the utility of systematic multi-omic profiling by revealing coordinated and distinct changes in chromatin, RNA and surface proteins during native hematopoietic differentiation and peripheral blood mononuclear cell stimulation and as a combinatorial decoder and reporter of multiplexed perturbations in primary T cells.

The recent explosion of technologies allowing detailed phenotypic measurements of single cells in high throughput has made the dissection of cell types and states in complex tissues widely accessible. Although measurement of single modalities has been highly informative for phenotyping, new techniques that allow detection of multiple modalities of information from single cells continue to be developed[1–4].

Multimodal approaches can couple sparse comprehensive measurements with more robust directed measurements that report on known cell types or states. For example, CITE-seq[5,6] and RNA expression and protein sequencing (REAP-seq)[7] couple the relatively sparse single-cell RNA sequencing (scRNA-seq) signal with robust detection of highly abundant and well-characterized surface proteins, using oligo-labeled antibodies, and enable more robust cell type discrimination. However, whereas mRNA and protein are the products of gene expression, their detection does not suffice to decipher underlying gene regulatory mechanisms.

Measurement of chromatin architecture allows detection of the earliest cellular states and responses to stimuli or developmental

decisions[8]. In particular, scATAC-seq is a widely used method to obtain a genome-wide snapshot of chromatin accessibility, signatures of active transcription and transcription factor (TF) binding[9,10]. Several single-cell methods developed to simultaneously capture mRNA with chromatin accessibility help correlate chromatin accessibility with gene expression and layer mRNA expression data on top of sparse ATAC-seq data[8,11–13] but do not report on established protein markers that, in combination, precisely define cell types and cell states.

Here we report ASAP-seq, a method that enables robust detection of surface and intracellular proteins using oligo-labeled antibodies together with scATAC-seq. ASAP-seq takes advantage of existing antibody reagents used for CITE-seq and related technologies, circumventing the need for specialized components. Notably, unlike co-assays of RNA and chromatin, where there is a tradeoff between enzymatic steps with different requirements, we leverage the enzymatic steps of the parent assay to detect multiple modalities and ensure high quality across both. ASAP-seq is compatible with quantification of intracellular proteins and mitochondrial DNA

[1]Technology Innovation Lab, New York Genome Center, New York, NY, USA. [2]Department of Pathology, Stanford University, Stanford, CA, USA. [3]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]Division of Hematology / Oncology, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [5]Department of Experimental Immunology, Immunology Frontier Research Center, Osaka University, Osaka, Japan. [6]Department of Experimental Pathology, Institute for Frontier Medical Sciences, Kyoto University, Kyoto, Japan. [7]BioLegend, Inc., San Diego, CA, USA. [8]Center for Genomics and Systems Biology, New York University, New York, NY, USA. [9]New York Genome Center, New York, NY, USA. [10]Laboratory of Human Immunology (Single Cell Immunology), Immunology Frontier Research Center, Osaka University, Osaka, Japan. [11]Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany. [12]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [13]Department of Biology and Koch Institute, Massachusetts Institute of Technology, Cambridge, MA, USA. [14]Present address: Genentech, South San Francisco, CA, USA. [15]These authors contributed equally: Eleni P. Mimitou, Caleb A. Lareau, Kelvin Y. Chen. [16]These authors jointly supervised this work: Shimon Sakaguchi, Leif S. Ludwig, Vijay G. Sankaran, Aviv Regev, Peter Smibert. ✉e-mail: smibertp@gmail.com
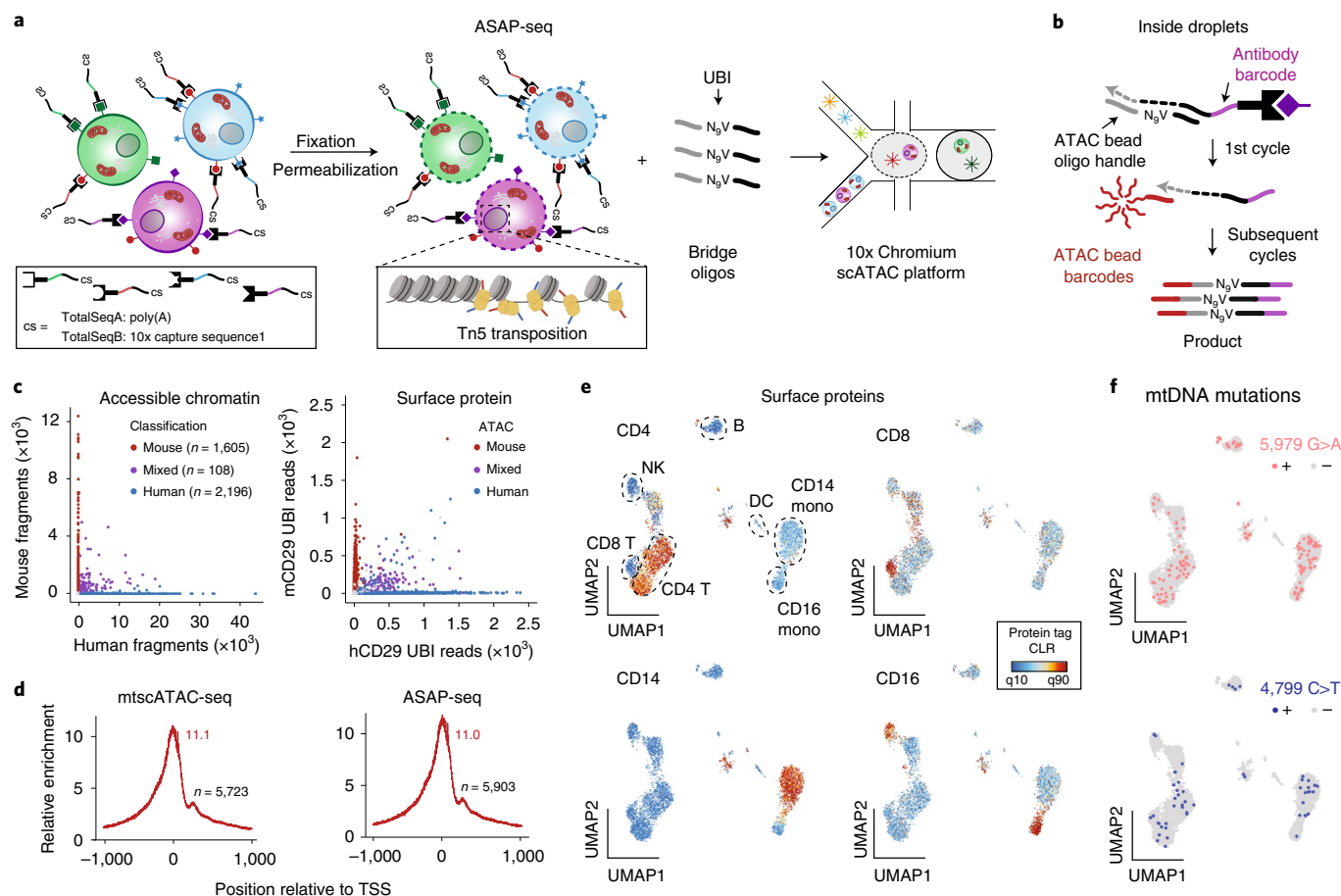
Fig. 1 | ASAP-seq incorporates protein detection in scATAC-seq workflows. a, Schematic of the cell processing steps that allow retention and joint profiling of cell surface markers with chromatin accessibility. Cells are stained with oligo-conjugated antibodies before fixation, permeabilization and transposition with Tn5. **b**, In droplets, bridge oligos spiked into the barcoding mix promote templated extension of the antibody tags during the first cycle of amplification, rendering them complementary to bead-derived barcoding oligos. Extended antibody tags are subsequently barcoded together with the transposed chromatin fragments. **c**, Species mixing experiment using the Pre-SPRI approach; number of unique nuclear fragments (left) and protein tag counts (right) associated with each cell barcode. Points are colored based on species classification using ATAC-derived fragments (97.4% agreement by assignment; all but one discrepancy was an errant doublet versus singlet classification) **d**, TSS enrichment scores of mtscATAC-seq without (left) or with (right) concomitant protein tag capture. *n* indicates the number of cells profiled. **e**, UMAP of chromatin accessibility-based clustering of PBMCs stained with a nine-antibody panel, with selected markers highlighted. Color bar: protein tag CLR values. **f**, Cellular distribution of the two most common mtDNA mutations in the population. Thresholds for + were 5% heteroplasmy based on empirical density.

(mtDNA) genotyping[14,15]. To demonstrate its utility, we applied ASAP-seq to the study of human hematopoiesis, where the combination of single-cell chromatin accessibility, surface marker profiles and mtDNA-based lineage tracing allowed us to resolve bone marrow heterogeneity. Separately, in a model of immune cell stimulation, we combined ASAP-seq with CITE-seq to reveal the distinct layers of regulation of chromatin accessibility, mRNA levels and protein. We further applied ASAP-seq in a multiplexed T cell perturbation assay to disentangle chromatin and protein phenotypes associated with specific signaling pathways. Finally, we introduce DOGMA-seq, a variant of CITE-seq, allowing co-measurement of chromatin accessibility, gene expression and protein from the same cells.

Results

Development and validation of ASAP-seq. To develop ASAP-seq, we built on mtscATAC-seq[14], a droplet-based scATAC-seq method that jointly profiles chromatin accessibility and mtDNA in fixed cells. We reasoned that the fixation and permeabilization before

Tn5 transposition would also retain cell surface markers, enabling their detection with oligo-conjugated antibodies (Fig. 1a)[5–7]. To test this, we stained peripheral blood mononuclear cells (PBMCs) and control beads with fluorophore-labeled antibodies against CD19, CD11c and CD4 and performed flow cytometry at subsequent protocol steps (Extended Data Fig. 1a). Although permeabilization of fixed cells had a minimal effect on signal intensity, incubation at 37 °C to mimic the transposition step led to a minor loss of intensity. Notably, the separation to background remained distinct on both cells with high (CD4 T cells) and moderate (CD4 on monocytes) levels of target proteins, suggesting sensitivity over a range of protein abundances.

We next devised a molecular bridging approach for protein detection that capitalizes on the large existing catalog of commercial antibody:oligo-conjugated products designed for scRNA-seq applications (TotalSeq products by BioLegend). In this approach, a short 3′ blocked oligo serves as a template to extend the antibody tag during the initial amplification cycles. The extended product subsequently anneals to the bead-derived barcoded oligo and becomes
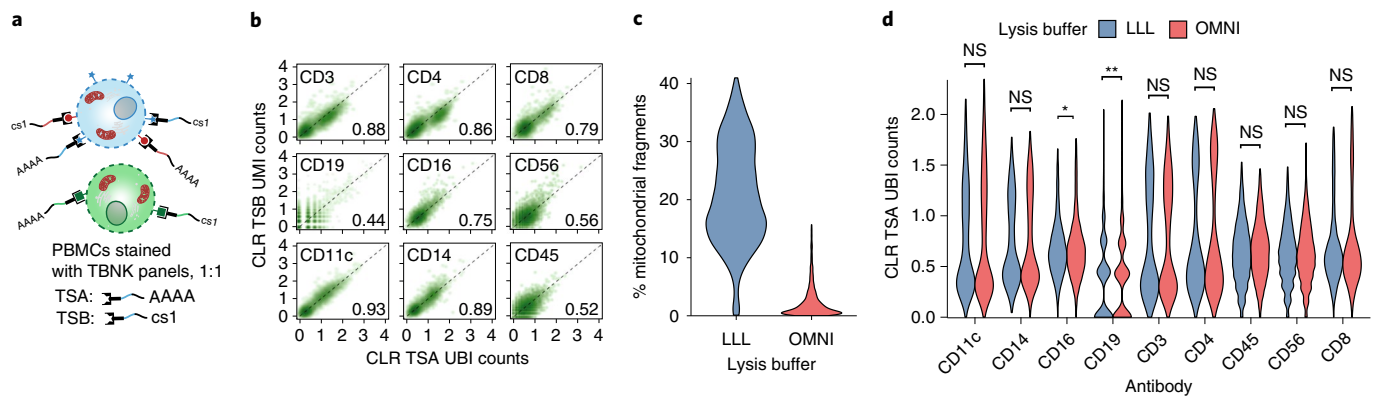
Fig. 2 | ASAP-seq enables a modular and versatile multi-omics toolkit. a, Schematic of experimental design. PBMCs were stained with TBNK panels of the TSA or TSB format at a 1:1 ratio, followed by fixation and permeabilization under mild (LLL) or strong (OMNI) conditions. **b,** Pairwise comparison of CLR-normalized TSA and TSB counts for indicated antibodies under mild lysis conditions ($n = 4,748$ cells). Counts were collapsed for unique molecules using UBIs (TSA panel) or UMIs (TSB panel). **c,** Distribution of percent of mtDNA fragments retained in the library under the two lysis conditions. **d,** Comparison of CLR-normalized TSA counts for indicated proteins under the two tested lysis conditions. Statistical comparisons are two-sided Wilcoxon rank-sum tests with Bonferroni-adjusted P values (NS, not significant; *$P_{adj} = 0.0002$ and **$P_{adj} = 8 \times 10^{-14}$).

linearly amplified along with accessible chromatin fragments (Fig. 1b and Extended Data Fig. 1b). As each antibody-derived oligo can be bridged only once, we introduced a unique bridging identifier (UBI) sequence via the bridge oligo for TotalSeq-A (BOA) to allow tag molecule counting when TotalSeq-A (TSA) products are used (Extended Data Fig. 1b). Alternate TotalSeq formats (for example, TotalSeq-B (TSB)) already contain UMI sequences[16] and do not require UBIs in their bridge oligos (Extended Data Fig. 2a).

To benchmark ASAP-seq, we stained a mix of human (HEK-293T) and mouse (NIH-3T3) cells with species-specific anti-CD29 antibodies (Supplementary Table 1, 'mixed species' tab), followed by fixation, permeabilization and transposition before barcoding in droplets in the presence of BOA (Methods). Species assignment by either modality yielded consistent results, demonstrating the specificity of protein detection in this assay (Fig. 1c, Extended Data Fig. 1c and Methods). Two identical barcoding reactions were run in separate lanes to test tag library preparation using alternate input—10% of either the purified fragments after emulsion breakage (Pre-SPRI) or the purified supernatant fraction after the SPRI purification step (Post-SPRI) (Methods). In both instances, we observed no substantial changes in ATAC fragment or protein tag complexity, suggesting that either fraction can be used to prepare the tag libraries (Extended Data Fig. 1d,e).

To perform additional benchmarking and optimization, we applied ASAP-seq to PBMCs stained with a TBNK panel (Supplementary Table 1, 'TBNK' tab) comprising nine major immune cell markers. The transcription start site (TSS) score and chromatin fragment complexity were similar to an unstained sample (Fig. 1d and Extended Data Fig. 1f). Furthermore, projection of antibody tag counts on cell types annotated by their chromatin accessibility profiles showed expected patterns, including mutual exclusivity of CD4 and CD8 expression in T cells, CD16 in natural killer (NK) cells and a subset of monocytes and CD14 in a non-overlapping set of monocytes, which generally corroborated chromatin activity scores at these loci (Fig. 1e and Extended Data Fig. 1h–j). As fixation and permeabilization conditions retain mitochondria[14], we further recover 31% mtDNA reads, allowing us to profile mtDNA mutations jointly with protein levels and chromatin accessibility (Fig. 1f).

Finally, we further expanded the utility of ASAP-seq by incorporating Cell Hashing[17,18]. In Cell Hashing[17,18] and related methods[19–22], sample multiplexing is enabled by barcoded oligo tags

(hashtags) that are attached by a variety of means to all cells of a specific sample, revealing both the sample identity and the presence of cross-sample multiplets. To demonstrate this, we stained PBMCs with four TSA hashing antibodies (Supplementary Table 1, 'Hashing' tab) and recovered 13,772 cells that were successfully demultiplexed in four distinct populations, with 1,396 detected doublets, consistent with the expected number of 1,138 doublets derived from a Poisson-based model (Extended Data Fig. 1g).

ASAP-seq is a modular toolkit. Next, we determined if UBIs, used in bridging TSA family antibody:oligo conjugates, perform similarly to UMIs (in TSB products). We designed 10-nt UBIs with complexity approaching or exceeding the UMI complexity commonly used in scRNA-seq[23,24]. To formally compare UMI versus UBI quantification, we co-stained PBMCs with a 1:1 ratio of TSA (UBI-based) and TSB (UMI-based) TBNK panel (Fig. 2a and Supplementary Table 1, 'TBNK' tab) and added both bridge oligos during the barcoding step (bridging schemes shown in Extended Data Figs. 1b and 2a). UBI-collapsed TSA counts showed good correlation with UMI-collapsed TSB counts across all nine antibodies (Pearson's $r = 0.44$–0.93, depending on antibody), verifying that the UBI can provide a reliable proxy for a UMI (Fig. 2b).

