Single-cell profiling of proteins and chromatin accessibility using PHAGE-ATAC

Evgenij Fiskin[1]✉, Caleb A. Lareau[2], Leif S. Ludwig [1,3], Gökcen Eraslan[1], Feimei Liu[4], Aaron M. Ring[4,5], Ramnik J. Xavier[1,6,7] and Aviv Regev[1,8,9]✉

Multimodal measurements of single-cell profiles are proving increasingly useful for characterizing cell states and regulatory mechanisms. In the present study, we developed PHAGE-ATAC (Assay for Transposase-Accessible Chromatin), a massively parallel droplet-based method that uses phage displaying, engineered, camelid single-domain antibodies ('nanobodies') for simultaneous single-cell measurements of protein levels and chromatin accessibility profiles, and mitochondrial DNA-based clonal tracing. We use PHAGE-ATAC for multimodal analysis in primary human immune cells, sample multiplexing, intracellular protein analysis and the detection of SARS-CoV-2 spike protein in human cell populations. Finally, we construct a synthetic high-complexity phage library for selection of antigen-specific nanobodies that bind cells of particular molecular profiles, opening an avenue for protein detection, cell characterization and screening with single-cell genomics.

Massively parallel single-cell profiling has become an invaluable tool for the characterization of cells by their transcriptome or epigenome, deciphering gene regulatory mechanisms and dissecting cellular ecosystems in complex tissues[1–4]. In particular, recent advances have highlighted the power of multimodal single-cell assays[5], such as cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq), which profile both transcriptome and proteins by DNA-barcoded antibodies[6–10]. Although the vast combinatorial space of oligonucleotide barcodes theoretically allows parallel quantification of an unrestricted number of epitopes, in practice these methods are limited by the availability of antigen-specific antibodies. Moreover, each antibody must be separately conjugated with a unique oligonucleotide (oligo)-barcode, which currently does not allow a scalable and pooled construction of barcoded antibody libraries. Recent technologies for the combined high-throughput measurement of the epigenome and proteome[9,10] also rely on oligo-barcoded antibody panels for epitope detection and thus face the outlined challenges for scaling up protein detection.

In the present study, we describe PHAGE-ATAC (Fig. 1a–c and Extended Data Fig. 1a,b), a multimodal single-cell approach for phage-based multiplex protein measurements and chromatin accessibility profiling using droplet-based single-cell (sc) ATAC-sequencing (10× Genomics scATAC[4]). PHAGE-ATAC enables sensitive quantification of accessible chromatin and proteins, captures mitochondrial (mt)DNA that can be used as a native clonal tracer[11,12], introduces phages as renewable reagents for high-throughput, single-cell epitope profiling and leverages phage libraries for the selection of antigen-specific antibodies[13,14]. Altogether, PHAGE-ATAC provides a method that will expand the scope of the single-cell profiling toolbox.

Results

Single-cell protein detection via Nb-displaying phages. Protein quantification in PHAGE-ATAC is based on epitope recognition by nanobody (Nb)[15]-displaying phages (Fig. 1a and Extended Data Fig. 1a,b), in contrast to recognition by oligonucleotide-conjugated antibodies in CITE-seq and related methods[7,8], or fluorescently labeled antibodies in other techniques[16,17]. The hypervariable complementarity-determining region 3 (CDR3) within each Nb-encoding phagemid acts as a unique genetic barcode[18] that is identified by sequencing in PHAGE-ATAC, and serves as a proxy for antigen detection and quantification (Fig. 1a and Extended Data Fig. 1a). To allow phage-based epitope quantification alongside accessible chromatin profiling, we engineered an M13 phagemid for the in-frame expression of: (1) an epitope-binding Nb; (2) a PHAGE-ATAC tag (PAC-tag) containing the Illumina Read 1 sequence (RD1); and (3) the phage coat protein p3 for surface display (Fig. 1a,b). This enables phage (p)Nb-based recognition of cell-surface antigens, simultaneous indexing of phagemids and ATAC fragments, as well as separate generation of phage-derived tag (PDT) and ATAC-seq libraries (Fig. 1c and Extended Data Fig. 1c–f; Methods).

We first validated that the PHAGE-ATAC-modified phagemid workflow allows specific pNb antigen recognition and pNb-based cell staining during scATAC cell lysis. As a proof of concept, we used HEK293T cells expressing surface-exposed, glycosylphosphatidyl-inositol (GPI)-anchored, enhanced green fluorescent protein (EGFP) (EGFP-GPI) that is specifically recognized by an anti-EGFP pNb[19] (Extended Data Fig. 2a–e). Notably, introduction of the PAC-tag did not impair Nb display and antigen recognition (Extended Data Fig. 2f,g). Moreover, fixation retained pNb-based cell staining after the lysis step (Extended Data Fig. 2h,i and see Methods).