Although ASAP-seq directly extends mtscATAC-seq, we sought to modulate mtDNA enrichment. To address this, we compared the original OMNI-ATAC-seq lysis protocol previously shown to deplete mtDNA[25], and currently recommended for the 10x Genomics scATAC-seq assay, to the effects of lysis conditions for mtscATAC-seq[14] (Methods). We fixed PBMCs stained with the TBNK panel (Fig. 2a), split into two aliquots, and lysed with the mild mtscATAC-seq conditions (referred to as low-loss lysis (LLL)) or with the stronger OMNI conditions. Although the lysis condition had a dramatic effect on mtDNA retention (Fig. 2c), we observed minimal effects on the distribution of UBI or UMI tag counts (Fig. 2d and Extended Data Fig. 2c). Moreover, the correlations between UBI- and UMI-collapsed tag counts under stronger permeabilization are similar to those in milder lysis conditions (Extended Data Fig. 2b; Pearson's $r = 0.38$–0.96), albeit with slight improvement for most antibodies. Overall, we conclude that the surface marker retention in ASAP-seq allows reliable measurement, irrespective of antibody:oligo reagent type or lysis conditions used, and without compromising ATAC-seq data (Extended Data Fig. 2d–f).
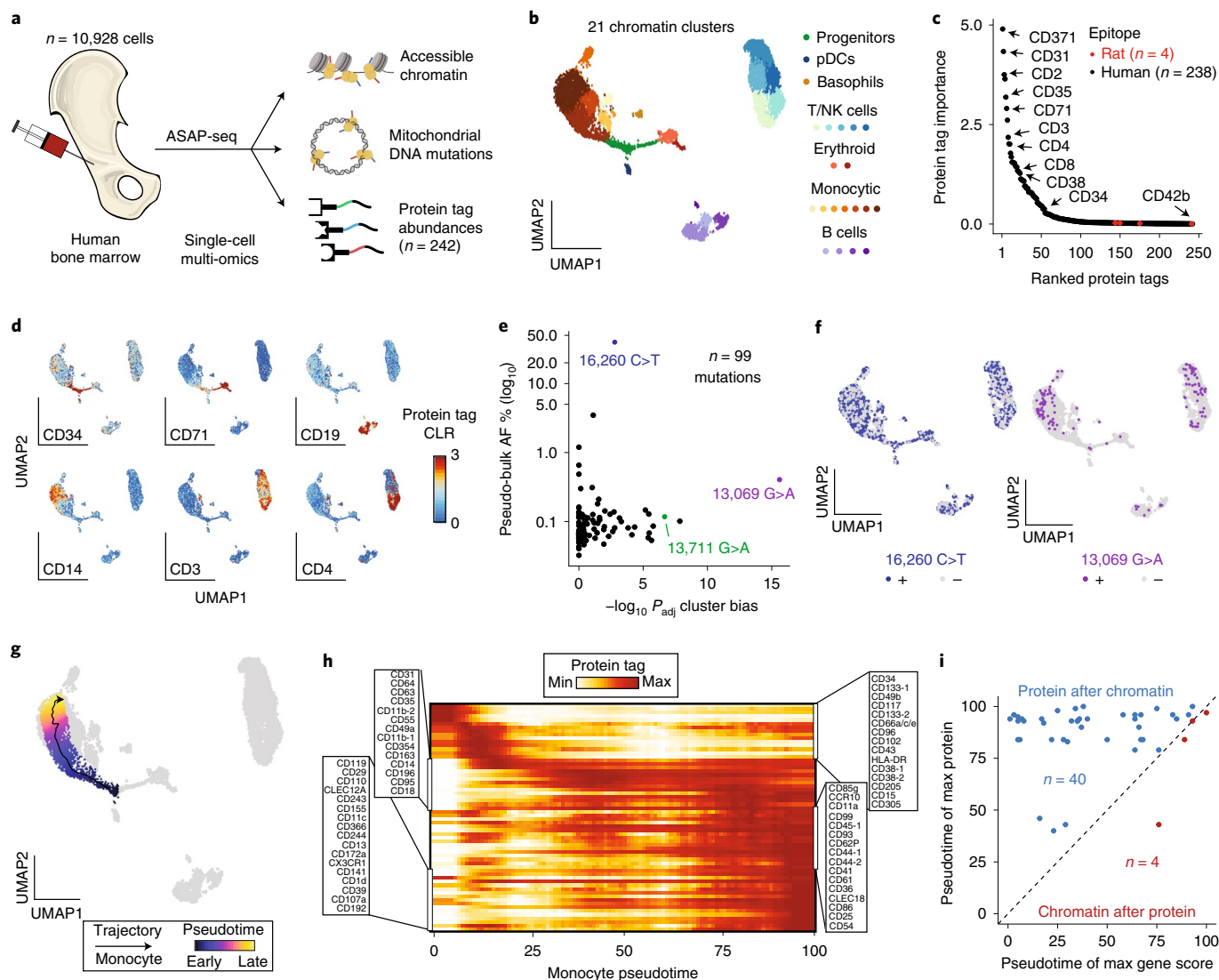
Fig. 3 | Dissection of native human hematopoiesis with multimodal cell state inference and mtDNA-based lineage tracing. a, Schematic of experimental design. Whole human bone marrow mononuclear cells (BMMCs) were stained with hashtag antibodies and a 242-antibody panel for ASAP-seq processing. **b**, Reduced dimension representation and cell clustering of high-quality cells (n=10,928) inferred using chromatin accessibility. **c**, Rank sorting of informative protein tags in distinguishing cell cluster identification. Negative controls (rat epitopes) are shown in red. **d**, Characterization of cell populations for six selected markers. **e**, Characterization of 99 somatic mtDNA mutations identified in the BMMCs. Selected mutations enriched for lineage bias (13,069 G>A and 13,711 G>A; x axis; Extended Data Fig. 3e) and highest for allele frequency (AF) (16,260 C>T; y axis) are highlighted. **f**, Projection of highlighted mutations from **e** on the reduced dimension space. Thresholds for + were 50% for 16,260 C>T and 5% for 13,069 G>A based on empirical density. **g**, Developmental trajectory of monocyte differentiation using semi-supervised pseudotime analysis. **h**, Expression of cell surface markers along the developmental trajectory highlighted in **g**. Rows are min–max normalized. **i**, Comparison of maximum gene activity scores (x axis) and protein (y axis) during pseudotime. Each dot is a gene/surface protein pair. pDC, plasmacytoid dendritic cell.

ASAP-seq reveals cell state and cell lineage in bone marrow. The multimodal readout of ASAP-seq uniquely enables high-throughput profiling of epigenomic, proteomic and clonal features (through mtDNA) of cells from human tissue. We applied ASAP-seq to profile bone marrow mononuclear cells from a healthy 24-year-old donor, using a TSA antibody panel (n=242 markers; Supplementary Table 1, 'BM' tab) and six hashtags. We permeabilized cells under LLL conditions to retain mtDNA fragments and retained 10,928 high-quality cells (Fig. 3a and Methods). Dimensionality reduction and clustering of chromatin accessibility partitioned the cells into 21 distinct clusters spanning the major hematopoietic lineages (Fig. 3b and Supplementary Table 2). Notably, we did not remove predicted cell doublets as these were enriched for monocytic

progenitors, a real cell state/type present at the expected frequencies (Extended Data Fig. 3a,b and Methods).

To identify protein markers associated with cell subsets, we used a random forest model trained on cluster labels using the scaled antibody tag abundances[26] (Methods). The model rediscovered many classical hematopoietic lineage markers, including CD3, CD4 and CD8 in lymphoid cells; CD371 (*CLEC12A*) and CD2 in myeloid cells; CD71 (*TFRC*) in erythroid cells; and CD38 in more mature progenitor cells (Fig. 3c,d and Extended Data Fig. 3c). Similarly, TF activities correlated with surface marker abundance as expected, including GATA1 in erythroid cells and CEBPA in myeloid progenitors and monocytes (Extended Data Fig. 3d). Furthermore, we mapped our ASAP-seq data to a CITE-seq bone marrow reference[27]
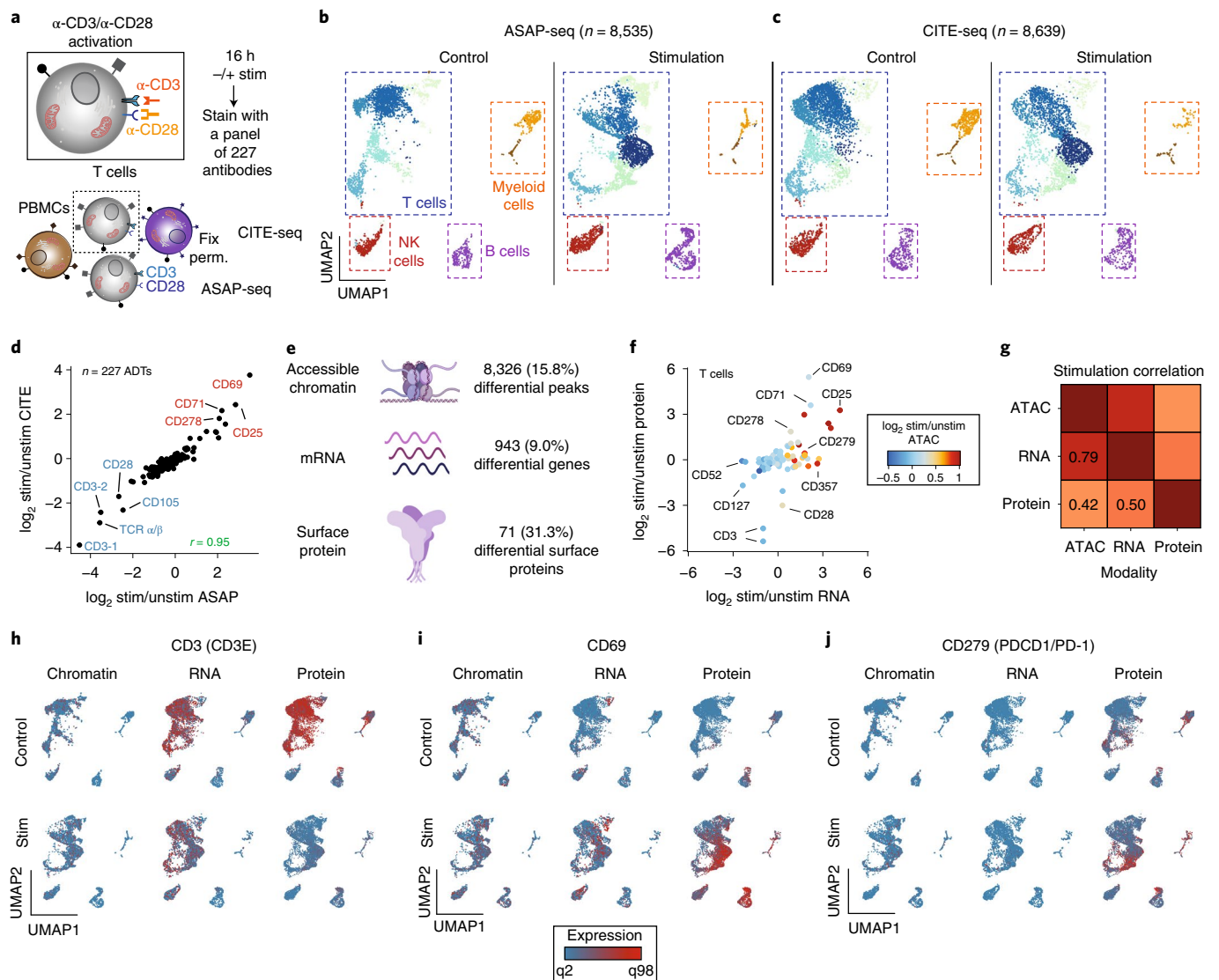
Fig. 4 | ASAP-seq and CITE-seq reveal coordinated and distinct changes in chromatin, RNA and protein levels. a, Schematic of the experimental design. PBMCs were incubated with (stimulation) or without (control) multimeric α-CD3/CD28 for 16 h, followed by staining with the 227-antibody panel. An aliquot of the cells was subjected to CITE-seq, whereas the remaining cells were fixed and subjected to ASAP-seq. **b**, **c**, Reduced dimension representations using data integration methods and UMAP for ASAP-seq (**b**) and CITE-seq (**c**) for both control (left) and stimulated (right) conditions. **d**, Correlation of surface marker fold changes (log$_2$) upon stimulation as detected by CITE-seq and ASAP-seq. Top upregulated markers are highlighted in red; downregulated markers are highlighted in blue. **e**, Schematic and summary of the number and proportion of differential features (chromatin accessibility peaks, genes and surface proteins) detected for T cells between the stimulation and control conditions. **f**, Summary of changes in chromatin accessibility, gene expression and surface protein abundance for 84 expressed genes during T cell stimulation. **g**, Pearson correlation between the log$_2$ fold changes for each modality as shown in **f**. **h**–**j**, UMAPs of single-cell chromatin accessibility, mRNA expression and surface protein levels for both the control (top) and stimulation (bottom) conditions shown on the reduced dimension space for CD3 (**h**), CD69 (**i**) and PD-1 (**j**).

and compared the cell populations defined by ATAC-seq only versus projected protein only, revealing that certain cell types poorly resolved in accessible chromatin space were readily resolved with the additional protein readout (Extended Data Fig. 3e).

Next, we used mtDNA genotypes for the clonal tracing of hematopoietic cells[14,28,29]. Using mgatk[14], we detected 99 heteroplasmic mtDNA mutations that were enriched for expected nucleotide substitutions[14] (Extended Data Fig. 3f). We used cell subset annotations to examine for putative lineage bias and detected lineage-restricted somatic mutations, such as 13,069 G>A and 13,711 G>A, that were relatively depleted in cells from the erythroid lineage but showed no predicted loss or gain of function (Fig. 3e,f and Extended Data Fig. 3g,h). Furthermore, one highly heteroplasmic variant, 16,260 C>T,

was present at ~40% heteroplasmy in the population yet was evenly distributed across the different lineages (Fig. 3e,f). Analysis of the donor mtDNA haplotype suggested that this mutation indeed arose somatically, potentially early during developmental hematopoiesis[30]. Overall, our observations support the utility of ASAP-seq to uncover mtDNA variants at single-cell resolution.

Dynamics of surface proteins during differentiation. We hypothesized that the integration of accessible chromatin and protein tags via ASAP-seq could refine our understanding of marker dynamics during the continuous lineage commitment and differentiation in hematopoiesis[31,32]. To examine this, we charted trajectories from CD34$^+$CD38$^-$ multipotent hematopoietic stem and progenitor cells

to differentiated monocytes (Fig. 3g) and erythroblasts (Extended Data Fig. 3j). Although markers associated with multipotent and other lineage progenitor cells, such as CD34 and CD49d (*ITGA4*), were downregulated early in the trajectory, monocyte markers, such as CD64 (*FCGR1A*) and CD31 (*PECAM1*), were quickly upregulated and persisted throughout differentiation (Fig. 3h and Supplementary Table 3). Conversely, CD11c (*ITGAX*) and CD371 (*CLEC12A*) were upregulated only toward the end of the trajectory. We also observed similar patterns of dynamic surface marker expression throughout erythroid differentiation (Extended Data Fig. 3k). Interestingly, among the proteins that were gained after commitment from the progenitor cluster, the increase in protein expression during monocyte differentiation was generally preceded by a gain of accessible chromatin at associated loci, which we corroborated in erythroid differentiation (Fig. 3i, Extended Data Fig. 3k–m and Methods). This result is consistent with a model where chromatin accessibility is the 'first mover' during differentiation, and the resultant changes in transcription prime cells for differentiation[8]. Taken together, our analyses showcase the versatility of ASAP-seq to measure multiple modalities of cell state alongside cell lineage, at greater surface marker diversity than in conventional cytometry approaches[33].

ASAP-seq and CITE-seq reveal three levels of regulation. ASAP-seq and CITE-seq are companion single-cell assays that combine highly multiplexed protein measurements with profiling epigenomic or transcriptional landscapes, respectively. We reasoned that the shared protein features can help connect scRNA-seq and scATAC-seq datasets.

We simultaneously applied ASAP-seq and CITE-seq to profile epigenomic, transcriptomic and proteomic changes after T cell stimulation. PBMCs were split into two aliquots: one stimulated with anti-CD3/CD28 and IL-2 for 16 h, the other cultured in the absence of stimulation (control), followed by staining of both with a TSA antibody panel (*n* = 227; Supplementary Table 1, 'PBMC' tab). Each of the samples was then split to run ASAP-seq and CITE-seq in parallel (Fig. 4a). We combined the RNA and ATAC profiles from the control and stimulated cells, revealing stimulation-dependent changes within T cells (Fig. 4b,c and Methods).

We next compared surface protein measurements by CITE-seq and ASAP-seq. As expected, we observed a decrease (~1.7–2×) in the tag molecule complexity in ASAP-seq compared to CITE-seq, consistent with the loss of signal observed in our initial flow cytometry tests (Extended Data Figs. 4a and 1a and Methods). However, both methods were highly concordant in the mean signal detected for variable proteins within each cluster (Extended Data Fig. 4a, lower panels), as well as the change in antibody signal stimulation across the panel (Pearson's *r* = 0.95; Fig. 4d), indicating that the cell processing-induced loss of sensitivity does not affect specific markers. Both assays detected substantial upregulation of canonical T cell activation markers, such as CD69, CD25, CD71 and CD278

(refs. [34–36]), at both the pseudo-bulk (Fig. 4d) and single-cell (Extended Data Fig. 4b,c) levels. Conversely, CD3 (protein $\log_2FC = -3.5$ and -4.5; $P < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test for ASAP-seq protein abundance), CD28 ($\log_2FC = -2.5$; $P < 2.2 \times 10^{-16}$) and T cell receptor (TCR) α/β ($\log_2FC = -2.9$; $P < 2.2 \times 10^{-16}$) antibody counts were noticeably reduced upon stimulation (Fig. 4d), likely due to internalization of the receptors upon triggering of the TCR complex[37]. An antibody prioritization approach using the random forest model[26] for ASAP-seq data verified that these markers were most associated with the stimulation (Extended Data Fig. 4d and Methods). Finally, embedding cells by protein abundance profiles intermixed cells profiled by the two assays, albeit with reduced separation of the activated T cell state (Fig. 4e and Extended Data Fig. 4e–g). Taken together, despite a lower tag complexity, ASAP-seq is similarly capable of capturing protein abundance associated with cell state and dynamic changes as measured with CITE-seq.
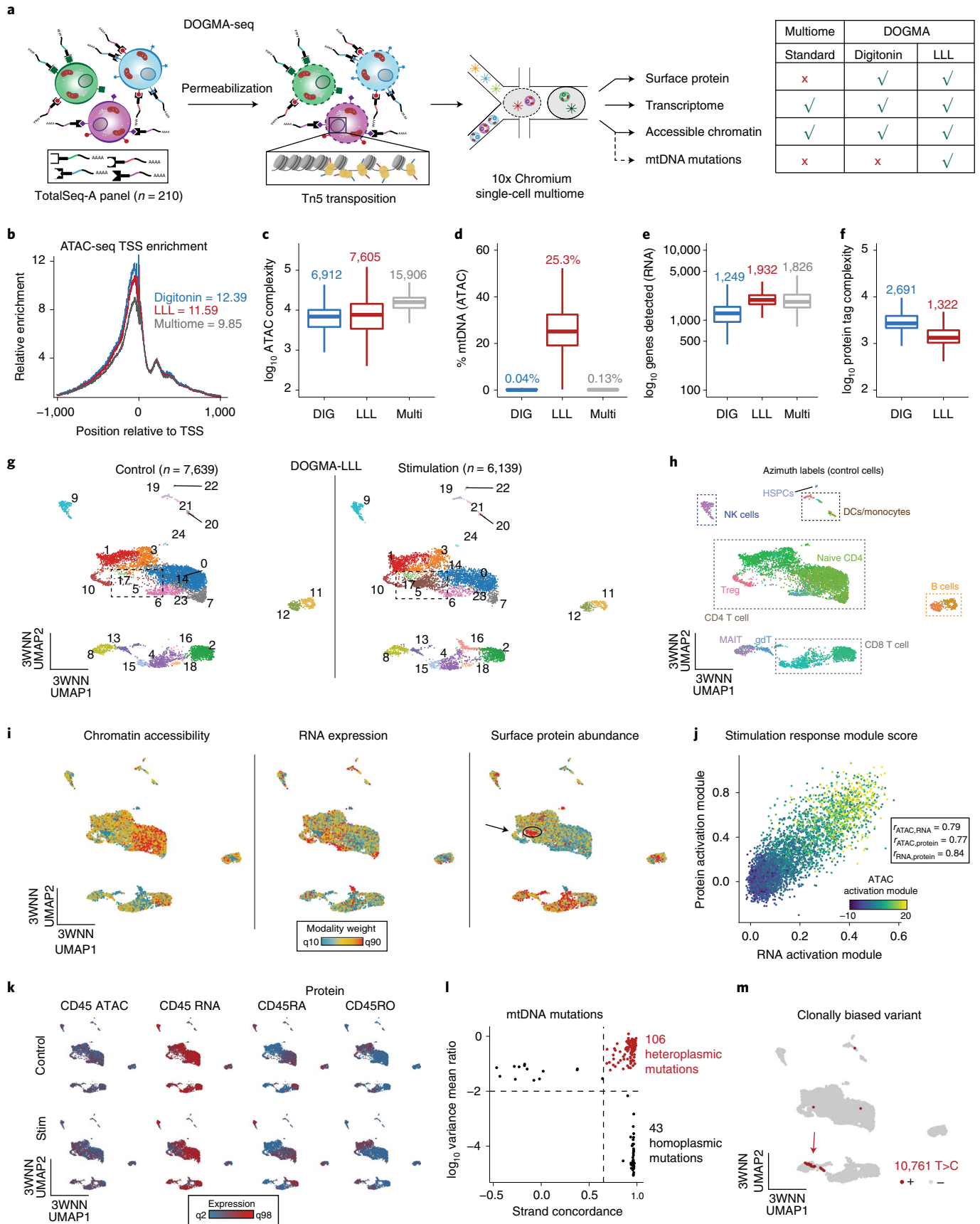
Next, we examined the dynamic changes in accessible chromatin, gene expression and protein abundance in stimulated versus control T cells. We detected 8,326 differential peaks, 943 differentially expressed genes and 71 differentially abundant surface proteins, consistent with previous unimodal analyses largely from bulk experiments[38,39] (Fig. 4e and Methods). Of the 84 cases where all three modalities were detected in T cells, we observed heterogeneous responses in gene expression, chromatin accessibility and surface protein abundance (Fig. 4f,g and Supplementary Table 4). Specifically, CD3 and CD28 downregulation along with CD69 upregulation are striking on the protein level, evident transcriptomically only for *CD3E* and *CD69*, but barely detectable at the chromatin level (Fig. 4h,i and Extended Data Fig. 4h). This can be due to true invariance in chromatin accessibility, such that gene expression is temporarily repressed without loss of accessibility, or to technical challenges—for example, given the higher sensitivity in capturing a modality with higher copy number (protein), as exemplified by CD4 and CD279 (Fig. 4j and Extended Data Fig. 4i). On the other hand, we observed RNA-specific changes in CD52 where chromatin accessibility and protein abundance were relatively constant pre- and post-stimulation (Extended Data Fig. 4j). Furthermore, as we activated T cells in a PBMC culture, we could observe secondary effects of the stimulations in other cell types, such as B cells, where CD25 and CD184 were altered (Extended Data Fig. 4k–m). Together, these analyses and anecdotes highlight the utility of combining ASAP-seq and CITE-seq to distinguish changes at three levels of gene regulation at single-cell resolution.