[1]Klarman Cell Observatory, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [2]Departments of Pathology, Stanford University, Stanford, CA, USA. [3]Berlin Institute of Health at Charité—Universitätsmedizin Berlin, Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany. [4]Department of Immunobiology, Yale School of Medicine, New Haven, CT, USA. [5]Department of Pharmacology, Yale School of Medicine, New Haven, CT, USA. [6]Department of Molecular Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [7]Center for Computational and Integrative Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [8]Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [9]Present address: Genentech, South San Francisco, CA, USA. ✉e-mail: efiskin@broadinstitute.org; aviv.regev.sc@gmail.com
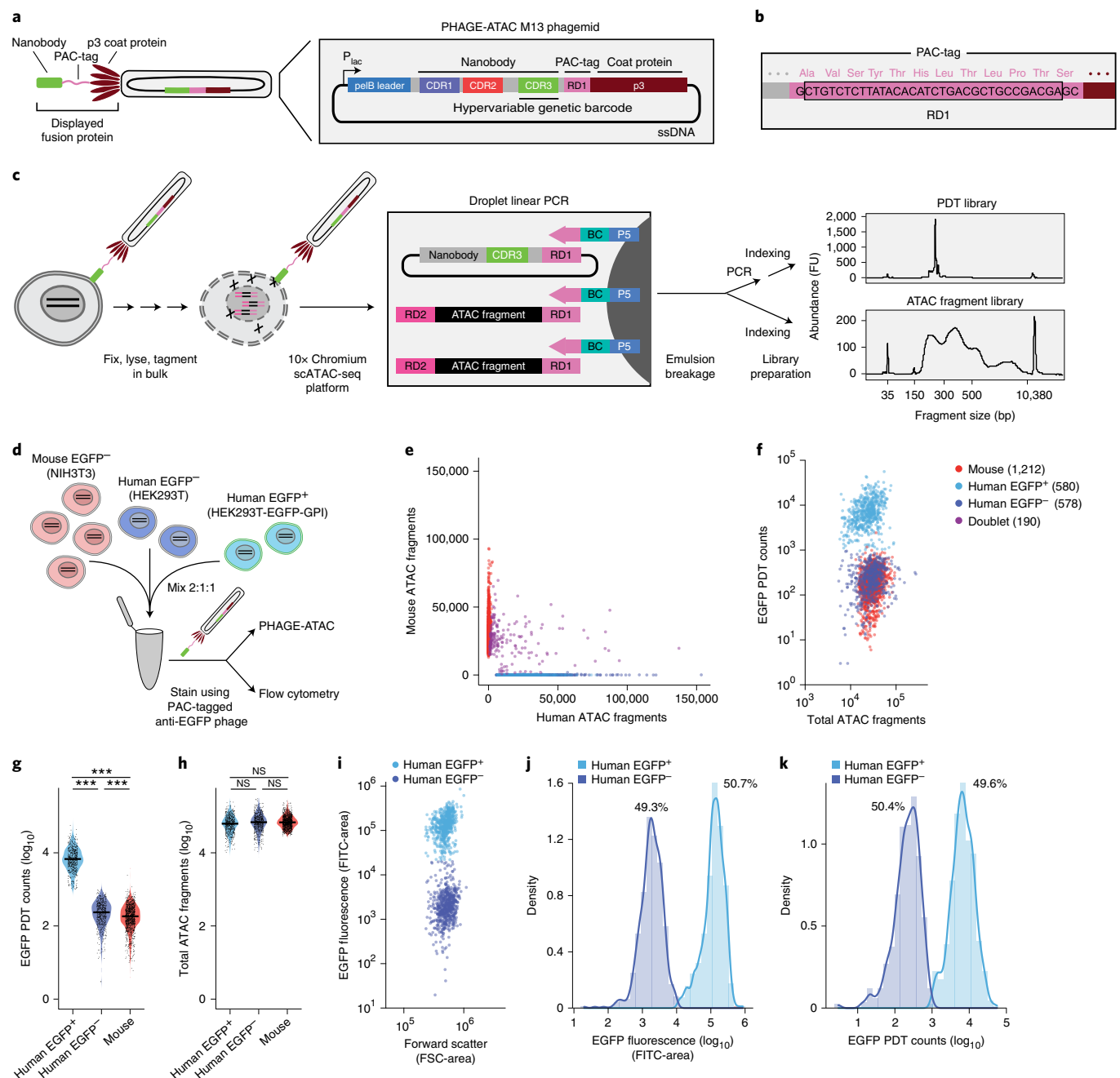
Fig. 1 | PHAGE-ATAC for massively parallel concomitant detection of protein epitopes and chromatin accessibility. a–c, PHAGE-ATAC overview.
a, Schematic of engineered Nb-displaying M13 phages used for PHAGE-ATAC. Nbs are displayed via fusion to the p3 coat protein; the PAC tag is placed in the linker between the Nb and p3. M13 phagemids contain a pelB leader for periplasmic secretion and incorporation of fusions during phage assembly.
b, PAC-tag RD1 sequence (pink) allows capture by 10× ATAC gel bead oligos (Extended Data Fig. 1c), without disrupting the open reading frame.
c, PHAGE-ATAC workflow. After phage Nb staining, fixation, lysis and tagmentation in bulk (leftmost), single cells and 10× ATAC gel beads are encapsulated into droplets using 10× Genomics microfluidics, followed by linear amplification with simultaneous droplet barcoding of chromatin fragments and phagemids via hybridization of 10× barcoding primers to RD1 sequences (middle). Separate PDT and ATAC-seq libraries are prepared (right and Extended Data Fig. 1f). Right: representative BioAnalyzer traces. BC, bead barcode; FU, fluorescence units. **d–k**, Single-cell ATAC-seq and EGFP specificity in a species-mixing experiment. **d**, Experimental scheme. **e**, Number of human (x axis) and mouse (y axis) ATAC fragments associated with each bead barcode (dots), colored by assignment as human EGFP$^+$ (light blue), human EGFP$^-$ (dark blue), mouse (red), doublet (purple, >10% human and mouse fragments). **f**, EGFP PDT counts (y axis, $\log_{10}$ scale) and number of ATAC fragments (x axis, $\log_{10}$ scale) for each bead barcode (dots) colored as in **e** (see color legend). **g,h**, Distributions of EGFP PDTs (**g**, y axis) and ATAC fragments (**h**, y axis) in each of the three populations (x axis, human EGFP$^+$ cells, n = 580; human EGFP$^-$ cells, n = 578; mouse cells, n = 1,212 from one experiment; one-tailed Mann–Whitney U-test, ***$P < 10^{-4}$). NS, not significant. In **g** human EGFP$^+$ versus human EGFP$^-$, $P = 5.039 \times 10^{-191}$; human EGFP$^-$ versus mouse, $P = 2.946 \times 10^{-10}$; human EGFP$^+$ versus mouse, $P = 1.101 \times 10^{-257}$. The line is the median. **i–k**, PDT quantification consistent with flow cytometry. EGFP fluorescence (**i**, y axis) and distribution (**j**, x axis) and distribution of EGFP PDT (**k**, x axis) in EGFP$^+$ (light blue) and EGFP$^-$ (dark blue) human cells.