DOGMA-seq enables co-measurement of four modalities. While this work was under review, 10x Genomics released the 'Multiome' product that captures the transcriptome and chromatin accessibility from the same cells. We recognized that the mechanism of protein barcode detection used in our previously described CITE-seq method[5], via the barcoded poly-T primer, would be compatible with the Multiome product and that our efforts to preserve cell surface antigens and mtDNA described for ASAP-seq would be transferable

Fig. 5 | DOGMA-seq enables a high-quality capture of multiple modalities sensitive to biological changes. a, Schematic of the workflow and the modality capture enabled by DOGMA-seq. **b**, TSS enrichment scores of DOGMA-seq variations on the control PBMC data compared to a Multiome PBMC dataset released by 10x Genomics. **c–f**, Additional quality control metric comparisons for the indicated conditions: ATAC fragment complexity (**c**), % mtDNA in ATAC library (**d**), number of genes per cell detected (**e**) and protein tag complexity across the different cell preparations (**f**); median values are indicated. Box plots: center line, median; box limits, first and third quartiles; whiskers, 1.5× interquartile range. **g**, 3WNN UMAP embedding of control and stimulated PBMCs under the LLL condition. Box indicates the activated T cell clusters. **h**, Control PBMC clusters labeled after projection into the azimuth reference. **i**, 3WNN UMAPs of stimulated PBMCs highlighting the weight of each modality. **j**, Stimulation module family transcription factors, includingscore for each of the three modalities quantified in stimulated T cells. Each dot is a single cell with the stimulation score for each modality. Per-pair Pearson correlation of the data shown is reported. **k**, 3WNN UMAPs of control and stimulated PBMCs highlighting chromatin accessibility, mRNA expression and surface protein levels for CD45 and isoforms CD45RA and CD45RO. **l**, Identification of high-confidence heteroplasmic (red; *n* = 106) and homoplasmic (black; *n* = 43) variants using mgatk. **m**, Cellular distribution of m.10,761 T>C in the 3WNN for all cells; the arrow points to a subset of γ/δ and MAIT cells. Threshold for + was 10% heteroplasmy based on empirical density. DC, dendritic cell; HSPC, hematopoietic stem and progenitor cell.

to this kit. The ability to capture readouts of gene activity from chromatin accessibility to mRNA expression to protein levels spans the central dogma of gene regulation, leading us to refer to this method as DOGMA-seq (Methods). To benchmark DOGMA-seq, we repeated the PBMC stimulation experiment, as described in Fig. 4, under two permeabilization conditions. The first is based on

Fig. (continued caption / panels a–m)

the mtDNA-retaining LLL preparation as performed in ASAP-seq with reduced amount of fixative; the second is based on a recently pre-printed condition using digitonin (DIG) permeabilization to allow detection of surface proteins alongside chromatin accessibility[40] (Fig. 5a). Using either preparation, we observed similar quality control metrics for TSS enrichment and genes detected compared to the standalone Multiome kit (Fig. 5b,e and Extended Data Fig. 5a,d), with the co-detection of mtDNA reads and protein tags being uniquely enabled by the LLL conditions (Fig. 5d,f and Extended Data Fig. 5c,e). Although LLL and DIG treatment led to similarly reduced accessible chromatin library complexity (Fig. 5c and Extended Data Fig. 5b), the overall complexity was similar to mtscATAC-seq and ASAP-seq (Extended Data Fig. 1f). We emphasize that the DIG and LLL treatments display differences with respect to mtDNA yield in ATAC libraries and complexity in protein tag libraries, which we attribute to the milder permeabilization properties of DIG, not lysing mitochondria[41], and improving preservation of the surface membrane and associated proteins (Fig. 5d,f and Extended Data Fig. 5c,e). As mitochondria in DIG-treated cells are ultimately lysed after droplet generation, we also observe a higher fraction of mtRNA, which are otherwise lost after LLL treatment and washes (Extended Data Fig. 5f). This is supported by the mitochondria-dependent increase in fraction of UMIs mapping to exons when using the DIG preparation (Extended Data Fig. 5g,h).

Next, we sought to chart biological variation in single cells. To achieve this, we extended the weighted nearest neighbor (WNN) approach[27] (Methods). Focusing on the LLL-DOGMA data, we performed three-modality WNN (3WNN), yielding 25 clusters, including activated T cell clusters that were corroborated by healthy PBMC reference data projection (Fig. 5g,h). To chart the contributions of each modality, we visualized the relative modality contribution weight from the 3WNN clustering (Fig. 5i), observing a strong effect from the chromatin accessibility component in a subpopulation of naive T cells. We further observed a protein marker-driven T cell cluster (cluster 17) that was delineated by CD138 (Extended Data Fig. 5i), which might mark an underappreciated, yet functional, population[42]. Further analyses indicated that DOGMA-seq showed sensitivity in detecting the molecular changes in all three modalities during the stimulation, at specific loci (Extended Data Fig. 5j) and globally via activation module scores (Fig. 5j and Methods). We further verified the distinct isoforms of CD45 that can be delimited only via our additional surface protein readout (Fig. 5k). Additionally, we compared three variations of WNN by holding out each of the accessible chromatin, transcriptome and surface protein measurements (Extended Data Fig. 5k), noting that the CD138+ T cell cluster (cluster 17) was not recovered in the absence of protein information.

As a fourth modality, we verified an approximately uniform coverage across the mitochondrial genome from this assay and a slight elevation of mean coverage in the stimulated cells (Extended Data Fig. 5l), consistent with increased mitochondrial biogenesis during

T cell activation[43]. Application of mgatk[14] revealed 106 heteroplasmic mutations (Fig. 5l), including an enrichment of specific nucleotide substitutions (Extended Data Fig. 5m) as observed in our bone marrow experiment (Fig. 3). Of note, we observed 13 variants with evidence of clonal lineage bias (Kruskal–Wallis-adjusted $P < 0.01$), including m.10761 T>C that was present in a putative subclone of largely gamma/delta (γ/δ) and mucosal-associated invariant T (MAIT) cells (Fig. 5m). Overall, mutational heteroplasmy did not significantly change between conditions (Kolmogorov–Smirnov test, $P = 0.83$), consistent with the short culture time limiting clonal expansion (Extended Data Fig. 5n). Finally, application of 3WNN to DIG-DOGMA and LLL-DOGMA together verified that the two approaches could be analyzed concomitantly with a clear preservation of biological signal (Extended Data Fig. 5o), and peak-to-gene linkages were largely concordant at loci with both surface protein and transcriptome measured (Extended Data Fig. 5p and Methods). Overall, our results show how extending the number of distinct measurements from the Multiome kit facilitates the interrogation of biological systems with greater nuance.

Multiplexed CRISPR perturbations in primary T cells. As the benchmarking of our approaches revealed distinct chromatin and protein changes underlying T cell activation, we sought to refine some of the underlying mechanisms by targeted perturbations via an arrayed CRISPR–Cas9 screening strategy. To this end, naive human CD4+ T cells were stimulated with anti-CD3/CD28, rested and subjected to CRISPR perturbations targeting single genes (*CD3E*, *CD4*, *ZAP70* and *NFKB2*) or *CD3E* + *CD4* (double knockout), followed by rest and re-stimulation (Extended Data Fig. 6a). Using a combination of hashing antibodies, we multiplexed each perturbation before antibody panel staining ($n = 37$) and processing by ASAP-seq (Fig. 6a and Supplementary Tables 1 and 5). Demultiplexing by hashtag reads enabled high-confidence identification of 5,825 perturbed cells with a median yield of 1.47×10^4 fragments mapping to the nuclear genome (Extended Data Fig. 6b).

Cells perturbed by guide RNAs (gRNAs) targeting critical regulators of TCR signal transduction (*CD3E* and *ZAP70*) had similar chromatin accessibility profiles (Fig. 6b). Moreover, these cells expressed characteristic markers of a resting state (for example, CD197 and CD62L), indicating a profound defect in TCR activation (Fig. 6c and Extended Data Fig. 6d). In contrast, cells with non-targeting (NTC) gRNAs or gRNAs targeting *CD4* or *NFKB2* clustered together and displayed high levels of classical T cell activation markers (for example, CD25, CD69 and CD137), indicating active TCR signaling. Notably, only cells with gRNAs targeting *CD4* exhibited substantial reduction in CD4 expression, further validating the robustness of the workflow (Fig. 6c and Extended Data Fig. 6c).
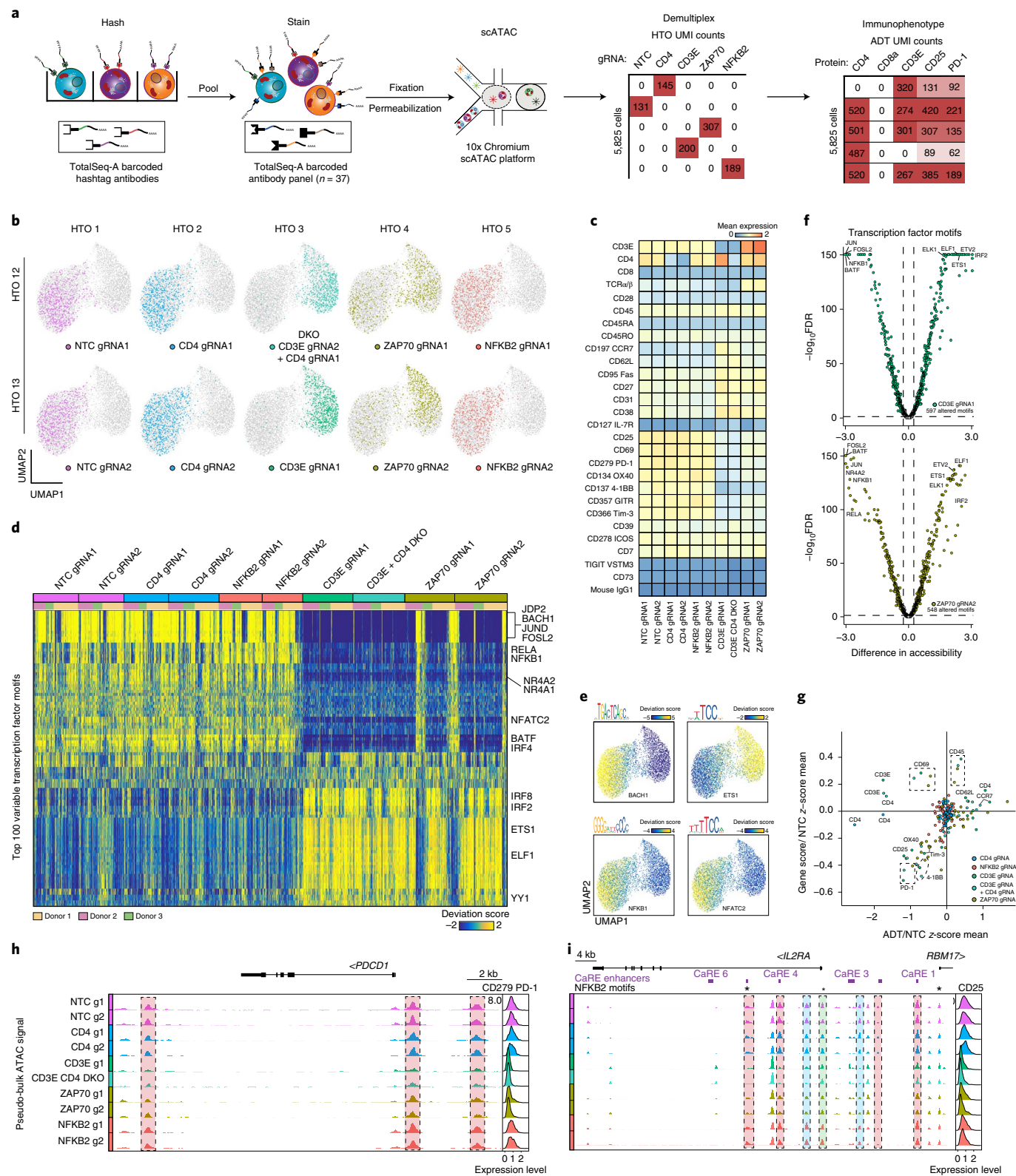
We next inferred changes in gRNA-dependent TF activities using chromVAR[44] and found that each perturbation had its predicted effect, despite varying targeting efficiencies (Fig. 6d–f and Extended Data Fig. 6e–g). As expected, depletion of *CD3E* resulted in a defective response to TCR re-stimulation and significantly

Fig. 6 | Multiplexed CRISPR perturbations with ASAP-seq in primary human T cells. a, Schematic workflow for combinatorial multiplexing with ASAP-seq. CRISPR-edited cells are first stained with oligo-conjugated hashtag antibodies and then pooled for processing by ASAP-seq. gRNA identities are demultiplexed using hashing antibody counts. **b**, UMAP embedding of $n = 5,825$ single cells and their associated gRNAs. **c**, Heat map showing mean expression for 27 surface protein markers across gRNA perturbations in stimulated cells. **d**, Heat map representation of chromVAR bias-corrected TF motif deviation scores for the top 100 most variable TFs across perturbation conditions. Associated gRNA and donor information are color-coded and indicated at the top of the plot. **e**, Overlay on ASAP-seq UMAP of chromVAR TF motif deviations. The motif for the given TF is indicated at the top of the plot. **f**, Volcano plots showing TF motifs with significantly changed chromatin accessibility profiles between NTC cells and guides targeting *CD3E* and *ZAP70* (FDR ≤ 0.05, chromVAR accessibility change ≥ 0.25). **g**, Scatter plot of mean gene activity scores for 22 individual gene loci plotted against CLR-normalized mean protein tag counts associated with each gRNA. Values are normalized against NTC cells. **h, i**, Genomic tracks of *PDCD1* (encoding PD-1) (**h**) and *IL2RA* (encoding CD25) (**i**), indicating pseudo-bulk ATAC signal tracks across gRNAs with corresponding CLR-normalized protein abundance ridge plots. Differentially accessible regions are highlighted in red. Differentially accessible regions not overlapping CaRE enhancers are highlighted in blue (**i**), and the TSS is highlighted in green (**i**). NFKB2 sequence motif matches are indicated by *. DKO, double knockout; HTO, hashtag oligo.

decreased accessibility in regions containing motifs of activator protein-1 (AP-1) family transcription factors, including c-JUN and BATF (median chromVAR accessibility loss, 10.24, and false discovery rate (FDR) < 0.0001; median chromVAR accessibility loss, 6.96, and FDR < 0.0001, respectively) compared to NTCs. Additional altered TF motifs in *CD3E*-targeted cells included NFAT family TFs, consistent with their crucial roles in chromatin remodeling and transcriptional regulation after TCR activation[45,46]. Interestingly,

disruption of *NFKB2* led to an increase of accessibility for NFKB family motifs, which could be reflective of competitive dimerization of p50 and p52 for common binding partners RELA and RELB (Fig. 6d and Extended Data Fig. 6f)[47].

Although methodology pairing measurements of perturbations and ATAC-seq profiles has enabled the dissection of molecular machinery governing cell state[48], our approach uniquely allows queries of how chromatin changes might relate to protein

a

b

c

d

e
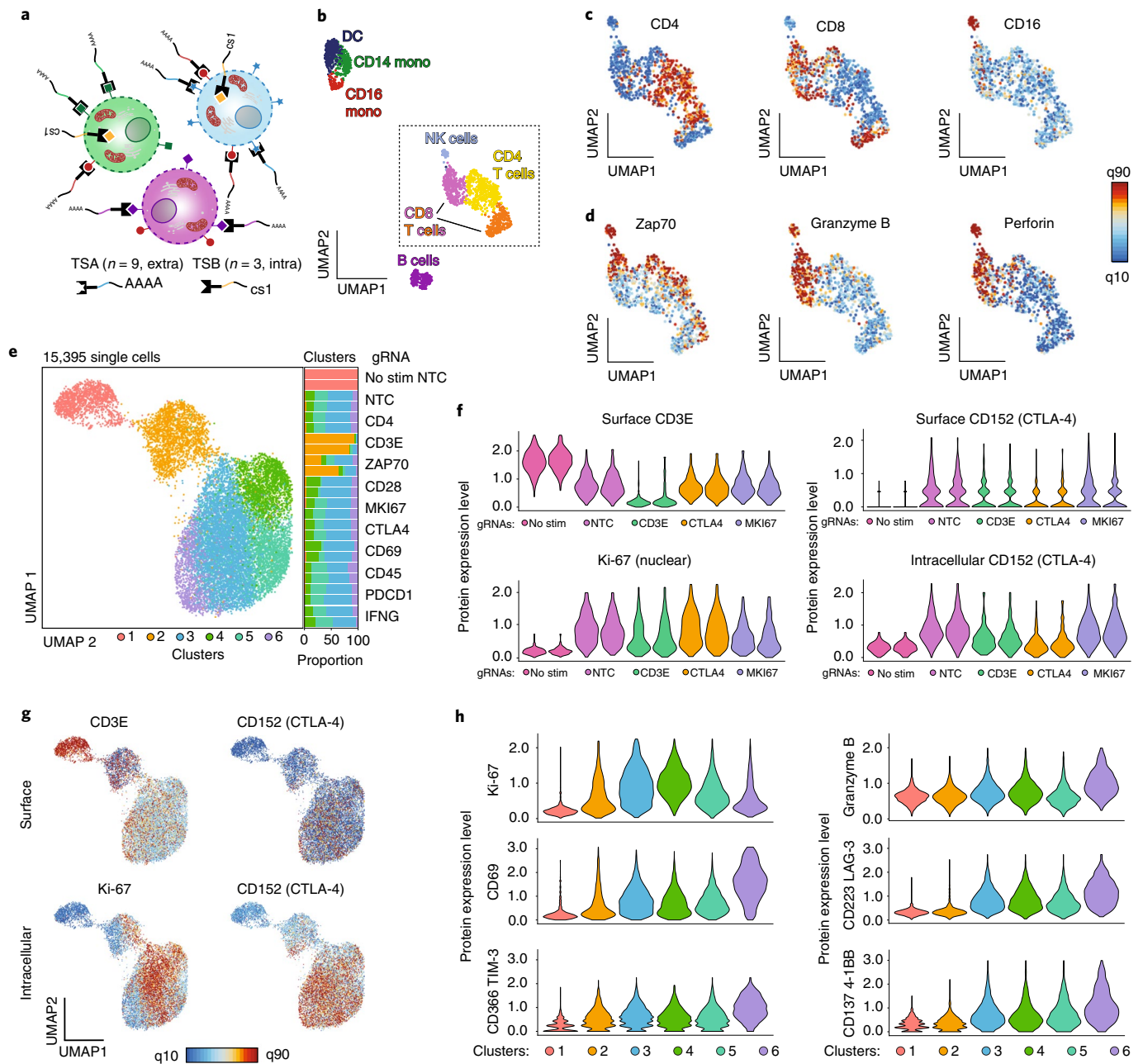
f

Transcription factor motifs

g

h

i

Fig. 7 | ASAP-seq enables detection of intracellular proteins with barcoded antibodies. a, Schematic of the intracellular staining experimental design. PBMCs stained with the TSA TBNK panel directed against cell surface markers were fixed, lysed and stained with TSB antibodies directed against intracellular markers, followed by transposition. **b**, Two-dimensional embedding of the PBMC chromatin accessibility data using UMAP, with major peripheral blood cell types highlighted. **c**, **d**, T cells and NK cells as highlighted in the dashed-line box from **b** with superimposed tag intensities for indicated cell surface (**c**) and intracellular (**d**) markers. Color bar: protein tag CLR values. **e**, UMAP embedding of $n = 15,395$ single cells targeted with distinct gRNAs. Colors indicate cluster identity (labeled on the bottom). The proportion of gRNA-targeted cells in each cluster is shown on the right. **f**, Violin plots showing the distribution of CLR-normalized protein counts for indicated proteins and their associated gRNA. **g**, UMAP embedding overlaid with expression of two surface (top) and two intracellular (bottom) protein markers. Color bar: protein tag CLR values. **h**, Violin plot showing CLR-normalized protein counts for indicated proteins across the six Louvain clusters. DC, dendritic cell.

expression dynamics. To examine this, we compared accessible chromatin scores with concomitant surface protein profiling across each perturbation condition (Supplementary Table 5). Overall, perturbation-induced changes in surface protein expression were correlated with changes in chromatin status ($r = 0.57$ in proteins not targeted by a perturbation; Fig. 6g,h). For example, many stimulation-responsive genes (for example, CD25, CD134 and CD279) were downregulated in both protein expression and

chromatin accessibility in *CD3E*- and *ZAP70*-targeted cells (compared to NTCs), largely mirroring our previous results (Figs. 4 and 5). Interestingly, we observed pronounced coordination between changes in protein expression and gene activity for CD357 and CD366 (centered log-ratio (CLR)-normalized mean protein tag difference of 0.84 and 0.66, respectively, between *CD3E*-targeted cells and NTCs; Extended Data Fig. 6h). This was not evident in our ASAP-seq PBMC stimulation experiment, where changes

in CD357 and CD366 protein levels were only modest, despite increased accessibility at associated stimulation-responsive enhancers (CLR-normalized mean protein tag difference of 0.18 and 0.17, respectively, between CD4 T cell stimulation and control), likely due to the shorter stimulation period.

Finally, as fine mapping of *cis*-regulatory elements has enhanced the capacity to uncover functional regulatory elements in different contexts[49–52], we reasoned that our perturbation screening approach could offer similar biological insights in identifying stimulation-responsive accessible chromatin regions. Examining pseudo-bulk ATAC signal at the *IL2RA* locus, we found strong depletion of chromatin accessibility at select regions with a concomitant decrease in the expression of CD25 protein for cells targeted by gRNAs against *CD3E* and *ZAP70*, suggesting a prerequisite of TCR stimulation in the activation of these putative enhancers (Fig. 6i). These affected enhancers largely overlapped the previously characterized *IL2RA* CRISPRa-responsive elements (CaREs)[51]. In particular, we observed marked accessibility changes overlapping CaRE4, a validated TCR stimulation-responsive enhancer for *IL2RA*. Moreover, we observed a decrease in CD25 expression in cells perturbed by gRNAs targeting *NFKB2*, despite relatively unchanged chromatin accessibility and the presence of compatible NFKB2 DNA-binding motifs within regulatory regions. These results suggest that, although NFKB2 does not actively regulate local chromatin accessibility at this locus, it might still play a role in coordinating and maintaining CD25 expression in activated T cells. Taken together, our integrated multimodal approach allows for the discovery of context-dependent coding and non-coding gene regulation modules.

ASAP-seq enables the detection of intracellular proteins. We hypothesized that ASAP-seq could provide an opportunity to detect intracellular epitopes, which has been generally inaccessible via droplet-based scRNA-seq[5–7]. To examine this, we stained PBMCs with different conjugate families against extracellular and intracellular markers to allow independent amplification and tunable sequencing depth of the two tag libraries. We labeled cells with the TSA TBNK panel comprising extracellular surface markers, followed by fixation, permeabilization and staining with TSB antibodies directed against intracellular epitopes, ZAP70, perforin and granzyme B (Fig. 7a and Supplementary Table 1). Accessible chromatin profile-based clustering and distribution of protein tags for extracellular markers (Fig. 7b,c) was consistent with previous experiments and corresponding gene activity scores (Extended Data Figs. 1i,j and 7a,b), verifying that the detection of these modalities remained robust with intracellular staining.

The distribution of protein tags for intracellular proteins displayed expected patterns, with ZAP70 present in activated NK and T cells and perforin and granzyme B most prominent in NK cells and a subset of cytotoxic CD8[+] T cells, and correlated with gene activity scores, ultimately validating the on-target activity for all three tested intracellular markers (Fig. 7c,d and Extended Data Fig. 7a,b).

To further assess the specificity of intracellular staining, we generated a series of targeted perturbations, including *CD4*, *CD3E*, *ZAP70*, *CD28*, *MKI67*, *CTLA4*, *CD69*, *CD45*, *PDCD1* and *IFNG* (two gRNAs per gene; Extended Data Fig. 7c,d), encoded each by hashing, before pooling of cells and staining with a TSA antibody surface marker panel (*n* = 53). After fixation and permeabilization, we stained with TSB antibodies including intracellular CD152 (*CTLA4*) and nuclear Ki-67 (*MKI67*), both known to be upregulated upon TCR activation (Supplementary Table 1)[53,54]. After hashtag demultiplexing, we identified six scATAC-seq Louvain clusters and visualized 15,395 single-cell profiles (Fig. 7e).

Consistent with its role as a proliferation marker, Ki-67 protein expression was highly upregulated in stimulated cells (Fig. 7f)[54].

This was not observed in cells with perturbed TCR signaling and non-stimulated control cells. Separate measurements of CD152 pre-permeabilization (staining CD152 on the cell's surface) and post-permeabilization (co-staining internalized CD152) revealed substantial upregulation of intracellular CD152 in TCR signaling-competent cells (Fig. 7f). For cell surface-localized CD152, we observed a sparse signal that lacked the dynamic range observed in its intracellular staining counterpart, consistent with previous reports on CD152 trafficking and localization (Fig. 7f,g)[54,55]. Notably, *CTLA4*- and *MKI67*-perturbed cells exhibited marked reduction in protein expression of CD152 and Ki-67, respectively, highlighting the specificity of intracellular epitope detection in both the cytoplasm and nucleus.

Among TCR-activated cells (Louvain clusters 3–6), we observed variation in Ki-67 levels, most notably with low expression in clusters 5 and 6 (Fig. 7e,g). Cells in cluster 6 exhibited high levels of T cell activation markers (for example, CD69, granzyme B and CD223), consistent with an effector memory phenotype (Fig. 7h)[56,57]. We did not observe overrepresentation of *MKI67*-targeted cells in this population, suggesting that the effect is biological and not due to perturbation. At the chromatin level, we found increased accessibility at *IFNG* and *GZMB* loci in cluster 6, consistent with an activated effector phenotype (Extended Data Fig. 7e). Together, these data demonstrate how ASAP-seq enables intracellular protein detection to facilitate the nuanced annotation of biologically relevant immune phenotypes and cell states.

Discussion

Here we present ASAP-seq, a unique approach that enables the concomitant detection of protein abundance alongside transposase-accessible chromatin and mtDNA in thousands of single cells[14,15]. As most cell atlases to date have characterized the distinct transcriptomes of single cells in complex tissue, ASAP-seq provides a complementary multi-omic approach to map regulatory elements, protein abundances and clonal relationships. The ASAP-seq workflow is directly compatible with related multimodal assays that simultaneously measure protein and RNA[5,6] (Fig. 4). Notably, our approach introduces a bridge oligo (Fig. 1) that enables the use of existing antibody conjugates, yielding an accessible and user-friendly protocol.

As a complement to our ASAP-seq assay, we describe an extension of CITE-seq to enable compatibility with the 10x Genomics Multiome product. The resulting trimodal assay, DOGMA-seq, enables measurement of chromatin accessibility, gene expression and protein levels in single cells, with optional detection of mtDNA genotypes. We expect future developments to mirror efforts and expand the functionality of ASAP-seq with respect to intracellular protein detection, localized transposition (for example, CUT&Tag[58–60]) and detection of engineered perturbations[61]. We note that, while this work was under revision, a similar strategy sharing several of the features of DOGMA-seq, referred to as TEA-seq, was posted as a preprint[62].

By examining our multimodal readouts in native and perturbed hematopoietic cells, our analyses reveal distinct cellular programming occurring in chromatin, transcriptional and post-translational regulation. In particular, we observe chromatin-based priming during a monocyte developmental trajectory in bone marrow (Fig. 3). Conversely, during T cell activation, we observe a more heterogeneous response where changes in chromatin, RNA and protein abundances become more uncoupled (Figs. 4 and 5). By further using CRISPR-based perturbations (Fig. 6), we disentangle downstream signaling of the TCR and reveal how ASAP-seq can enable the fine mapping of regulatory elements in various cell states that directly affect protein expression (for example, *IL2RA*). Future extensions of ASAP-seq that incorporate direct detection of gRNA sequences[6,61,63]

or encoded guide barcodes[64–66] will further enable pooled screens at a substantially increased scale and phenotypic depth.

Furthermore, we show that ASAP-seq allows direct detection and quantification of cytoplasmic and nuclear intracellular markers (Fig. 7). Although other protocols have achieved concomitant quantification of intracellular protein abundance and gene expression with plate-based methods[67], using isolated nuclei[68] or combining fluorescence-activated cell sorting (FACS)-based enrichment of cells with scRNA-seq[69], ASAP-seq provides a more parsimonious approach to concomitant estimation of both surface and intracellular markers of whole cells alongside chromatin accessibility profiles on a widely used commercial platform. We anticipate the intracellular protein detection capacity of ASAP-seq to spur the development of large antibody panels targeting intracellular epitopes including signaling molecules, specific phospho-epitopes and TFs. In total, our methodological approach and analyses demonstrate the utility of ASAP-seq and DOGMA-seq as modular and powerful tools for charting the complex interplay of gene and protein regulatory layers in single cells.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-021-00927-2.

References

1. Nam, A. S., Chaligne, R. & Landau, D. A. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nat. Rev. Genet.* **22**, 3–18 (2021).
2. Zhu, C., Preissl, S. & Ren, B. Single-cell multimodal omics: the power of many. *Nat. Methods* **17**, 11–14 (2020).
3. Schier, A. F. Single-cell biology: beyond the sum of its parts. *Nat. Methods* **17**, 17–20 (2020).
4. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat. Rev. Genet.* **20**, 257–272 (2019).
5. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
6. Mimitou, E. P. et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 (2019).
7. Peterson, V. M. et al. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936–939 (2017).
8. Ma, S. et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116 (2020).
9. Lareau, C. A. et al. Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat. Biotechnol.* **37**, 916–924 (2019).
10. Satpathy, A. T. et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).
11. Cao, J. et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018).
12. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* **37**, 1452–1457 (2019).
13. Zhu, C. et al. An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nat. Struct. Mol. Biol.* **26**, 1063–1070 (2019).
14. Lareau, C. A. et al. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat. Biotechnol.* **39**, 451–461 (2021).
15. Walker, M. A. et al. Purifying selection against pathogenic mitochondrial DNA in human T cells. *N. Engl. J. Med.* **383**, 1556–1563 (2020).
16. Granja, J. M. et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.* **37**, 1458–1465 (2019).
17. Stoeckius, M. et al. Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).
18. Gaublomme, J. T. et al. Nuclei multiplexing with barcoded antibodies for single-nucleus genomics. *Nat. Commun.* **10**, 2907 (2019).
19. McGinnis, C. S. et al. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).
20. Gehring, J., Hwee Park, J., Chen, S., Thomson, M. & Pachter, L. Highly multiplexed single-cell RNA-seq by DNA oligonucleotide tagging of cellular proteins. *Nat. Biotechnol.* **38**, 35–38 (2020).
21. Srivatsan, S. R. et al. Massively multiplex chemical transcriptomics at single-cell resolution. *Science* **367**, 45–51 (2020).
22. Shin, D., Lee, W., Lee, J. H. & Bang, D. Multiplexed single-cell RNA-seq via transient barcoding for simultaneous expression profiling of various drug perturbations. *Sci. Adv.* **5**, eaav2249 (2019).
23. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
24. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
25. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
26. Kim, H. J., Lin, Y., Geddes, T. A., Yang, J. Y. H. & Yang, P. CiteFuse enables multi-modal analysis of CITE-seq data. *Bioinformatics* **36**, 4137–4143 (2020).
27. Hao, Y., Hao, S., Andersen-Nissen, E. & Mauck, W. M. Integrated analysis of multimodal single-cell data. Preprint at *bioRxiv* https://doi.org/10.1101/2020.10.12.335331 (2020).
28. Ludwig, L. S. et al. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325–1339 (2019).
29. Lareau, C. A., Ludwig, L. S. & Sankaran, V. G. Longitudinal assessment of clonal mosaicism in human hematopoiesis via mitochondrial mutation tracking. *Blood Adv.* **3**, 4161–4165 (2019).
30. Ju, Y. S. et al. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature* **543**, 714–718 (2017).
31. Buenrostro, J. D. et al. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. *Cell* **173**, 1535–1548 (2018).
32. Velten, L. et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).
33. Hartmann, F. J. & Bendall, S. C. Immune monitoring using mass cytometry and related high-dimensional imaging approaches. *Nat. Rev. Rheumatol.* **16**, 87–99 (2019).
34. Sancho, D., Gómez, M. & Sánchez-Madrid, F. CD69 is an immunoregulatory molecule induced following activation. *Trends Immunol.* **26**, 136–140 (2005).
35. Caruso, A. et al. Flow cytometric analysis of activation markers on stimulated T cells and their correlation with cell proliferation. *Cytometry* **27**, 71–76 (1997).
36. Dong, C. et al. ICOS co-stimulatory receptor is essential for T-cell activation and function. *Nature* **409**, 97–101 (2001).
37. San José, E., Borroto, A., Niedergang, F., Alcover, A. & Alarcón, B. Triggering the TCR complex causes the downregulation of nonengaged receptors by a signal transduction-dependent mechanism. *Immunity* **12**, 161–170 (2000).
38. Calderon, D. et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat. Genet.* **51**, 1494–1505 (2019).
39. Gate, R. E. et al. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. *Nat. Genet.* **50**, 1140–1150 (2018).
40. Swanson, E., Lord, C., Reading, J. & Heubeck, A. T. Integrated single cell analysis of chromatin accessibility and cell surface markers. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.04.283887 (2020).
41. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
42. Liu, L., Takeda, K. & Akkoyunlu, M. Disease stage-specific pathogenicity of CD138 (syndecan 1)-expressing T cells in systemic lupus erythematosus. *Front. Immunol.* **11**, 1569 (2020).
43. Tan, H. et al. Integrative proteomics and phosphoproteomics profiling reveals dynamic signaling networks and bioenergetics pathways underlying T cell activation. *Immunity* **46**, 488–503 (2017).
44. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
45. Hogan, P. G. Transcriptional regulation by calcium, calcineurin, and NFAT. *Genes Dev.* **17**, 2205–2232 (2003).
46. Macian, F. NFAT proteins: key regulators of T-cell development and function. *Nat. Rev. Immunol.* **5**, 472–484 (2005).
47. Basak, S., Shih, V. F.-S. & Hoffmann, A. Generation and activation of multiple dimeric transcription factors within the NF-κB signaling system. *Mol. Cell. Biol.* **28**, 3139–3150 (2008).
48. Rubin, A. J. et al. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell* **176**, 361–376 (2019).
49. Korkmaz, G. et al. Functional genetic screens for enhancer elements in the human genome using CRISPR–Cas9. *Nat. Biotechnol.* **34**, 192–198 (2016).