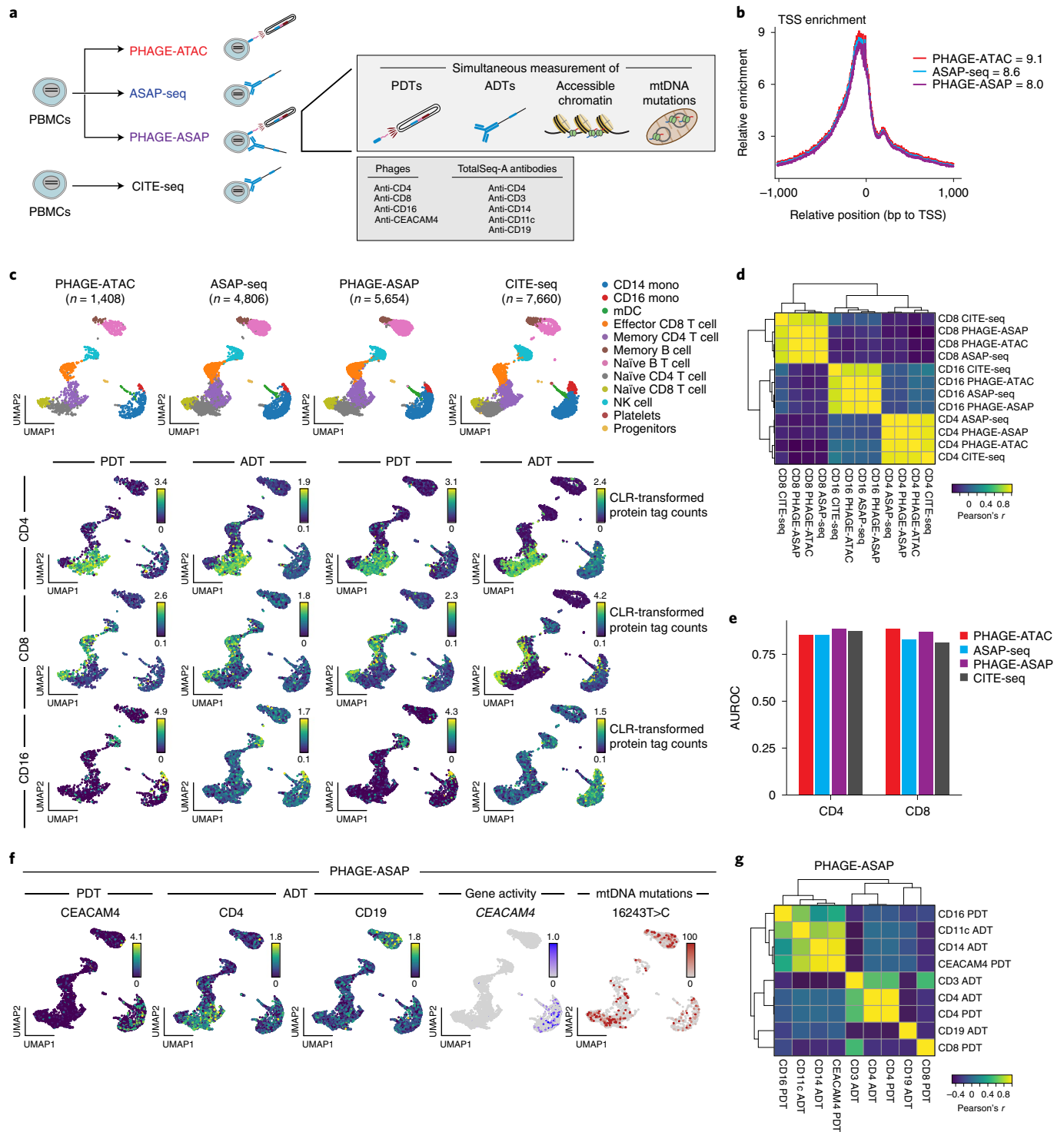
Fig. 2 | Benchmarking of PHAGE-ATAC and PHAGE-ASAP for epitope, accessible chromatin profiling and mtDNA genotype capture versus ASAP-seq and CITE-seq. a, Comparison of multimodal workflows. **b**, Comparable TSS enrichment across assays. Mean relative enrichment (*y* axis) of ATAC fragments at positions relative to TSSs (*x* axis) for each method. **c**, Comparable cell type and protein detection across assays. The 2D joint embedding of scRNA-seq profiles are from published CITE-seq[8] and scATAC-seq profiles from PHAGE-ATAC, ASAP-seq and PHAGE-ASAP, colored by annotated cell types (top) or the level of protein marker ADTs or PDTs (bottom). **d**, Agreement between protein level estimates (PDT or ADT) from different assays. Pearson's *r* (color) is given for each pair of indicated ADTs or PDTs (columns and rows). **e**, Comparable cell-type classification accuracy across assays. Classification accuracy (AUROC, *y* axis) for CD4+ (left) or CD8+ (right) T cells is based on measured CD4 and CD8 protein tag levels (ADT or PDT) in each assay. **f**, Diverse data profiles detected by PHAGE-ASAP. The 2D embedding that showed PHAGE-ASAP scATAC-seq profiles (as in **c**), colored by the level of protein marker ADTs or PDTs (left), normalized gene activity score (middle) or mtDNA mutation frequency (right). **g**, Agreement between protein level estimates (PDT or ADT) from PHAGE-ASAP for each marker across all cell states. Pearson's *r* (color) is given for each pair of indicated ADTs or PDTs (columns and rows).
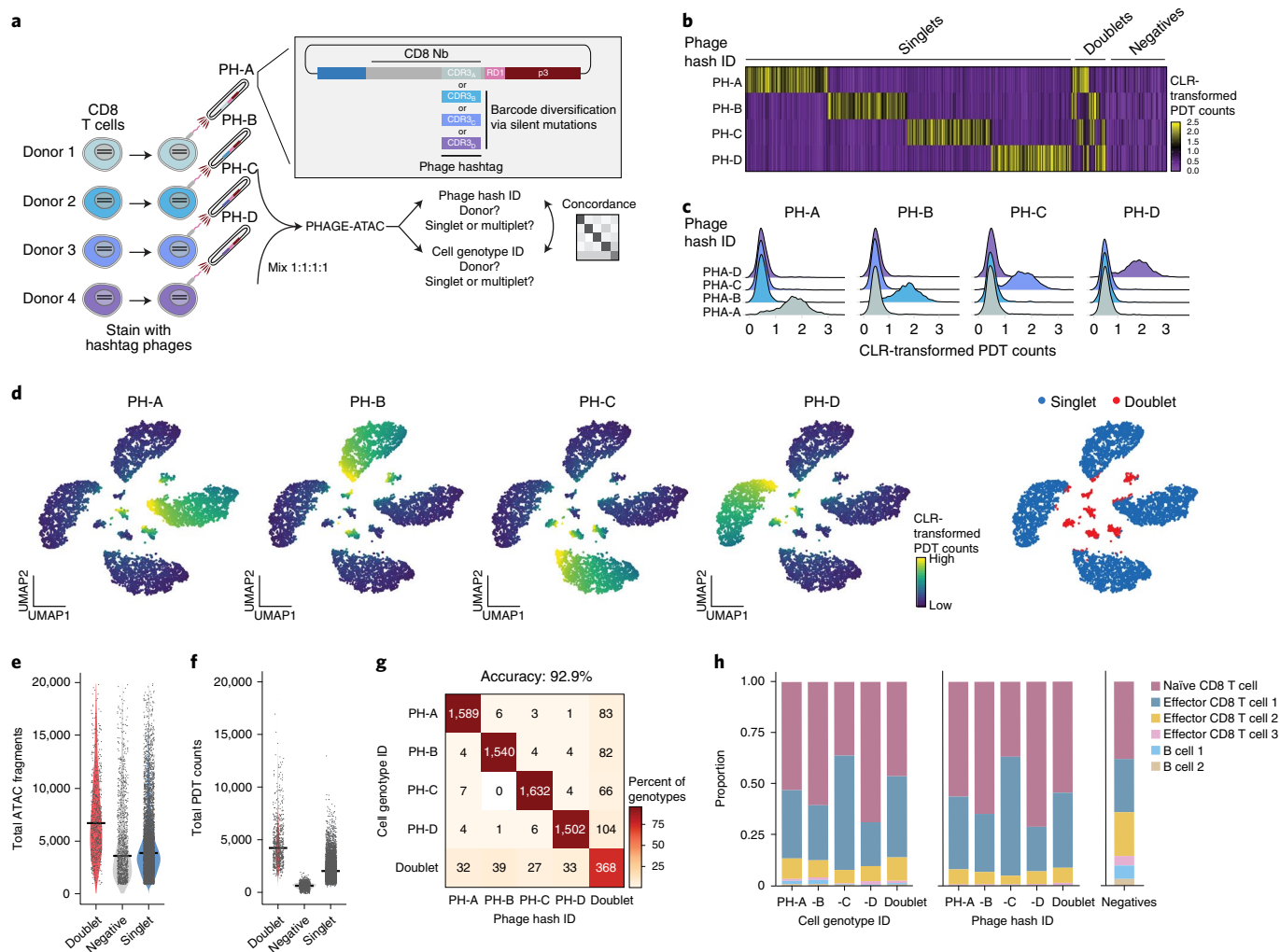
Fig. 3 | Phage nanobodies harboring silent mutations enable sample multiplexing. a, Generation of phage hashtags by silent mutations. Schematic shows four anti-CD8 phage hashtags and a subsequent hashing experiment using CD8 T cells from four human donors. **b–h**, Effective demultiplexing of phage hashtags. **b**, PDT counts (color bar, CLR) for each hashtag (rows) across cells (columns) sorted by their HTODemux classification (phage hash ID). **c**, PDT count distributions for each hashtag (colored histograms) across the four phage hash IDs (Wilcoxon's two-tailed test, ***$P < 10^{-4}$). **d**, The 2D embedding of cell barcodes by PDT count data, colored by PDT count for the marked hashtag (four left panels) or by singlet/doublet classification (right). **e,f**, Distribution of the number of ATAC fragments per barcode (**e**, y axis) or PDT counts (**f**, y axis) in cell barcodes in each category (x axis) (two-tailed Mann–Whitney U-test, ***$P < 10^{-4}$). NS, not significant. The line is the median. **g**, Number and percentage (color) of barcodes shared between each genotype-based (genotype ID, rows) and phage hashtag ID-based (columns) assignments. Top: overall accuracy. **h**, Proportion of cells of each type (y axis) within each assigned barcode category (x axis) based on genotype (left) and/or hashtags (right), and in the negative fraction (far right).