50. Gasperini, M. et al. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* **176**, 1516 (2019).
51. Simeonov, D. R. et al. Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549**, 111–115 (2017).
52. Fulco, C. P. et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
53. Perkins, D. et al. Regulation of CTLA-4 expression during T cell activation. *J. Immunol.* **156**, 4154–4159 (1996).
54. Scholzen, T. & Gerdes, J. The Ki-67 protein: from the known and the unknown. *J. Cell. Physiol.* **182**, 311–322 (2000).
55. Valk, E., Rudd, C. E. & Schneider, H. CTLA-4 trafficking and surface expression. *Trends Immunol.* **29**, 272–279 (2008).
56. Kaech, S. M. et al. Selective expression of the interleukin 7 receptor identifies effector CD8 T cells that give rise to long-lived memory cells. *Nat. Immunol.* **4**, 1191–1198 (2003).
57. Soares, A. P. et al. Longitudinal changes in CD4+ T-cell memory responses induced by BCG vaccination of newborns. *J. Infect. Dis.* **207**, 1084–1094 (2013).
58. Kaya-Okur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
59. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell profiling of histone modifications in the mouse brain. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.02.279703 (2020).
60. Wu, S. J. et al. Single-cell analysis of chromatin silencing programs in developmental and tumor progression. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.04.282418 (2020).
61. Pierce, S. E., Granja, J. M. & Greenleaf, W. J. High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.02.364265 (2020).
62. Swanson, E. et al. TEA-seq: a trimodal assay for integrated single cell measurement of transcription, epitopes, and chromatin accessibility. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.04.283887 (2020).
63. Replogle, J. M. et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* **38**, 954–961 (2020).
64. Wroblewska, A. et al. Protein barcodes enable high-dimensional single-cell CRISPR screens. *Cell* **175**, 1141–1155 (2018).
65. Dixit, A. et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866 (2016).
66. Jaitin, D. A. et al. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* **167**, 1883–1896 (2016).
67. Gerlach, J. P. et al. Combined quantification of intracellular (phospho-) proteins and transcriptomics from fixed single cells. *Sci. Rep.* **9**, 1469 (2019).
68. Chung, H. et al. Simultaneous single cell measurements of intranuclear proteins and gene expression. Preprint at *bioRxiv* https://doi.org/10.1101/2021.01.18.427139 (2021).
69. Katzenelenbogen, Y. et al. Coupled scRNA-seq and intracellular protein activity reveal an immunosuppressive role of TREM2 in cancer. *Cell* **182**, 872–885 (2020).

Methods

Protocols. See Protocol Exchange (doi pending), https://cite-seq.com/protocols/ or Supplementary Information for step-by-step ASAP-seq and DOGMA-seq protocols.

Cells. Cryopreserved healthy donor PBMCs and bone marrow cells were obtained from AllCells or Cellular Technology Limited and processed immediately after thawing. NIH-3T3 and HEK-293FT cells were maintained according to standard procedures in DMEM (Thermo Fisher Scientific) supplemented with 10% FBS (Thermo Fisher Scientific) at 37 °C with 5% CO_2.

Cell staining with barcoded antibodies. TSA- and TSB-conjugated antibodies and panels were obtained from BioLegend (see Supplementary Table 1 for a list of antibodies, clones and barcodes used for ASAP-seq). Cells were stained with barcoded (and fluorophore-conjugated where indicated) antibodies as previously described for CITE-seq[5,6]. Briefly, approximately 1.5–2 million cells per sample were resuspended in 1× CITE-seq staining buffer (2% BSA and 0.01% Tween in PBS) and incubated for 10 min with Fc receptor block (TruStain FcX, BioLegend) to block Fc receptor-mediated binding. Subsequently, cells were incubated with indicated antibodies or panels for 30 min at 4 °C, as recommended by the manufacturer (BioLegend). After staining, cells were washed 3× by resuspension in 1× CITE-seq staining buffer followed by centrifugation (300*g* for 5 min at 4 °C) and supernatant exchange. After the final wash, cells were resuspended in PBS and subjected to fixation and permeabilization as described in the 'Cell fixation and permeabilization' section.

Intracellular staining was performed in fixed and permeabilized cells that were resuspended in Intracellular Staining Buffer (BioLegend, custom part no. 900002577), with the addition of TruStain FcX and True-Stain Monocyte Blocker as recommended by the manufacturer (BioLegend).

Cell fixation and permeabilization for ASAP-seq. Cells were fixed in 1% formaldehyde (Thermo Fisher Scientific, no. 28906) in PBS for 10 min at room temperature and quenched with glycine solution to a final concentration of 0.125 M, followed by washing twice in PBS via centrifugation at 400*g* for 5 min at 4 °C. Cells were subsequently treated with the appropriate lysis buffer depending on downstream application. If mtDNA retention was desired, permeabilization was performed as described in mtscATAC-seq[14] with 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl$_2$, 0.1% NP40 and 1% BSA (referred to as LLL conditions). When mtDNA depletion was desired, cells were lysed in 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl$_2$, 0.1% NP40, 0.1% Tween 20, 0.01% DIG and 1% BSA (referred to as OMNI conditions). Permeabilization was performed on ice—3 min for primary cells and 5 min for cell lines—followed by adding 1 ml of chilled wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl$_2$ and 1% BSA) and inversion before centrifugation at 500*g* for 5 min at 4 °C. The supernatant was discarded, and cells were diluted in 1× diluted nuclei buffer (10x Genomics) and filtered through a 40-µm Flowmi cell strainer before counting using trypan blue and a Countess II FL Automated Cell Counter.

Transposition and barcoding for ASAP-seq. Cell were subsequently processed according to the Chromium Single Cell ATAC Solution user guide (versions CG000168 Rev D for v1 and CG000209 Rev D for v1.1, 10x Genomics) with the following modifications:

1. During the barcoding reaction (Step 2.1), 0.5 µl of 1 µM bridge oligo was added to the barcoding mix. The sequences of the bridge oligos are: BOA (bridge oligo for TSA): TCGTCGGCAGCGTCAGATGTGTATAAGAGACA-GNNNNNNNNNNVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT/3InvdT/ and BOB (bridge oligo for TSB): TCGTCGGCAGCGTCAGATGTGTATAAG AGACAGTTGCTAGGACCGGCCTTAAAGC/3InvdT/.

2. To facilitate bridge oligo annealing during GEM incubation (Step 2.5), a 5-min incubation at 40 °C was added at the beginning of the amplification protocol (40 °C for 5 min, 72 °C for 5 min, 98 °C for 30 s; 12 cycles of 98 °C for 10 s, 59 °C for 30 s and 72 °C for 1 min; ending with hold at 15 °C). This extra annealing step was not essential when using TSA products but increased efficiency in TSB tag capture.

3. During silane bead elution (Step 3.1o), beads were eluted in 43.5 µl of elution solution I, and 3 µl was kept aside to use as input in the tag library polymerase chain reaction (PCR), whereas the remaining 40 µl was used to proceed with SPRI cleanup as the protocol describes. We reasoned that some tag fragments could stay in the bound fraction during the 1.2× SPRI separation, so, to maximize tag capture, we recommend to include a small portion (up to 10%) of the silane bead elution as input in the tag indexing reaction.

During SPRI cleanup (Step 3.2d), the supernatant was saved, and an additional 0.8× reaction volume of SPRI beads (32 µl) was added to bring the ratio up to 2.0×. Beads were washed twice with 80% ethanol and eluted in EB. This fraction can be combined with the few microliters left aside after the silane purification to be used as input in the protein tag indexing reaction, or either source can be used alone with minimal effect on tag complexity (Extended Data Fig. 1). PCR reactions were set up to generate the protein tag library (P5 and RPI-x primers for TSA conjugates, P5 and D7xx_s for TSB conjugates) and the hashtag library (P5 and

D7xx_s) with the following program: 95 °C for 3 min; 14–16 cycles of 95 °C for 20 s, 60 °C for 30 s and 72 °C for 20 s; followed by 72 °C for 5 min and ending with hold at 4 °C. Example of an RPI-x primer (TruSeq Small RNA handle, present in TSA tags; 'x' nucleotides represent a user-defined sample index): CAAGCAGAAGACGGC-ATACGAGATxxxxxxxxGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA. Example of a D7xx_s primer (TruSeq DNA handle, present in TSB tags or TSA hashing): CAAGCAGAAGACGGCATACGAGATxxxxxxxxGTGAC-TGGAGTTCAGACGTGTGC. The final libraries were quantified using a Qubit dsDNA HS Assay Kit (Invitrogen) and a High Sensitivity DNA chip run on a Bioanalyzer 2100 system (Agilent).

Note: Both versions 1 and 1.1 of the scATAC kit were successfully used throughout this study, with no discernible differences with respect to protein tag detection.

Cell permeabilization for DOGMA-seq. After staining cells with the antibody panel as described above, cells were washed twice with CITE-seq buffer and FACS sorted to remove dead cells/debris and CD66b$^+$ cells. Sorted cells were harvested and permeabilized under two conditions: one that includes fixation and preserves mtDNA in the ATAC-seq libraries (termed LLL, similarly to mtscATAC-seq[14] and ASAP-seq) and one that omits fixation and treats cells with 0.01% DIG, as described in ICICLE-seq[40].

For LLL permeabilization, cells were first fixed with 0.1% formaldehyde (Thermo Fisher Scientific, no. 28906) in PBS/RI (PBS supplemented with 0.1% BSA and 0.2 U µl^{-1} of RNAse inhibitor) for 5 min at room temperature and quenched with glycine solution to a final concentration of 0.125 M, followed by washing twice in PBS/RI via centrifugation at 400*g* for 5 min at 4 °C. Fixed cells were subsequently treated with LLL lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl$_2$, 0.1% NP40, 1% BSA, 1 mM DTT and 2 U µl^{-1} of RNAse inhibitor) for 3 min on ice, followed by adding 1 ml of chilled LLL wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl$_2$, 1% BSA, 1 mM DTT and 1 U µl^{-1} of RNAse inhibitor) and inversion before centrifugation at 500*g* for 5 min at 4 °C. The supernatant was discarded, and cells were diluted in 1× diluted nuclei buffer supplemented with 1 mM DTT and 1 U µl^{-1} of RNAse inhibitor (as described by 10x Genomics), followed by counting using trypan blue and a Countess II FL Automated Cell Counter.

For DIG permeabilization, cells were treated with DIG lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 3 mM MgCl$_2$, 0.01% DIG and 2 U µl^{-1} of RNase inhibitor) for 5 min on ice, followed by adding 1 ml of chilled DIG wash buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 3 mM MgCl$_2$ and 1 U µl^{-1} of RNAse inhibitor) and inversion before centrifugation at 500*g* for 5 min at 4 °C. The supernatant was discarded, and cells were resuspended in DIG wash buffer, followed by counting using trypan blue and a Countess II FL Automated Cell Counter.

Transposition and barcoding for DOGMA-seq. Cell were processed according to the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression user guide (version CG000338 Rev A, 10x Genomics) with the following modifications:

1. During pre-amplification PCR (Step 4.1), 1 µl of 0.2 µM ADT additive primer (CCTTGGCACCCGAGAATT*C*C) was spiked in the reaction mix.

2. After pre-amplification PCR and SPRI cleanup (Step 4.3.k), the beads were eluted in 100 µl of EB buffer. Then, 25% of the pre-amplified sample was used as input for the ATAC library indexing, and 35% was used as input in the cDNA amplification reaction.

3. To amplify and index protein tags, 35% of the pre-amplified sample was used in PCR reactions with SI-PCR (AATGATACGGCGACCACCGAGATCTA-CACTCTTTCCCTACACGACGCTC) and RPI-x primers, with the following program: 95 °C for 3 min; ten cycles of 95 °C for 20 s, 60 °C for 30 s and 72 °C for 20 s; followed by 72 °C for 5 min and ending with hold at 4 °C. Example of an RPI-x primer (TruSeq Small RNA handle, present in TSA tags, 'x' nucleotides present a user-defined sample index): CAAGCAGAAGACGGCATAC-GAGATxxxxxxxxGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA. The final libraries were quantified using a Qubit dsDNA HS Assay Kit (Invitrogen) and a High Sensitivity DNA chip run on a Bioanalyzer 2100 system (Agilent).

Flow cytometry. For flow cytometry analysis, PBMCs were washed in FACS buffer (2% FBS in PBS) before antibody staining using BV421-conjugated CD19 (HIB19, 302233, BioLegend), CD3 (UCHT1, 300433, BioLegend), CD4 (RPA-T4, 300531, BioLegend) and CD11c (Bu15, 337225, BioLegend), each at a 1:100 dilution. After washing, fixation and permeabilization were conducted as described in the 'Cell fixation and permeabilization' section above, before cells were resuspended in nuclei dilution and ATAC buffer and incubated for 1 h at 37 °C in a thermocycler to mimic the Tn5 transposition step during (mt)scATAC-seq. Aliquots for flow cytometry analysis were processed at indicated stages as schematically depicted in Extended Data Fig. 1a. Bead staining was similarly performed using BD CompBeads (552843, BD Biosciences) at a 1:100 antibody dilution. Analysis was conducted on a BD Biosciences FACSAria III system. Data were analyzed using FlowJo software version 10.4.2. Briefly, lymphocytes and monocytes were separated using forward versus side scattering gating, followed by sub-gating on

the fluorophore-positive fraction (for stained cells/beads) or fluorophore-negative fraction (for unstained cells/beads). Fluorescence intensity histograms were produced from these gated fractions using the 'Histogram' function.

PBMC stimulation. Cryopreserved PBMCs were thawed and washed in complete medium (RPMI GlutaMAX, supplemented with 10% FCS and 50 IU ml^{-1} of IL-2). Cells were allowed to rest in complete medium for 30 min at 37 °C before filtering through a 70-μm cell strainer to remove aggregates. PBMC aliquots were split in half and resuspended to a final density of 1×10^6 per milliliter in either complete medium (unstimulated control) or complete medium supplemented with ImmunoCult Human CD3/CD28 T cell Activator (stimulated sample) according to manufacturer instructions (STEMCELL Technologies). Then, 200-μl cell suspension aliquots were deposited in a 96-well round-bottom plate and placed in a humidified 5% CO$_2$ incubator at 37 °C for 16 h. Cells from respective wells were pooled, harvested, washed 2× with media and resuspended in 1 ml of media before filtering through a 70-μm cell strainer to remove cell aggregates. About 1×10^6 cells of each condition were then harvested and resuspended in 100 μl of CITE-seq staining buffer in preparation for staining. We note that these stimulation conditions result in a depletion of monocytes due to their adherence to the plastic[70].

Arrayed Cas9 ribonucleotide protein preparation and electroporation. Lyophilized CRISPR RNAs (crRNAs) and trans-activating CRISPR RNAs (tracrRNAs) (Integrated DNA Technologies) were reconstituted to a concentration of 400 μM and stored at −80 °C until use. crRNAs and tracrRNAs were mixed at a 1:1 vol/vol ratio, transferred into a 96-well plate and heated at 95 °C for 5 min, followed by incubation at room temperature for 15 min to complex the gRNAs. Then, 30 μg of Cas9 protein (Takara Bio, cat. no. Z2640N) was added to each well and mixed by gentle pipetting, followed by incubation at room temperature for 15 min. Complexed ribonucleotide proteins (RNPs) were then dispensed in a 96-well V-bottom plate at 12.7 μl per well. Cells were resuspended in Lonza P2 primary nucleofection buffer at 1×10^6 cells per 20 μl and added to the RNP-containing V-bottom plate. The mixture was gently mixed by pipetting and then transferred into a 16-well electroporation cuvette plate (Lonza, cat. no. V4XP-2032) and pulsed with the EH100 program. Immediately after electroporation, 100 μl of pre-warmed T cell culture medium was gently added to each well, and cells were incubated at 37 °C for 10 min. Cells were then transferred into 96-well U-bottom plates for culture at 1×10^6 cells per milliliter, supplemented with 500 IU ml^{-1} of IL-2. A list of all crRNAs used in this study can be found in Supplementary Table 5.

Multiplexed perturbation workflow. Primary human CD4$^+$ T cells were enriched by magnetic negative selection using the human CD4$^+$ T Cell Isolation Kit (Miltenyi Biotec, cat. no. 130-096-533) as per manufacturer instructions. Cells were then stained, and naive CD4 T cells were sorted on a BD FACSAria SORP system on the basis of CD4 and CD45RA expression. After isolation, cells were cultured in T cell culture medium consisting of RPMI with 10% FBS, 10 mM HEPES, 2 mM GlutaMAX (Gibco, cat. no. 35050-061), 1× MEM non-essential amino acids (Gibco, cat. no. 11140-050), 1 mM sodium pyruvate, 55 μM 2-mercaptoethanol and 100 IU ml^{-1} of IL-2 at a density of 1×10^6 cells per milliliter and stimulated with anti-human CD3/CD28 Dynabeads (Thermo Fisher Scientific, cat. no. 11131D) at a 1:1 cells-to-beads ratio. Seventy-two hours after stimulation (Day 3), beads were removed, and cells were rested in media containing IL-2 for expansion, while maintaining at a density of 1×10^6 cells per milliliter. On Day 7, cells were electroporated with Cas9 RNP complexes. After electroporation, cells were cultured in media with 500 IU ml^{-1} of IL-2 and split regularly to maintain a density of 1×10^6 cells per milliliter. On Day 15, cells were re-stimulated with anti-human CD3/CD28 Dynabeads (Thermo Fisher Scientific, cat. no. 11132D), supplemented with 100 IU ml^{-1} of IL-2. Seventy-two hours later, beads were removed, and cells for each condition were stained and washed as described above with a combination of two specific TSA hashtag antibodies (0.25–0.5 μg per antibody). Live cells were enriched and pooled by cell sorting on a BD FACSAria SORP and then processed as per the ASAP-seq protocol described above using OMNI lysis conditions.

Next-generation sequencing of DNA amplicons. Next-generation sequencing of gDNA was performed essentially as previously described[71]. Cells transfected with Cas9 were harvested 8 d after electroporation, enriched for live cells by cell sorting on a BD FACSAria SORP and then processed for gDNA extraction using the DNeasy Blood & Tissue Kit (Qiagen, cat. no. 69504) following the manufacturer's instructions. Genomic sites of interest were first amplified by PCR with Phusion High-Fidelity DNA Polymerase (NEB) using gene-specific primers (primer sequences are listed in Supplementary Table 5). A second round of PCR was performed using 1 μl of product of the first PCR reaction to barcode the samples for next-generation sequencing. PCR products of the barcoded reaction were verified by running on agarose gel and then extracted using the MinElute Gel Extraction Kit (Qiagen, cat. no. 28604) as per manufacturer recommendations with a final elution volume of 30 μl in EB buffer. Amplicon libraries were sequenced single ended 1 × 150 bp on an Illumina NextSeq machine. After demultiplexing, FASTQ files were analyzed using CRISPResso2 (ref. [72]).

Sequencing data pre-processing. Raw sequencing data for both scATAC-seq and antibody tag libraries were demultiplexed using cellranger-ATAC mkfastq. For the ATAC data, sequencing reads for all libraries were aligned to the hg38 or hg38/mm10 reference genomes using cellranger-ATAC count. To eliminate barcode multiplets[73], all libraries were processed with Cell Ranger ATAC v1.2, which uses shared Tn5 transposition events to identify and remove barcodes with low tag abundance. Protein tag abundances were estimated using the kallisto, bustools and kite frameworks[74,75]. To make the protein tag reads compatible with the kallisto framework processing, we developed an accessory script, ASAP_to_kite. py, that converts FASTQ files into a format similar to the 10x scRNA-seq format, enabling tag abundance quantification. For CITE-seq data, raw sequencing reads were aligned using Cell Ranger version 3 to the hg38 reference genome. For DOGMA-seq analyses, raw sequencing reads were processed using Cell Ranger ARC (v1.0) for ATAC and RNA using a modified reference genome with mtDNA mapping regions masked[14]. Tag abundances for all experiments were computed directly using the kallisto, bustools and kite frameworks[74,75].

Analysis of species mixing experiment. Cells that passed the Cell Ranger ATAC knee call were assigned as putative human cells when at least 100 fragments overlapped accessibility peaks in the human reference genome and putative mouse for at least 100 fragments in peaks in the murine reference genome. Similarly, cells were annotated as putative mouse or human cells based on protein abundance based on a minimum count of 100 for human CD29 and 50 for mouse CD29. Doublets were assigned for cells that consisted of less than 95% (ATAC; fragments in peaks) or 90% (protein; CD29 abundance) of the corresponding molecule. All thresholds were determined after evaluation of empirical densities of these measurements. The percent agreement between the multimodal assays was determined using cells that had corresponding labels (mouse and human doublet), which was 97.4% for the pre-SPRI experiment and 97.1% for the post-SPRI experiment. For each experiment, only one cell was observed that was annotated as mouse in one modality and human in the other; the rest of the discrepancies were due to edge cases associated with doublet assignments.

Complexity analyses. For both protein tag and chromatin complexity estimations, we used the number of unique and duplicate fragments as part of the Cell Ranger ATAC (chromatin) and bustools (tag) output as inputs into the Lander–Waterman equation[76], which estimates the total number of unique molecules present given these two measurements. For chromatin, we used the 'total' and 'passed_filters' columns from the singlecell.csv file. For the tag libraries, we converted the corrected bustools file into a TSV file to manually assign and de-duplicate reads based on error-corrected barcode, UMI/UBI and feature assignments. For species mixing experiments, comparisons were performed by selecting the top 1,000 cells ranked by library complexity per condition per species to minimize differences due to variable cell yield (Extended Data Fig. 1d,e).

Resting PBMC analyses. For all analyses in Figs. 1–3, gene activity scores, cell clusters and reduced dimension representations were computed using ArchR[77] with the default workflow. Visualizations of gene activity scores and protein tag abundances were performed using unsmoothed values after CLR normalization for the protein tags. From the cell hashing experiment (Extended Data Fig. 1g), we assigned putative cell doublet identities using HTODemux[17] for all barcodes passing the Cell Ranger ATAC knee call. Heteroplasmic mtDNA mutations were determined using the mgatk pipeline and variant calling parameters as previously described[14]. The two mutations shown in Fig. 1f were selected as they had the highest mean allele frequency among high-confidence heteroplasmic mutations. Violin plots depicting the proportion of mtDNA fragments (Fig. 2c) and tag abundances (Fig. 2d and Extended Data Fig. 2c) were plotted after removing the top 1% of barcodes from the Cell Ranger ATAC knee called for each value to minimize the visual effect of artifacts, such as cell doublets.

Bone marrow mononuclear cell analyses. We identified high-quality cells that satisfy three criteria: (1) minimum TSS score >4 and 1,000 fragments from ArchR[77]; (2) are not doublets based on hashtag oligos/HTODemux[17]; and (3) have fewer than 10,000 total tags or 50 tags in rat antibodies (cutoffs inferred from density distributions). These steps resulted in 10,928 cells. We then performed latent semantic indexing (LSI), uniform manifold approximation and projection (UMAP) and clustering with ArchR using default settings, which includes calls to the Seurat FindClusters function for Louvain cluster determination on the shared nearest neighbors graph[77]. Annotations of cellular protein tags were performed using CLR-normalized counts among these barcodes. Tag importance was determined after fitting a random forest model using the chromatin-derived cluster labels as outcomes and scaled, CLR-normalized protein tag abundances as input features, an approach inspired by the CiteFuse workflow[26].

Monocytic and erythroid pseudotime was determined using the semi-supervised functionality in ArchR[77]. Protein tag/pseudotime heat maps were computed by dividing cells into 100 bins, computing means and then performing a rolling average over 11 consecutive bins as implemented in ArchR[77]. The subset of proteins shown for each lineage were selected such that (1) the mean scaled protein tag value exceeded 1 across cells in the trajectory, and (2) the ratio of means

between cells included and excluded in the trajectory exceeded 1. These filtering steps for included proteins were incorporated to minimize the contribution of factors not specifically expressed in these differentiation trajectories. The paired gene score heat maps were computed using the same procedure but using the single-cell, unsmoothed gene activity scores. Finally, we further restricted the set of genes for the comparison of max protein tag and max gene activity score (Fig. 3i) to genes where the protein peaked after 0.25 in the pseudotime directory to eliminate factors associated with multipotent or erythroid-biased progenitors.

Among the cells passing accessible chromatin and protein quality control, 6,797 had a minimum 10× mtDNA coverage, which were considered for downstream mutation analysis. Heteroplasmic mtDNA mutations were determined using the mgatk pipeline and variant calling parameters as previously described[14]. Putative lineage-biased variants were identified using a per-mutation Kruskal–Wallis test of association between heteroplasmy and cell lineage, which were assigned to individual cells based on chromatin clusters (Supplementary Table 2).

Analysis of PBMC stimulation experiments. Control and stimulated CITE-seq cells were filtered using the following criteria: predicted singlets using Scrublet[78], maximum 10% mtRNA reads and minimum 500 genes detected and minimum 1,000 total UMIs observed. Cells were further filtered out if they had excess abundance of total protein tags (>25,000 or >30,000 in control and stimulated conditions, respectively) or tags measured from the rat isotype controls (>55 or >65 in the control and stimulated, respectively). Similarly, we identified high-quality cells from the ASAP-seq dataset such that each cell had a TSS score exceeding 4 and a minimum of 1,000 fragments. Cells were further filtered out if they had excess abundance of total protein tags (>25,000 in either condition) or tags measured from the rat isotype controls (>75 in either condition). All thresholds for both the ASAP-seq and CITE-seq filtering were determined by evaluating the per-cell empirical density.

We performed two-stage data integration for the ASAP-seq and CITE-seq datasets to preserve the biological effect of the stimulation and residualize differences between the RNA and ATAC assays. First, we created a union of variable genes from the CITE-seq stimulated and control datasets along with genes whose proteins were measured as part of the antibody panel. Using these ~2,700 genes, we performed canonical correlation analysis (CCA) between the stimulated ASAP-seq (gene scores computed from Signac[79]) and CITE-seq (RNA abundance) datasets and a second round of CCA between the control ASAP-seq and CITE-seq datasets[80]. For both datasets, we imputed RNA expression for the ASAP-seq objects using transfer anchors as described in Seurat version 3 (ref. [81]). In RNA space for these two merged objects, we performed principal component analysis (PCA), before using Harmony to integrate the stimulated and control integrated datasets[82]. A final dimensionality reduction and clustering using Harmony components was performed to summarize both modalities (ASAP-seq and CITE-seq) and both biological conditions (stimulated and control) in one setting. Finally, the embedding and clustering of ASAP-seq and CITE-seq based on protein tag abundances was performed using the 100 most variable features across the merged ASAP-seq and CITE-seq datasets as inputs to PCA and then Harmony[82] to account for the technology and stimulation status as two group variables.

In determining the relative changes among chromatin accessibility, RNA and protein abundance between the stimulated and unstimulated conditions, we generated counts per million-normalized pseudo-bulk abundances, which were used to determine the log₂ fold changes. Although these measures were computed for both the B cell and T cell clusters separately, we note that many changes in the B cell population mirrored that of the T cells, which we attributed to low-frequency cell doublets that persisted even after our computational filtering. This inference was based on the presence of markers such as CD4 and CD8 appearing in the B cell clusters, which are markers restricted to T cells and largely unchanged in the stimulation.

Separately, the numbers of differential peaks, genes and proteins were computed using a per-peak permutation test[9], the edgeRQLFDetRate for differential gene expression[83] and a Mann–Whitney test for the CLR protein abundances. The number of significant differential features (Fig. 4) was determined using consistent thresholds of a Benjamini–Hochberg-adjusted P value of 0.01 and a minimum magnitude of log₂ change exceeding 0.5. The proportion of differential features was computed out of 52,551 peaks, 10,533 genes and 227 proteins. For accessibility peaks and genes, the universe of those tested were selected based on a mean count per million exceeding 2 across the stimulated and control samples. For the proteins, none of the 71 differentially expressed markers using these criteria was the rat isotype antibodies (known negative controls).

Tag importance (Extended Data Fig. 4d) for either the stimulation or cluster identities was determined by fitting two random forest models to (1) the cluster labels and (2) the experimental control/stimulation conditions. Again, the scaled, CLR-normalized protein tag abundances were used as input features, an approach inspired by the CiteFuse workflows[26].

UMAPs showcasing changes in chromatin accessibility, RNA and protein abundance were consistently displayed using the 2nd and 98th percentiles as minimum and maximum values on the color scale. Cells depicted were displayed in random order. Furthermore, only cells where the modality was directly measured (that is, chromatin accessibility: ASAP-seq; RNA: CITE-seq; protein: ASAP-seq and CITE-seq) were displayed, and no gene smoothing was applied in any display.

For the DOGMA-seq analyses, we first performed cell filtering on each modality, requiring cells to have high-quality measurements for all modalities (Online Code). To perform 3WNN dimensionality reduction, we first ran Harmony[82] on the linear components for each reduction with the stimulation/control status as the covariate. Next, we used the FindMultiModalNeighbors function in Seurat version 4 (ref. [27]) using the first 30 Harmony-adjusted components for all three modalities (although excluding the first scATAC component due to correlation with sequencing depth). To examine the effect of each modality, we computed the adjusted Rand index per pair of modalities (Extended Data Fig. 5k) on clusters defined from the default FindClusters execution of the FindMultiModalNeighbors function for two modalities. To determine a per-cell, per-modality activation score, we used the AddModuleScore and AddChromatinModule functions for the top features upregulated in the stimulation from the Fig. 4 ASAP-seq/CITE-seq experiment (top 2,000 peaks; top 1,000 genes; top 20 proteins). Heteroplasmic and homoplasmic mtDNA mutations were determined using the mgatk pipeline and variant calling parameters as previously described[14]. Putative lineage-biased variants were again identified using a per-mutation Kruskal–Wallis test of association.

Multiplexed perturbation analyses. We first de-noised the hashtag count matrix using dsb[84], and, then, we assigned perturbation identity using HTODemux[17]. Donor ID per cell was further inferred using popscle, which extends the demuxlet toolkit[85]. High-quality cells were determined based on quality control criteria using the Signac[79] workflow, focusing on high-quality cells to maximize the inference of the CRISPR perturbations. Subsequently, these quality-controlled cells were used in generating LSI dimensions and the UMAP embedding using ArchR with default settings[77]. TF accessibility deviation scores were computed using chromVAR with default settings for known human TF motifs, including the inference of the top 100 most variable[44]. Downstream analyses of protein tag abundances were performed on CLR-normalized tag abundances. NFKB-compatible motifs were discovered in chromatin accessibility peaks using the motifmatchr framework as part of the chromVAR[44] suite of tools. Pseudo-bulk genomic loci tracks were generated by first subsetting gRNA-specific cell barcode reads using sinto (https://github.com/timoast/sinto), followed by processing with MACS2 (ref. [86]) under the options -B --no model, --extsize 150, --shift 75, --SPMR.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data are available at the Gene Expression Omnibus under accession number GSE156478.

Code availability

Custom code to reproduce all analyses and figures is available at https://github.com/caleblareau/asap_reproducibility.

References

70. Thurner, B. et al. Generation of large numbers of fully mature and stable dendritic cells from leukapheresis products for clinical application. *J. Immunol. Methods* **223**, 1–15 (1999).
71. Joung, J. et al. Genome-scale CRISPR–Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).
72. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
73. Lareau, C. A., Ma, S., Duarte, F. M. & Buenrostro, J. D. Inference and effects of barcode multiplets in droplet-based single-cell assays. *Nat. Commun.* **11**, 866 (2020).
74. Melsted, P., Ntranos, V. & Pachter, L. The barcode, UMI, set format and BUStools. *Bioinformatics* **35**, 4472–4473 (2019).
75. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
76. Lander, E. S. & Waterman, M. S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231–239 (1988).
77. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
78. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291 (2019).
79. Stuart, T., Srivastava, A., Lareau, C. & Satija, R. Multimodal single-cell chromatin analysis with Signac. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.09.373613 (2020).
80. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
81. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
82. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

83. Soneson, C. & Robinson, M. D. Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* **15**, 255–261 (2018).
84. Mulè, M. P., Martins, A. J. & Tsang, J. S. Normalizing and denoising protein expression data from droplet-based single cell profiling. Preprint at *bioRxiv* https://doi.org/10.1101/2020.02.24.963603 (2020).
85. Kang, H. M. et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).
86. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).

Author contributions
E.P.M. and P.S. conceived and designed the methods with input from L.S.L., K.Y.C. and C.A.L. E.P.M., K.Y.C., L.S.L. and P.S. designed experiments with input from C.A.L., R.S., V.G.S. and A.R. E.P.M., K.Y.C., E.P., W.L., P.I.T., T.K., M.H. and L.S.L. performed experiments. C.A.L. led data analysis with substantial contributions from E.P.M., K.Y.C., Y.H. and Y.T. A.L.Z.-F., T.-S.H., B.Y. and K.L.N. provided insights and developed key reagents and protocols for protein detection. J.B.W. provided insights and discussions for experimental planning. R.S., S.S., L.S.L., V.G.S., A.R. and P.S. each supervised various aspects of the work. E.P.M., C.A.L., K.Y.C., L.S.L. and P.S. drafted the manuscript with input from all other authors.

Competing interests
C.A.L., L.S.L., V.G.S. and A.R. are listed as co-inventors on a patent related to mtscATAC-seq (U.S. provisional patent application 62/683,502). In the past 3 years, R.S. has worked as a consultant for Bristol-Myers Squibb, Regeneron and Kallyope and served as a scientific advisory board member for Immunai and Resolve BioSciences. A.R. is a founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas Therapeutics and, until August 31, 2020, was a scientific advisory board member of Syros Pharmaceuticals, Neogene Therapeutics, Asimov and Thermo Fisher Scientific. From August 1, 2020, A.R. is an employee of Genentech. P.S. is listed as co-inventor on a patent related to this work (U.S. provisional patent application 62/515-180).
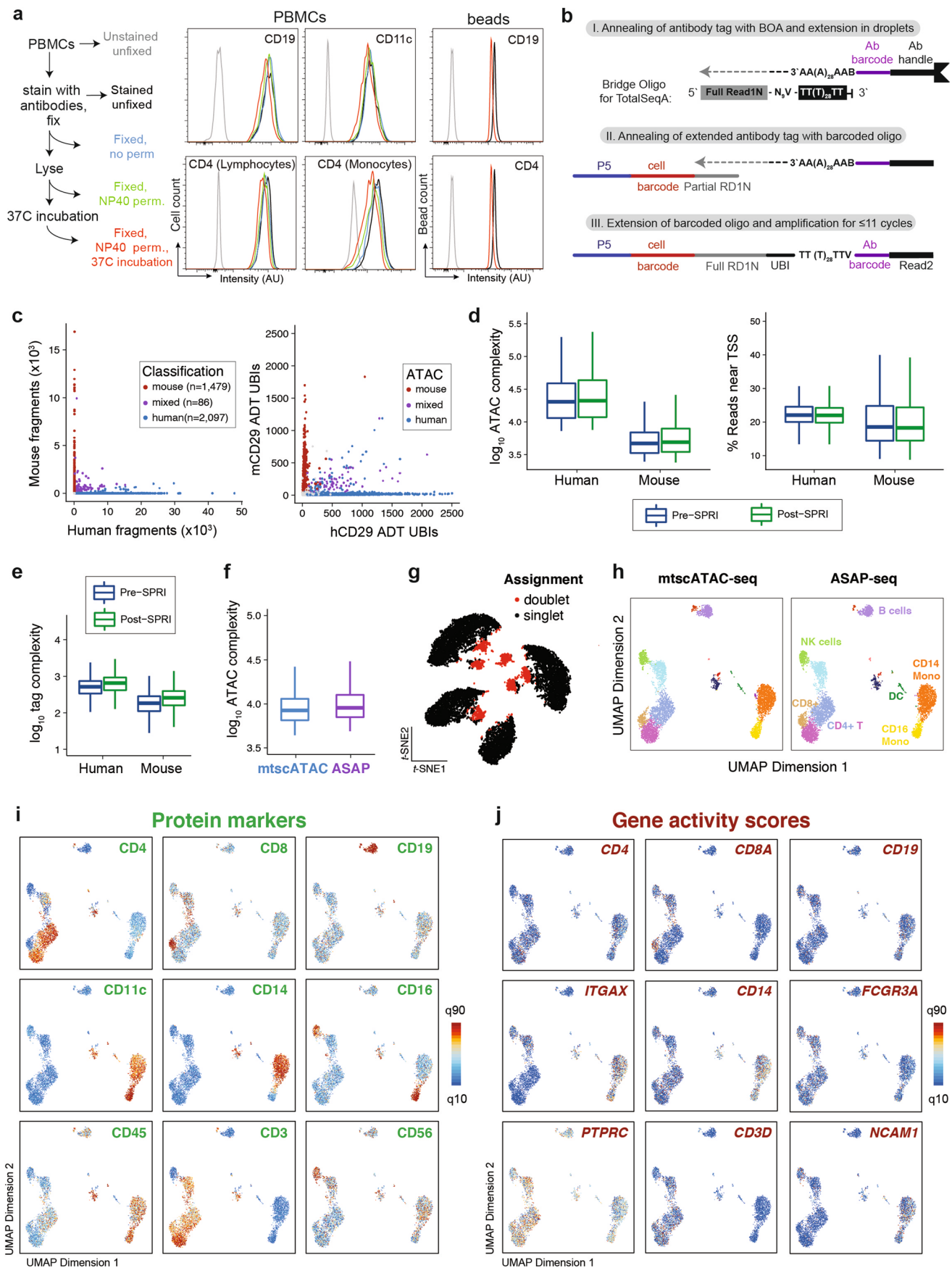
Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41587-021-00927-2.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-021-00927-2.