Phage-based single-cell profiling of proteins and accessible chromatin. To benchmark PHAGE-ATAC for single-cell profiling, we performed a 'species-mixing' experiment, in which we pooled mouse (NIH3T3), human EGFP⁻ (HEK293T) and human EGFP⁺ (HEK293T-EGFP-GPI) cells at a 2:1:1 ratio, followed by anti-EGFP pNb staining, library preparation and analysis using a customized computational workflow (Fig. 1d and Extended Data Fig. 3a; Methods). After filtering, we recovered 1,212 mouse and 1,158 human cell barcodes (Fig. 1e), with good quality based on a high mean transcription start site (TSS) enrichment score (8.3), fraction of fragments in peaks (54.7%), DNase-hypersensitive sites (64.6%) and TSSs (36.8%) (Extended Data Fig. 3b–d and Supplementary Table 6), all comparable to gold-standard, published reference data without additional protein detection[4,11]. Analysis of EGFP PDT counts confirmed the presence of EGFP⁺ and EGFP⁻ cells (Fig. 1f,g) which, together with mouse cell barcodes, were recovered at expected input ratios (observed 2.09:1:1, expected 2:1:1), with no

meaningful differences in scATAC-seq data quality metrics (Fig. 1h and Extended Data Fig. 3b–d). EGFP PDT levels by PHAGE-ATAC (Fig. 1f,g) and EGFP fluorescence intensities by standard flow cytometry (Fig. 1i) were highly concordant (flow cytometry EGFP⁻ 49.3% versus PHAGE-ATAC EGFP⁻ 50.4% and flow cytometry EGFP⁺ 50.7% versus PHAGE-ATAC EGFP⁺ 49.6%) (Fig. 1j,k).

As PHAGE-ATAC enables access to the cell interior before droplet encapsulation and barcoding (Methods), we reasoned that pNb staining after cell lysis could allow single-cell detection of intracellular antigens. To measure both intracellular epitopes and chromatin accessibility profiles in single cells using PHAGE-ATAC (Methods), we analyzed HEK293T cells expressing cytosolic EGFP, an antigen that is inaccessible to pNb staining in intact cells (Extended Data Fig. 2d). Although phage binding to EGFP was not detectable in intact cells, there was robust anti-EGFP pNb staining postfixation and lysis (Extended Data Fig. 3e) and EGFP PDTs showed faithful bimodal cytosolic EGFP fluorescence at the expected ratio

(Extended Data Fig. 3f). Thus, PDTs can accurately and sensitively quantify either extracellular or intracellular epitopes.

PHAGE-ATAC for protein, chromatin and mtDNA genotype quantification. To further expand the number of epitopes detectable by PHAGE-ATAC, we introduced PHAGE-ATAC with select antigen profiling (PHAGE-ASAP), a combined workflow for costaining with pNbs and oligo-labeled antibodies (commercial TotalSeq antibodies) (Fig. 2a; Methods), and further optimized the PHAGE-ATAC lysis buffer to better preserve phage staining[11] (Extended Data Fig. 4; Methods).

Next, we showed that PHAGE-ATAC (and PHAGE-ASAP) can simultaneously measure accessible chromatin, surface epitopes and mitochondrial genotypes[11,12] in peripheral blood mononuclear cells (PBMCs), at a quality comparable to the oligo-barcoded antibody approaches CITE-seq and ASAP-seq[9] (Fig. 2a). We used four pNbs: three previously reported high-affinity Nb sequences against CD4, CD8 and CD16 and a selected anti-CEACAM4 Nb (Extended Data Fig. 5a–d and Methods). For PHAGE-ASAP, we also included five TotalSeq antibodies (to CD4, CD3, CD14, CD11c and CD19). Flow cytometry of pNb-stained PBMCs and side-by-side comparison between pNbs and conventional antibody-stained cells confirmed the antigen specificity of the produced phages (Extended Data Fig. 5e). PHAGE-ATAC, ASAP-seq and PHAGE-ASAP had comparable data quality (Fig. 2b, Extended Data Fig. 6a–c and Supplementary Table 6). Integrative canonical correlation analysis[20], clustering and dimensionality reduction of PHAGE-ATAC, ASAP-seq, PHAGE-ASAP and published CITE-seq data[8] identified the same set of expected cell states (Fig. 2c; Methods). PDT (from PHAGE-ATAC and PHAGE-ASAP) and antibody-derived tag (ADT, from ASAP-seq and CITE-seq) values for each cell-surface marker were highly correlated (Pearson's $r = 0.89–0.99$; Fig. 2d), and CD4 and CD8 PDT distribution across cells reflected the expected cell populations (Extended Data Fig. 6d,e). Differential gene activity scores from CD4 and CD8 PDT abundances (Extended Data Fig. 6f) identified *CD4* and *CD8* loci as top hits and recovered many known bona fide markers of CD4[+] and CD8[+] T cells (for example, CD4: *CD40LG*, *ANKRD55*, *TSHZ2*; CD8: *PRF1*, *EOMES*, *RUNX3*; Extended Data Fig. 6f). Finally, CD4[+] and CD8[+] T cells were classified correctly by CD4 and CD8 PDT levels alone (area under the receiver operating characteristics (AUROC) = 0.85–0.89) (Fig. 2e and Extended Data Fig. 6g; Methods). PHAGE-ASAP successfully captured PDTs, ADTs, accessible chromatin and mitochondrial genotypes[11] (median fraction of mtDNA-derived fragments = 22.1–23.0%) (Fig. 2f and Extended Data Fig. 6h–j; Methods), with high concordance between anti-CD4 pNb and anti-CD4 TotalSeq antibody signal across cell states (Pearson's

$r = 0.99$; Fig. 2g). Thus, PHAGE-ATAC reliably and specifically detects cell-surface proteins, epigenomic profiles and mtDNA genotypes in single cells.

Phage Nbs enable cell hashing and sample multiplexing. To scale PHAGE-ATAC, we introduced an alternative for sample multiplexing in scATAC-seq using pNbs for cell hashing. Several methods allow 'overloading' antibody-tagged cells into droplets to increase throughput and mitigate batch effects[2,21–23]. To demonstrate hashtags for PHAGE-ATAC, we generated four anti-CD8 hashtag pNbs (henceforth referred to as hashtags) by introducing distinct silent mutations into the anti-CD8 CDR3 (Fig. 3a; Methods), allowing sequencing-based identification of each hashtag. As expected, the hashtags displayed comparable CD8 recognition within PBMCs (Extended Data Fig. 7a). To demonstrate phage-based hashing, we stained CD8 T cells from four different healthy donors with a unique hashtag each, pooled them and processed the pool by PHAGE-ATAC overloading 20,000 cells (Fig. 3a) (versus ~6,000 cells without overloading). These yielded high-quality data for 8,366 cell barcodes to which we assigned donor and singlet/doublet status from hashtag counts (Methods), identifying the sample of origin for 6,438 singlets and 703 doublets (observed doublet rate 8.4% versus 10% expected) (Fig. 3b,c). Singlet and doublet assignments were concordant with a two-dimensional (2D) embedding of hashtag count data (Fig. 3d), with the expected higher numbers of chromatin fragments and hashtag counts in doublets ($P < 2.2 \times 10^{-16}$, Mann–Whitney *U*-test; Fig. 3e,f). The hashtag-based assignments were also highly concordant with assignments based on computationally derived donor genotypes from accessible chromatin profiles[24] (Methods), with a singlet classification accuracy of 99.3% and an overall classification accuracy of 92.9% (Fig. 3g). Notably, chromatin accessibility analyses revealed a small set of putative B cells (Extended Data Fig. 7b,c), consistent with the presence of a minor contaminating population after CD8 T-cell enrichment. Although B cells were classified as hashtag negative, genotype- and hashtag-based classifications were highly consistent across CD8 T-cell states (Fig. 3h and Extended Data Fig. 7d–f), further confirming hashtag antigen specificity. Mitochondrial genotyping using mgatk[11] was broadly concordant with the hashtag assignments, but showed that two donors (PH-B and PH-C) shared the same haplotype, whereas each of the other two donors carried distinguishing variants (Extended Data Fig. 7g). Collectively, these results establish the use of hashtag pNbs for sample multiplexing in scATAC-seq.