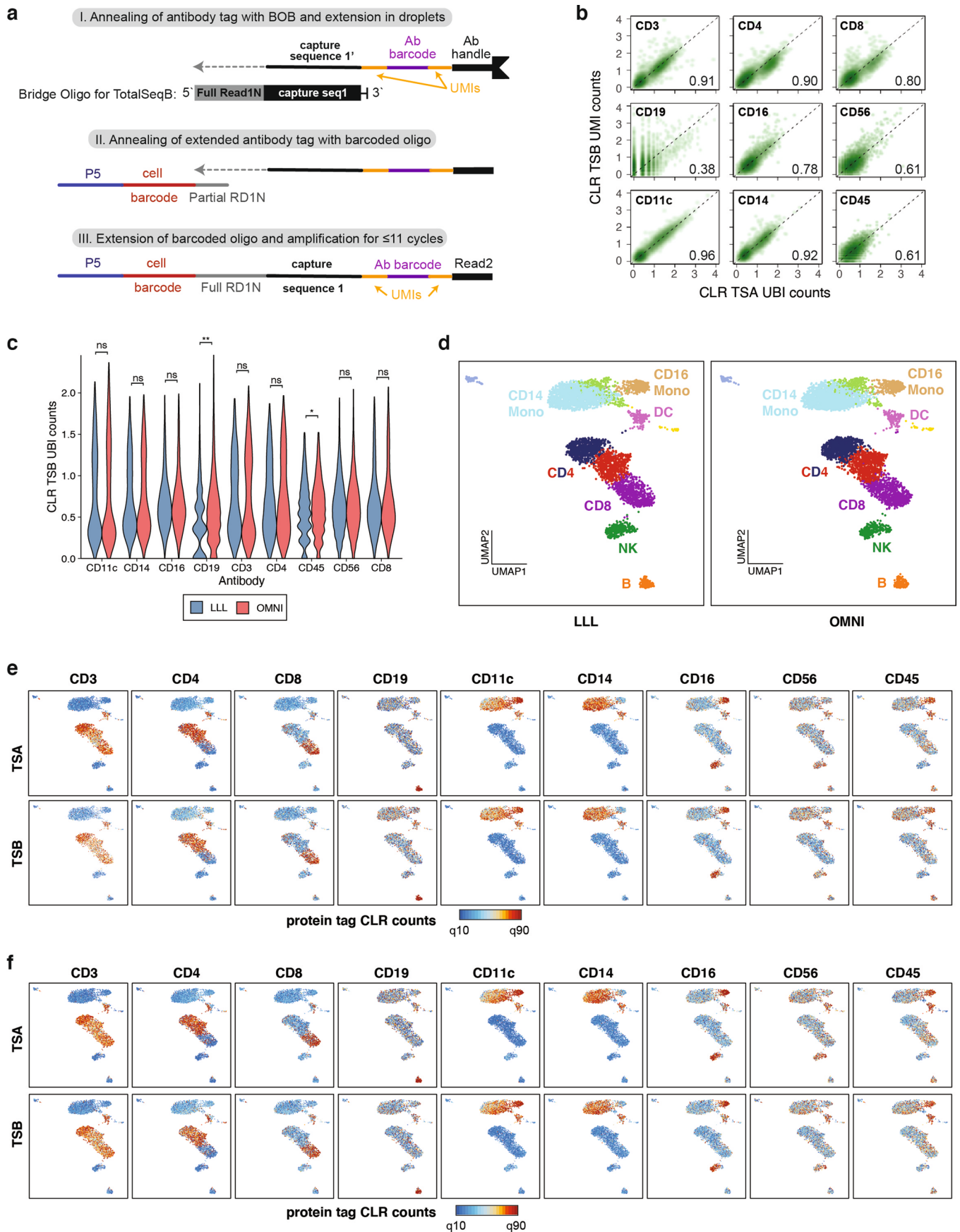**Correspondence and requests for materials** should be addressed to P.S.

Peer review information *Nature Biotechnology* thanks Dan Xie, Golnaz Vahedi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

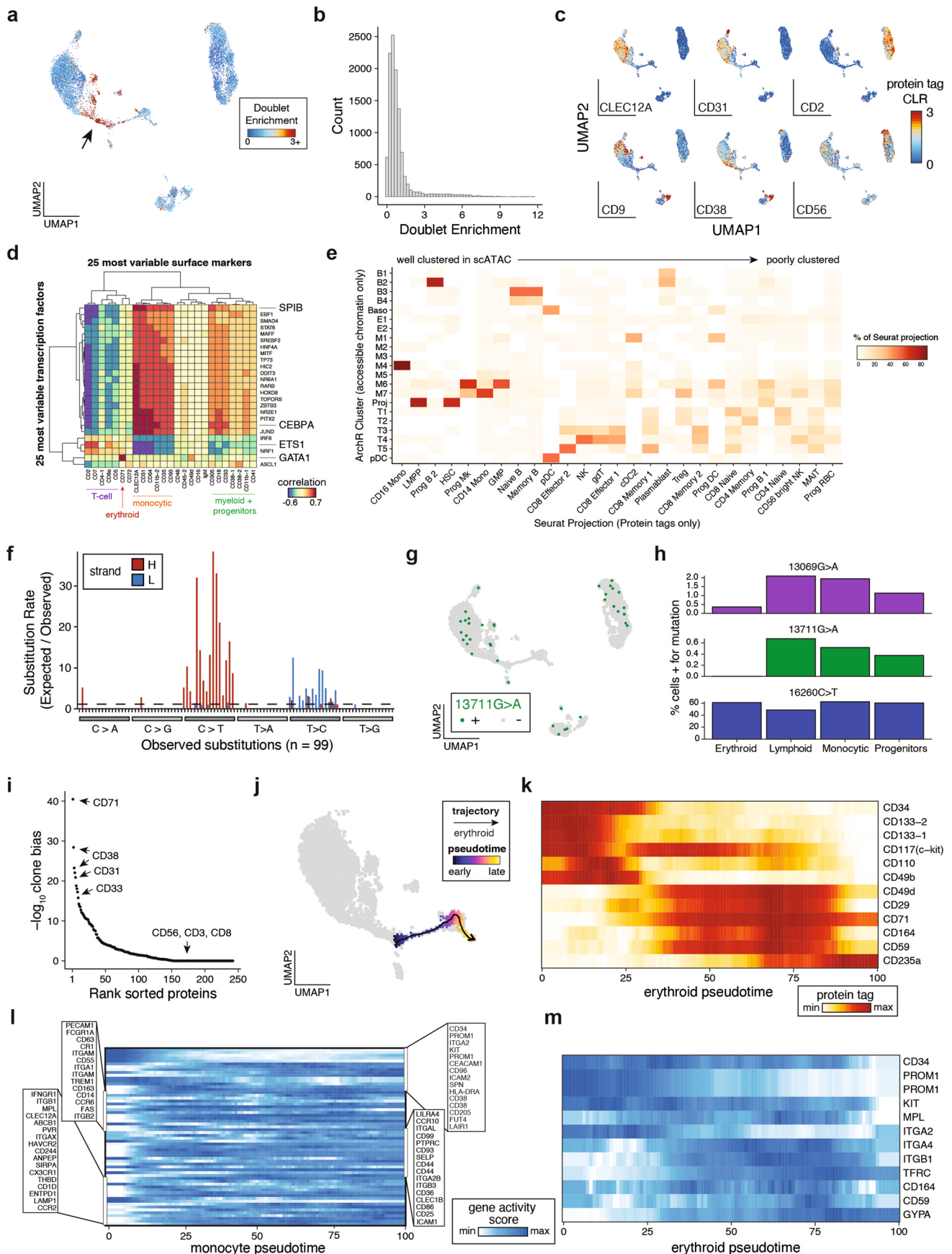**Reprints and permissions information** is available at www.nature.com/reprints.

Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Additional technical and computational validation of ASAP-seq workflows. a. PBMCs and compensation beads were stained with fluorophore-conjugated antibodies and subjected to the ASAP-seq workflow with samples withdrawn at the indicated steps and assessed for fluorophore intensity by flow cytometry. CD19 (staining B cells), CD11c (dendritic cells) and CD4 (lymphocytes and monocytes) signal on fixed cells is hardly affected by permeabilization alone, but after the 37 °C incubation for 1 h to mimic the Tn5 transposition reaction, some signal reduction is observed. **b**. Barcoding scheme of TSA tags using the bridge oligo for TotalSeq™-A (BOA). TSA tags do not contain UMIs, so to allow molecule counting, UBIs (N9V) are incorporated via the bridge oligo. **c**. Species mixing experiment as in Fig. 1c, using the Post-SPRI approach for tag recovery. Points are colored based on species classification using ATAC fragments. **d**. ATAC library complexity and TSS enrichment for fragments from each species under the two protein-tag library approaches. **e**. Comparison of protein tag complexity between libraries prepared using the pre- and post-SPRI approach. **f**. Comparison of ATAC library complexity between mtscATAC-seq and ASAP-seq. Boxplots: center line, median; box limits, first and third quartiles; whiskers, 1.5× interquartile range. **g**. Two-dimensional embedding of the PBMC hashing data using *t*-SNE. The four major clusters (black) correspond to the four hashing antibodies used to stain the PBMCs. 13,772 cells were recovered and 1,396 doublets (red) were detected. **h**. UMAP embedding resolving PBMC cell types based on chromatin accessibility for cells processed by mtscATAC-seq and ASAP-seq. Data for the two different samples were processed together using cell ranger-atac aggr before dimensionality reduction. **i,j**. Selected protein markers (**i**) and corresponding gene score activities (**j**) superimposed on the ATAC-clustered PBMCs (for the ASAP-seq sample) as in (**h**).