Phage display selection of pNbs for single-cell protein profiling. The production of high-quality antigen-specific antibodies is laborious, expensive and limited by animal immunization,

Fig. 4 | Phage display selection of antigen-specific phage Nbs using PANL for multiplexed PHAGE-ATAC. a–g, Phage display selection of PHAGE-ATAC Nbs. **a**, Selection using PANL (Methods). PANL is panned against HEK293T-EGFP-GPI with counterselection against antigen-devoid HEK293T. Bound phages are eluted, bacteria infected and output libraries generated. After multiple rounds, the output is Illumina sequenced and phage clones are picked, phagemids isolated and inserts sequenced. **b**, Flow cytometry analysis of selection progress. Flow cytometry plots of EGFP fluorescence (*y* axis) and phage binding (*x* axis, Alexa Fluor-647 area) to HEK293T-EGFP-GPI (EGFP[hi] and EGFP[lo]) in, from left, the input library and after consecutive selection cycles (see also Extended Data Fig. 2c and Methods). **c–e**, Illumina sequencing analysis of selection progress. **c**, Proportion of top 35 selected clones (**c**, *y* axis), top 500 and 1,000 clones (**d**, *x* axis), and Lander–Waterman estimates of library complexity (**e**, *y* axis) across phage libraries (*x* axis). Est./est., estimated. **f**, Flow cytometry screen of 94 pNbs derived from selection round 3. *Q2/Q1* signal (as defined in **b**) shown when staining HEK293T-EGFP-GPI (EGFP[hi] and EGFP[lo]) cells with individual pNbs after selection round 3. Dashed line: *Q2/Q1* = 1 threshold used for calling positive clones. **g**, Flow cytometry plots of EGFP fluorescence (*y* axis) and phage binding (*x* axis, Alexa Fluor-647 area) to HEK293T-EGFP-GPI cells (EGFP[hi] and EGFP[lo]) using an immunization-based[19] anti-EGFP pNb (middle), selected clone C5 (right) and anti-mCherry control (left). **h–m**, Multiplexed PHAGE-ATAC with a 12-plex panel including anti-SARS-CoV-2-S pNbs, anti-EGFP and PBMCs, recognizing pNbs in a mixed-cell population. **h**, Schematic overview. **i**, The 2D embedding of scATAC-seq profiles from a mixture of PBMCs, HEK293T-EGFP-GPI[+] cells and HEK293T SARS-CoV-2-S[+] cells colored by cellular components (left), annotated cell types (right) or level of measured PDTs (bottom). **j**, Mutually exclusive PDT detection. PDT counts (geometric mean) from seven anti-SARS-CoV-2-S pNbs (*y* axis) and two anti-EGFP pNbs (*x* axis) are given in each cell (dots, colored by type as in **i**). **k–m**, Agreement in protein level estimates between different pNbs to the same protein. **k**, Pearson's *r* (color) for each pair of PDTs (columns and rows) across cells. **l,m**, PDTs of pNbs (*x* and *y* axes) recognizing SARS-CoV-2-S (**l**) or EGFP (**m**) across all cells (dots, colored as in **i**). Upper left: Pearson's *r*.

generating a bottleneck for antibody-based protein profiling[25]. In contrast, recombinant antibody technology based on phage display has allowed fast selection of high-affinity binders against purified

targets produced, for example, in bacterial or insect hosts or even against antigens expressed on whole intact mammalian cells[14,26]. To enable rapid generation of antigen-specific pNbs for PHAGE-ATAC
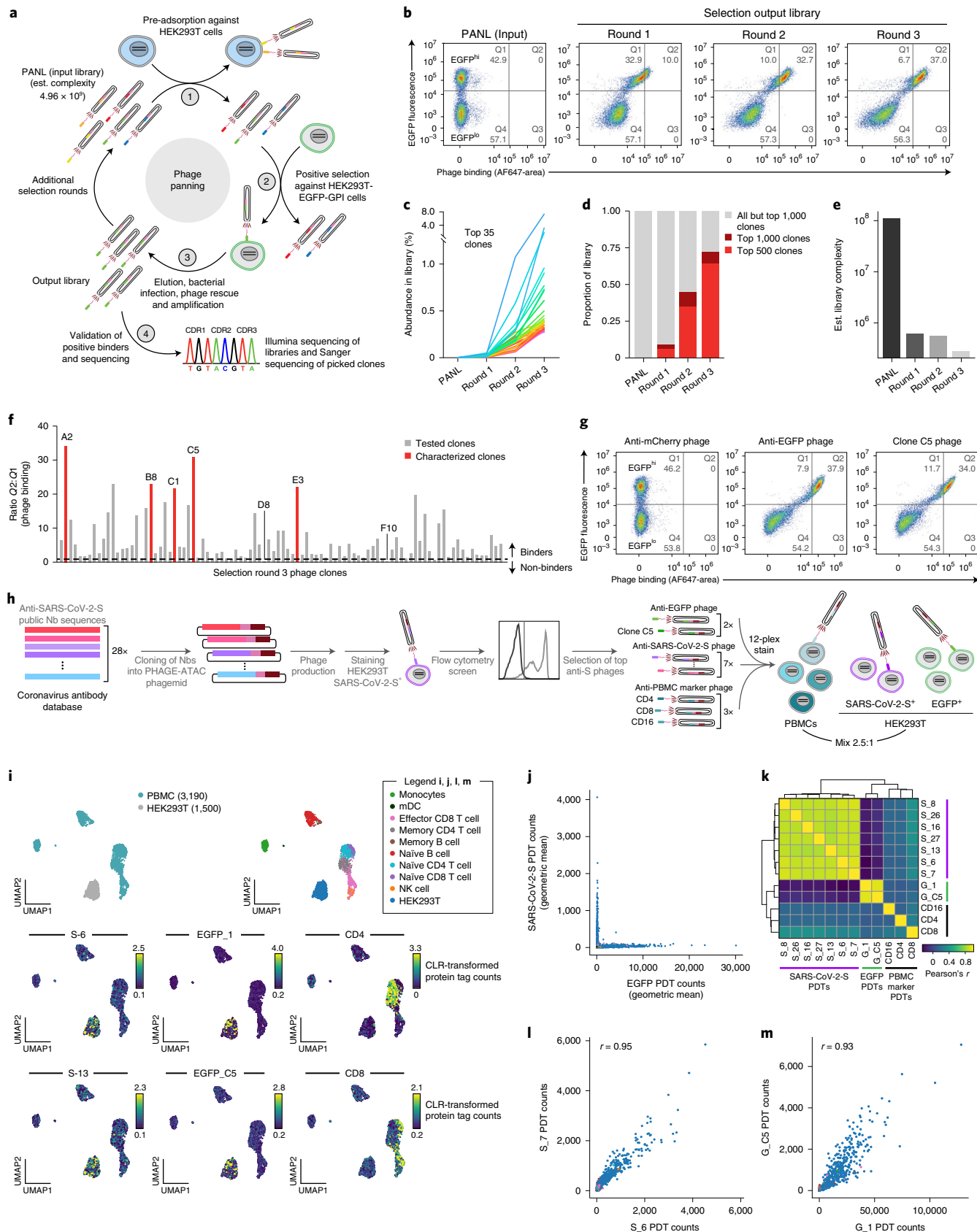
within just a few weeks, we developed a PHAGE-ATAC Nanobody Library (PANL), a synthetic high-complexity (4.96×10^9) pNb library (Extended Data Fig. 8). To demonstrate identification of pNbs using PANL, we performed a selection against EGFP-GPI-expressing HEK293T cells, while counterselecting using parental HEK293T cells (Fig. 4a). Over three selection rounds, we monitored the enrichment of pNbs by staining EGFP-GPI$^+$ cells, revealing a steady increase of antigen-recognizing pNbs with each consecutive round (Fig. 4b). In addition, we assessed the selection by sequencing both PANL and each output library (Methods), finding a relative increase of short CDR3-harboring pNbs and an enrichment of full-length pNbs after each selection round (Extended Data Fig. 9a). Output libraries also had a shifted amino acid distribution at each random and semirandom CDR position relative to the input PANL (Extended Data Fig. 9b). There was a continuous expansion of individual pNb clones, with the most abundant clones representing 0.2–8% and the top 1,000 clones constituting 72.6% of the final output library (Fig. 4c,d). As expected, the emergence of dominant clones coincided with a reduction in overall library complexity, with each selection round illustrating selection-driven convergence (Fig. 4e). Screening of 94 clones after the final (third) selection demonstrated that at least 95% of clones recognized EGFP-GPI$^+$ cells with strong binding ($Q2/Q1 > 1$; Fig. 4f and Extended Data Fig. 9c,d). As clones varied in their ability to bind EGFP-GPI$^+$ cells, we Sanger sequenced the phagemid inserts of seven clones (five strong and two weak binders), uncovering multiple clones harboring identical Nb inserts (A2 and C1, B8 and E3; Extended Data Fig. 9e). These picked clones included the second (clone A2/C1), third (C5) and fourteenth (B8/E3) most abundant clones within the final output library (Extended Data Fig. 9f), confirming the successful selection of target antigen-binding pNbs. Finally, side-by-side comparison of a selected clone (C5) and a reported high-affinity anti-EGFP Nb derived from immunized animals[19] indicated similar binding to EGFP-GPI$^+$ cells (Fig. 4g). These results demonstrate the utility of PANL for the rapid selection of pNbs to detect and quantify cellular antigens of interest, and its potential for the generation of a toolbox of barcoded affinity reagents for single-cell genomics.

Measurement of host and SARS-CoV-2 epitopes using recombinant pNbs. Finally, we demonstrated the ability to multiplex epitope detection using PHAGE-ATAC and to monitor antigens with multiple target-specific pNbs. We focused on the severe acute respiratory syndrome (SARS)-CoV-2 spike protein (SARS-CoV-2-S), the major target of natural and vaccine-induced COVID19 immunity[27]. Based on published anti-SARS-CoV-2-S Nb sequences[28] (Methods), we produced 28 pNbs and screened them against SARS-CoV-2-S-expressing HEK293T cells (Fig. 4h). Flow cytometry revealed specific staining of SARS-CoV-2-S$^+$ cells by all tested pNbs (Extended Data Fig. 10a) and we selected the top seven pNbs with the highest signal for PHAGE-ATAC. We constructed a twelve-plex panel of seven anti-SARS-CoV-2-S Nbs, two anti-EGFP Nbs (including PANL-derived clone C5) and three PBMC-recognizing pNbs (anti-CD4, CD8 and CD16), and profiled a mixture of PBMCs, SARS-CoV-2-S$^+$ HEK293T cells and EGFP-GPI$^+$ HEK293T cells by PHAGE-ATAC (Fig. 4h). The filtered dataset of 4,690 cell barcodes revealed PBMCs and HEK293T cells at the expected ratio, recovered the expected cell states and successfully detected the entire pNb panel (Fig. 4i and Extended Data Fig. 10b,c). There was mutually exclusive detection of anti-SARS-CoV-2-S, anti-EGFP and anti-PBMC PDTs (Fig. 4j and Extended Data Fig. 10d–g) and highly correlated PDT levels for all pNbs recognizing the same antigen (Pearson's $r = 0.87$–0.95; Fig. 4k–m). Thus, PHAGE-ATAC can be used for multiplex single-cell detection of host and viral antigens with both published and PANL-derived pNbs.