a

I. Annealing of antibody tag with BOB and extension in droplets

Bridge Oligo for TotalSeqB: 5` Full Read1N | capture seq1 | 3`

II. Annealing of extended antibody tag with barcoded oligo

III. Extension of barcoded oligo and amplification for ≤11 cycles

b

c

d

e

protein tag CLR counts   q10   q90

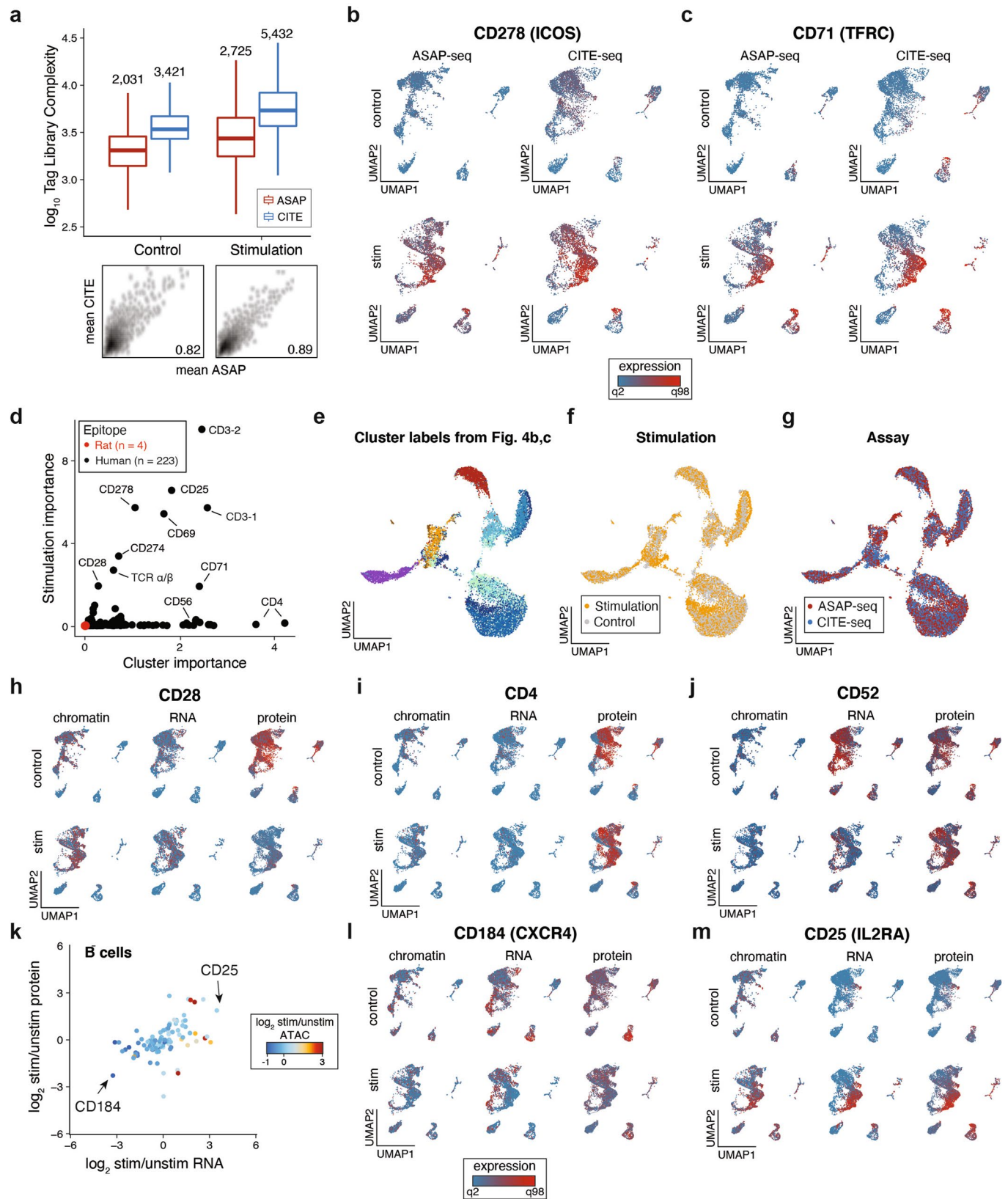**f**

protein tag CLR counts   q10   q90

Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Additional validation and comparison of modular ASAP-seq workflows. a. Barcoding scheme of TSB tags using the bridge oligo for TotalSeqB (BOB). TSB tags contain UMIs (encompassing the antibody barcode), negating the requirement for a UBI on the bridge oligo. **b**. Pairwise comparison of centered log-ratio (CLR) normalized TSA and TSB counts under OMNI lysis conditions (n = 5,236 cells). Counts were collapsed for unique molecules using UBIs (TSA panel) or UMIs (TSB panel). **c**. Comparison of CLR normalised TSB counts under the two lysis conditions. Statistical comparisons are two-sided Wilcoxon rank sum test with Bonferroni adjusted p-values (ns = not significant; * p_{adj} = 0.0002; ** p_{adj} = 2.2×10^{-16}). **d**. UMAP embedding and cluster annotation of the LLL (n = 5,236) and OMNI (n = 4,748) processed cells. Data for the two different samples were processed together using cell ranger-atac aggr before dimensionality reduction. **e**. TSA and TSB CLR counts projected on the LLL embeddings. **f**. TSA and TSB CLR counts projected on the OMNI embeddings.
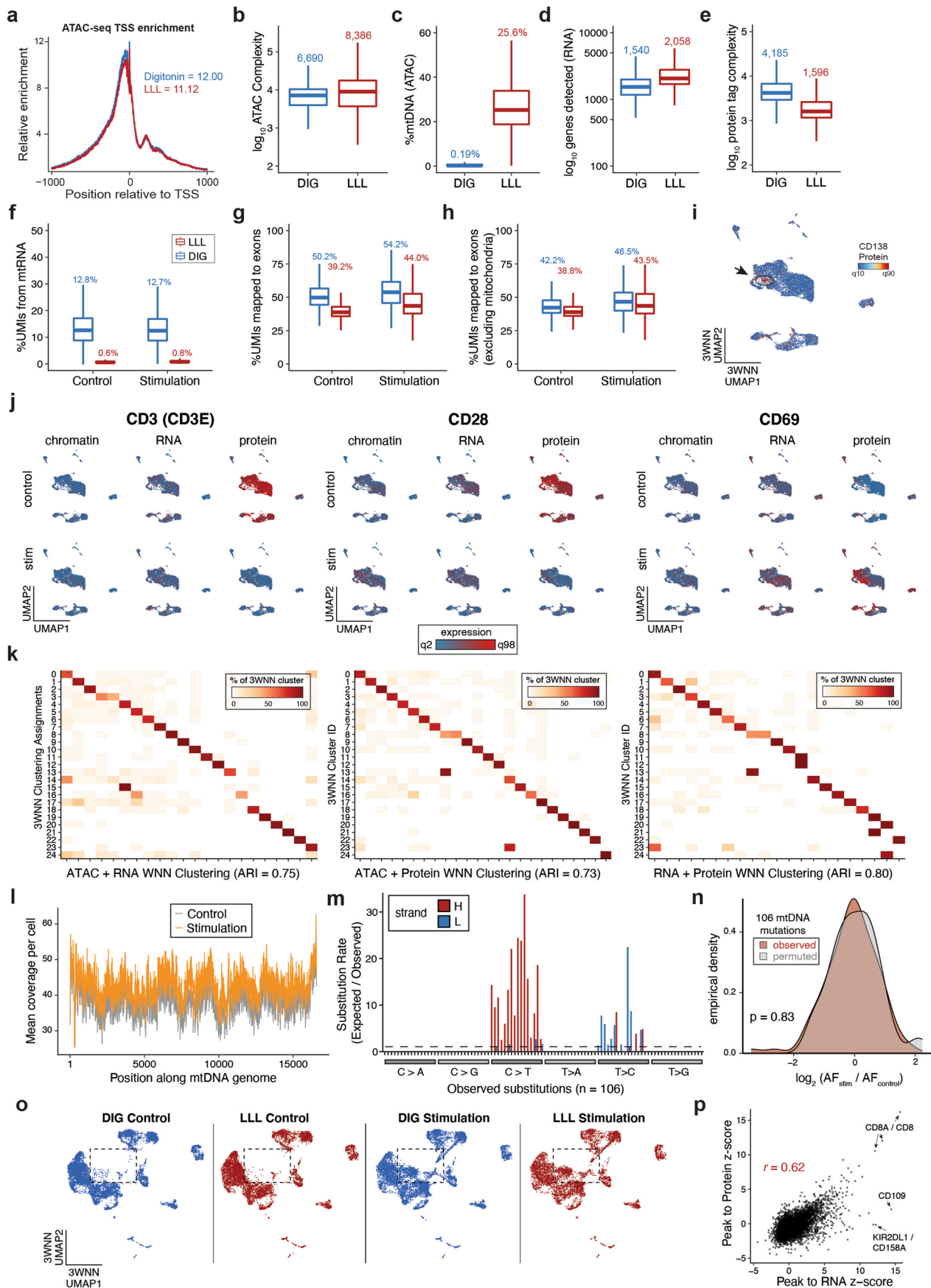
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Supporting information for ASAP-seq bone marrow analyses. a. Annotation of reduced dimension space with the Doublet Enrichment score from ArchR. Arrow indicates the monocytic progenitor population. **b**. Histogram of scores from panel (**a**). **c**. Feature plots for six additional antibody tags in the reduced dimension space. **d**. Correlation heatmap between 25 most variable TF activities and surface markers. **e**. Percent of cells in each ArchR cluster (y axis) mapping to the indicated Seurat cluster (x axis) after label transfer using the protein tags only **f**. Substitution rate (observed over expected) of mgatk-identified heteroplasmic mutations (y axis) in each class of mononucleotide and trinucleotide change resolved by the heavy (H) and light (L) strands of the mitochondrial genome. **g**. Projection of 13711 G > A in single cells; threshold for + was 5% heteroplasmy. **h**. Distribution of observed mtDNA mutations in cells among major cell lineages. **i**. Association of antibody tag abundance to cell clones determined by mtDNA genotypes, highlighting the erythroid marker CD71. **j**. Developmental trajectory of erythroid differentiation using semi-supervised pseudotime analysis. **k**. Expression of select cell surface markers along the erythroid developmental trajectory highlighted in (**j**). Rows are min-max normalized. **l**. Expression of chromatin activity scores along the monocytic developmental trajectory for genes encoding proteins shown in Fig. 3h. **m**. Expression of chromatin activity scores along the erythroid developmental trajectory for genes encoding proteins shown in (**k**).
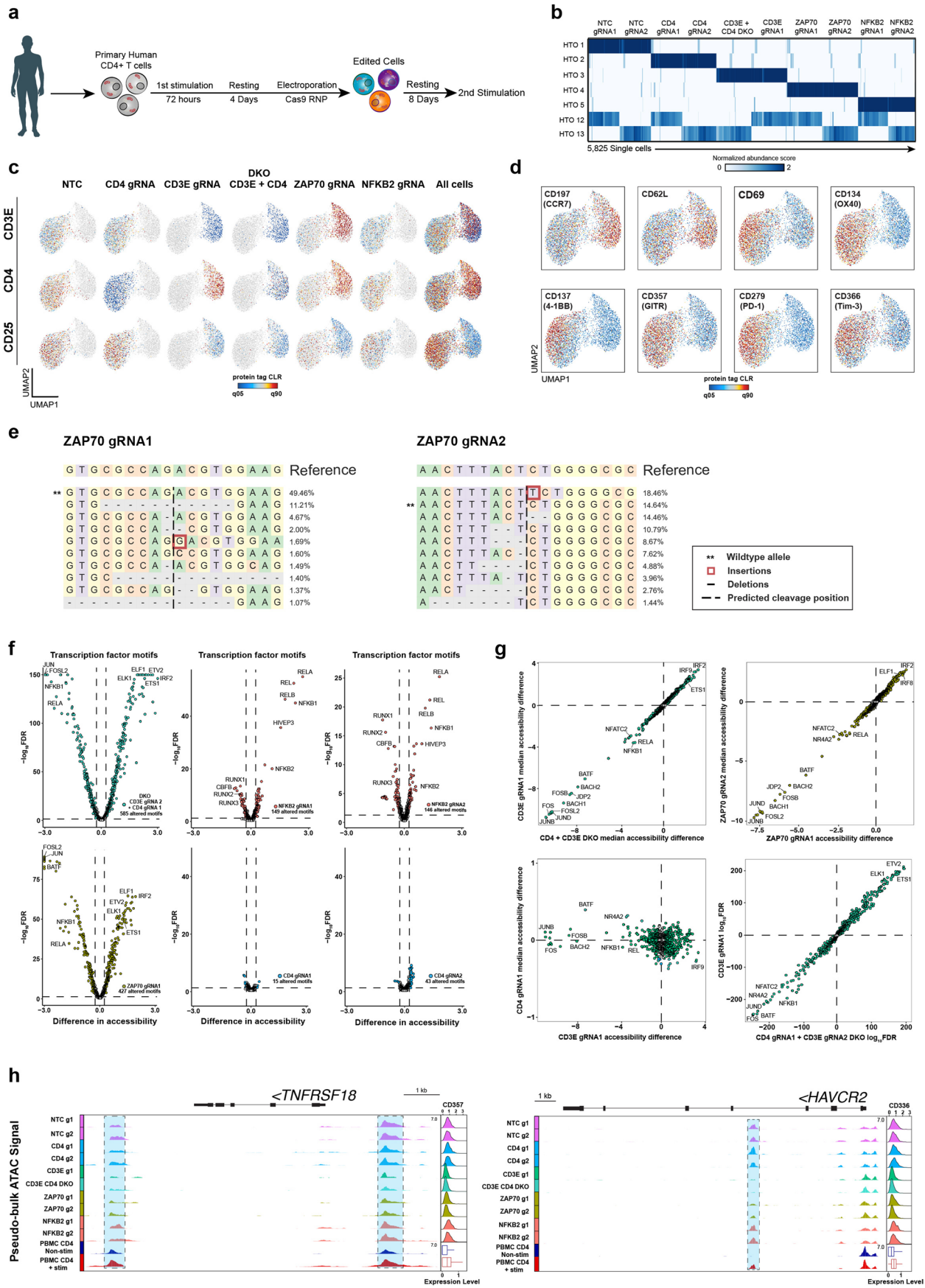
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Supporting information for combined ASAP-seq and CITE-seq readouts. a. Antibody tag complexity per condition and technology. Median tag complexity is 1.7-2x higher in CITE-seq compared to ASAP-seq and 1.3-1.6x higher in stimulation compared to control sample. Boxplots: center line, median; box limits, first and third quartiles; whiskers, 1.5× interquartile range. The lower panels show the per-cluster mean tag abundance for the 50 most variable antibodies and corresponding Pearson correlations. **b,c**. Cellular distribution of protein tags measured by ASAP-seq (left) and CITE-seq (right) for control (top) and stimulated conditions (bottom) for, (**b**) CD278 (*ICOS*) and (**c**) CD71 (*TFRC*). **d**. Protein tag measurement importance in predicting cell cluster and stimulation from two different Random Forest models. Negative controls (rat epitopes) are shown in red. **e-g**. ASAP-seq and CITE-seq data co-embedding utilizing protein abundances. Cells are highlighted by (**e**) chromatin/RNA cluster identity, (**f**) stimulation condition and (**g**) technology assayed. **h-j**. UMAPs of chromatin accessibility, mRNA expression, and surface protein levels for (**h**) CD28, (**i**) CD4, and (**j**) CD52. **k**. Summary of changes in chromatin accessibility, gene expression and surface protein abundance for 103 expressed genes in B cells following T cell stimulation. **l,m**. UMAPs of chromatin accessibility, mRNA expression, and surface protein levels for genes with differential expression in B cells, including (**l**) CD184 (*CXCR4*) and (**m**) CD25 (*IL2RA*).
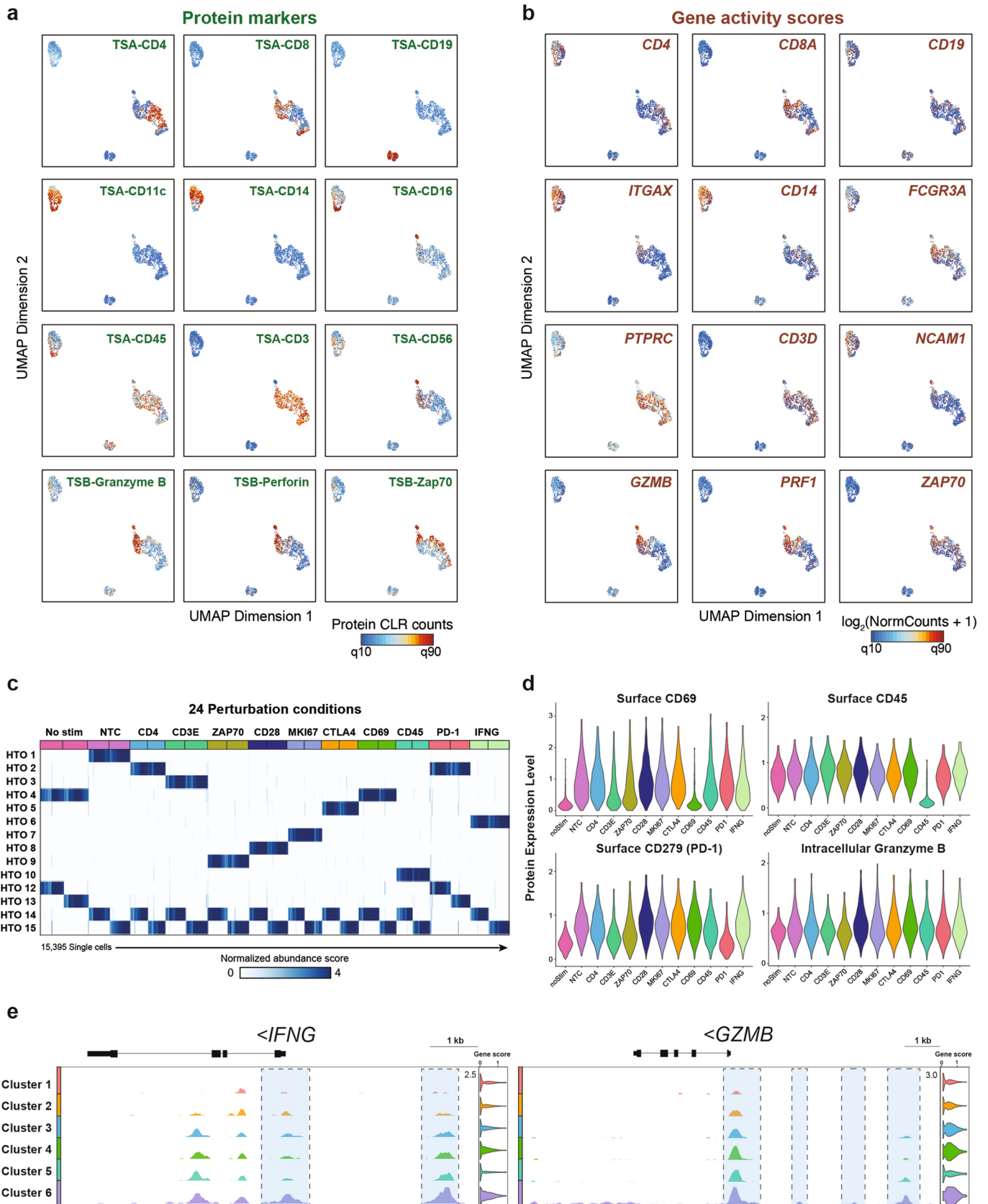
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Supporting information for DOGMA-seq. a-e. QC metrics of indicated modalities captured by DOGMA-seq applied on the stimulated PBMC sample. (**a**) TSS score, (**b**) ATAC fragment complexity, (**c**) % mtDNA content, (**d**) number of genes/cell and (**e**) protein tag complexity in the two different cell preparations compare similarly to the control PBMC sample in Fig. 5b-f. **f**. Percent of UMIs detected in the GEX library that map to mtRNA is higher in the digitonin-treated cells. **g-h**. Percent of UMIs mapping to exons is higher in the digitonin-treated (DIG) compared to LLL-treated cells (**g**), but similar when mitochondrial transcripts are excluded (**h**). **i**. CD138 tag counts projected on the three modality WNN stimulation clusters. **j**. Gene activity scores, transcript and protein tag counts projected for the indicated markers on the control and stimulated 3WNN clusters. **k**. Heatmaps showing percent overlap between clusters detected by 3WNN compared to 2WNN variations applied on the control PBMC dataset. **l**. Mean coverage along the mtDNA genome in control and stimulated PBMCs. **m**. Substitution rate (observed over expected) of mgatk-identified heteroplasmic mutations (y axis) in each class of mononucleotide and trinucleotide change resolved by the heavy (H) and light (L) strands of the mitochondrial genome for all cells in the PBMC-LLL condition. **n**. Observed (red) and permuted (gray) log2 heteroplasmy changes across the 106 identified variants. Statistical test: two-sided Kolmogorov–Smirnov Test. **o**. 3WNN UMAP embedding of control and stimulated PBMC samples under LLL and DIG processing. Dashed box indicates activated T cell clusters. **p**. Comparison of peak to gene linkage for genes detected in both protein and RNA modalities. Each dot is a peak to gene link with the z score representing the magnitude of the association. Boxplots: center line, median; box limits, first and third quartiles; whiskers, 1.5× interquartile range.

Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Supporting information for ASAP-seq based decoding of perturbations in primary T cells. a. Schematic for CRISPR perturbation experiment in primary human T cells. CD4 + T cells from healthy donors were stimulated for 72 hours, followed by a resting period of four days to enable expansion. On Day 7, cells were electroporated with Cas9 RNPs and then rested for an additional 8 days before secondary stimulation. **b**. Heatmap of cell demultiplexing with hashing antibodies, indicating normalized abundance of each hashtag. **c**. Assessment of the effect of CRISPR perturbations on three indicated protein surface markers. **d**. UMAP embedding overlaid with expression of the eight indicated surface protein markers. **e**. Allele-specific CRISPR editing outcomes for ZAP70 gRNA1 (left) and ZAP70 gRNA2 (right). The wildtype allele is indicated by **. **f**. Volcano plots showing TF motifs with significantly changed chromatin accessibility profiles between NTC cells and the indicated gRNAs (FDR < = 0.05, chromVAR accessibility change > = 0.25). **g**. Correlation of chromVAR median accessibility changes or FDR (bottom right panel) between the indicated gRNAs. **h**. Genomic tracks of TNFRSF18 and HAVCR2 loci with corresponding CLR-normalized protein abundance ridge plots. CLR-normalized protein abundance from the PBMC stimulation experiment is indicated by the corresponding boxplots. Differentially accessible regions are highlighted in blue.

a Protein markers

TSA-CD4 · TSA-CD8 · TSA-CD19
TSA-CD11c · TSA-CD14 · TSA-CD16
TSA-CD45 · TSA-CD3 · TSA-CD56
TSB-Granzyme B · TSB-Perforin · TSB-Zap70

UMAP Dimension 2

UMAP Dimension 1

Protein CLR counts
q10   q90

b Gene activity scores

CD4 · CD8A · CD19
ITGAX · CD14 · FCGR3A
PTPRC · CD3D · NCAM1
GZMB · PRF1 · ZAP70

UMAP Dimension 2

UMAP Dimension 1

$\log_2(\text{NormCounts} + 1)$
q10   q90

c 24 Perturbation conditions

No stim · NTC · CD4 · CD3E · ZAP70 · CD28 · MKI67 · CTLA4 · CD69 · CD45 · PD-1 · IFNG

HTO 1 – HTO 15

15,395 Single cells

Normalized abundance score
0   4

d

Surface CD69 · Surface CD45 · Surface CD279 (PD-1) · Intracellular Granzyme B

Protein Expression Level

noStim, NTC, CD4, CD3E, ZAP70, CD28, MKI67, CTLA4, CD69, CD45, PD1, IFNG

e

<IFNG · <GZMB

1 kb · Gene score 0 1

Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6

2.5 · 3.0

Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Supporting information for intracellular ASAP-seq workflow. a,b. Selected protein markers (**a**) and corresponding gene activity scores (**b**) superimposed on the ATAC-clustered PBMCs from the intracellular staining experiment (see Fig. 3a). **c**. Heatmap of cell demultiplexing with hashing antibodies, indicating normalized abundance of each hashtag for 24 different perturbation conditions. **d**. Violin plots showing distribution of CLR normalized protein counts for indicated proteins and their associated gRNA. **e**. Genomic tracks of IFNG and GZMB loci, indicating pseudo-bulk ATAC signal tracks across six Louvain clusters with corresponding log-normalized gene activity score violin plots shown to the right. Differentially accessible regions are highlighted in blue.

Corresponding author(s):	Peter Smibert
Last updated by author(s):	Apr 9, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	CellRanger-ATAC v.1.2.0; CellRanger v.3.0; MITOMAP v. r102; Seurat v.3.1; Seurat v.4.0.; Signac v.0.2.; chromVAR v1.10.0; AchrR v. 0.9.2; harmony v1.0; mgatk v0.5.8; kite | kallisto | bustools v0.02, v0.44.0, v0.39.3, respectively.
Data analysis	Custom code to reproduce all analyses is available at https://github.com/caleblareau/asap_reproducibility

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing and processed data generated as part of this work is available at GEO accession GSE156478

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed a priori. Analyses involved thousands of cells per comparison, providing a robust sample size in-line with similar high-throughput single-cell genomics assays.
Data exclusions	No experimental replicates were excluded from analysis. Single-cell filtering was performed in a dataset-specific manner. Common scATAC-seq quality control cutoffs as well as filters for protein abundances detected per single cell were applied to identify true cells as outlined in the methods.
Replication	All attempts at replication were successful. Verification of a successful library varied depending on the cell input materials but generally consisted of a pseudobulk TSS score > 5 and an average mtDNA content exceeding 15% for ATAC capture using mtscATAC and <1% using OMNI lysis. Verification of protein abundance using two different barcoded antibody captures are shown for 9 proteins in Figure 2.
Randomization	There were no variables or interventions to randomize in this study.
Blinding	Blinding is not relevant to our study, as our tools are not dependent on blinding. Investigators could not be blinded during data collection or analysis as there was no intervention. Further, analyses were performed in an exploratory manner where blinding is not possible.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	All antibodies used for this study are commercially available and listed on Supplemental Table 1
Validation	All are well-established clones, validated by the provider

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK-293 and NIH-3T3 were obtained from ATCC
Authentication	Cells were not authenticated. They were used for a proof of principle experiment where only species was important. Correct species was verified by species specific antibody staining and mapping of genomic reads.
Mycoplasma contamination	The aforementioned lines were not tested, but other lines used in the same hood/incubator have been routinely tested and have always been negative.
Commonly misidentified lines (See ICLAC register)	N/A