Discussion
PHAGE-ATAC uses the power of recombinant phage display technology as the basis for single-cell profiling of cell-surface and intracellular proteins, chromatin accessibility and mtDNA. This allows users to leverage the renewable nature and scalability of pooled phage library preparation as well as the compact size and stability of Nbs[15]. Despite differences in affinities and sizes between phage Nbs and antibodies, our extensive benchmarking of PHAGE-ATAC demonstrated its highly specific and sensitive monitoring of protein expression. As we have previously shown, the robust capture of mtDNA fragments and the detection of mitochondrial mutations can be used for clonal tracing[11,12,29] and will enable the integration of lineage information with protein expression and cell state using PHAGE-ATAC. Although measured data modalities in PHAGE-ATAC are analogous to information recorded by recently reported oligo-barcoded–antibody-based multimodal methods[9,10], the unique use of genetic barcoding and recombinant affinity reagents in PHAGE-ATAC paves the way for ultra-highly multiplexed epitope measurements using complex phage libraries. We further envisage that PHAGE-ATAC will be of particular interest for the future study of solid tissues, including tumors. We also envisage PHAGE-ATAC as an adaptive tool that may be further combined with unique molecular identifiers for phagemid counting and other engineerable scaffolds used in phage display applications (for example, scFv, Fab)[30]. In the future, the use of large panels of recombinant affinity reagents could substantially enhance multimodal, single-cell characterization of the proteome, epigenome and other readouts at high throughput and specificity.

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/s41587-021-01065-5.

References
1. Klein, A. M. et al. Droplet barcoding for single–cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
2. Lareau, C. A. et al. Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat. Biotechnol.* **37**, 916–924 (2019).
3. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
4. 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).
5. Ma, A., McDermaid, A., Xu, J., Chang, Y. & Ma, Q. Integrative methods and practical challenges for single-cell multi-omics. *Trends Biotechnol.* **38**, 1007–1022 (2020).
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. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
9. Mimitou, E. P. et al. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat. Biotechnol.* **39**, 1246–1258 (2021).
10. Swanson, E. et al. Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. *Elife* **10**, (2021).
11. Lareau, C. A. et al. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat. Biotechnol.* **39**, 451–461 (2021).
12. Ludwig, L. S. et al. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325–1339 e1322 (2019).

13. Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315–1317 (1985).

14. Hoogenboom, H. R. Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.* **23**, 1105–1116 (2005).

15. Ingram, J. R., Schmidt, F. I. & Ploegh, H. L. Exploiting nanobodies' singular traits. *Annu. Rev. Immunol.* **36**, 695–715 (2018).

16. Katzenelenbogen, Y. et al. Coupled scRNA-seq and intracellular protein activity reveal an immunosuppressive role of TREM2 in cancer. *Cell* **182**, 872–885 e819 (2020).

17. Paul, F. et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell* **163**, 1663–1677 (2015).

18. Pollock, S. B. et al. Highly multiplexed and quantitative cell-surface protein profiling using genetically barcoded antibodies. *Proc. Natl Acad. Sci. USA* **115**, 2836–2841 (2018).

19. Rothbauer, U. et al. Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat. Methods* **3**, 887–889 (2006).

20. 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).

21. 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).

22. McGinnis, C. S. et al. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).

23. Stoeckius, M. et al. Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).

24. Heaton, H. et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620 (2020).

25. Bradbury, A. R., Sidhu, S., Dubel, S. & McCafferty, J. Beyond natural antibodies: the power of in vitro display technologies. *Nat. Biotechnol.* **29**, 245–254 (2011).

26. Miersch, S. & Sidhu, S. S. Synthetic antibodies: concepts, potential and practical considerations. *Methods* **57**, 486–498 (2012).

27. Dai, L. & Gao, G. F. Viral targets for vaccines against COVID-19. *Nat. Rev. Immunol.* **21**, 73–82 (2021).

28. Raybould, M. I. J., Kovaltsuk, A., Marks, C. & Deane, C. M. CoV-AbDab: the coronavirus antibody database. *Bioinformatics* **37**, 734–735 (2021).

29. Walker, M. A. et al. Purifying selection against pathogenic mitochondrial DNA in human T cells. *N. Engl. J. Med.* **383**, 1556–1563 (2020).

30. Gebauer, M. & Skerra, A. Engineered protein scaffolds as next-generation antibody therapeutics. *Curr. Opin. Chem. Biol.* **13**, 245–255 (2009).

Methods

Oligonucleotides. Oligonucleotide sequences are listed in Supplementary Table 1. Oligonucleotides were ordered from Integrated DNA Technologies (IDT) unless indicated otherwise.

Cloning of phagemids for display of PAC-tagged Nb–p3 fusions for PHAGE-ATAC. Based on the 10× scATAC bead oligo design (Extended Data Fig. 1c), we hypothesized that introduction of an RD1 flanking the Nb CDR3 barcode would enable barcode capture alongside accessible chromatin fragments during droplet-based indexing. To avoid premature termination of Nb–p3 fusion translation due to the introduction of RD1, we modified the RD1-spanning reading frame, which resulted in the expression of a 12-amino acid PHAGE-ATAC tag (PAC-tag). To generate a phagemid for C-terminal fusion of both PAC-tag and p3, 20 ng of pDXinit (Addgene, catalog no. 110101) was subjected to site-directed mutagenesis with primers EF77 and EF78 using PfuUltraII (Agilent) in 50-μl reactions. PCR conditions were: 95 °C for 3 min; 19 cycles at 95 °C for 30 s, 60 °C for 1 min and 68° for 12 min; and a final extension at 72 °C for 14 min. Template DNA was digested for 1.5 h at 37 °C by addition of 1.5 μl of DpnI (Fastdigest, Thermo Fisher Scientific). PCR reactions were then purified using GeneJet Gel Extraction Kit (Thermo Fisher Scientific) and eluted in 45 μl of water, after which 20 μl of eluate was transformed into chemically competent *Escherichia coli* (NEB Stable Competent) and plated on lysogeny broth containing ampicillin (LB-Amp), yielding pDXinit-PAC. For cloning of Nb-PAC–p3, fusion-encoding phagemids, Nb sequences (Supplementary Table 3) were ordered as gBlocks from IDT. Then, 25-ng Nb gBlocks were first amplified by PCR to introduce SapI restriction sites. Primers EF87 and EF88 were used for CD4 Nb, primers EF87 and EF89 for CD16 Nb, primers EF104 and EF105 for CD8 Nb, primers EF299 and EF300 for CEACAM4 Nb, and primers EF176-EF213 for all 28 SARS-CoV-2-S-recognizing Nbs. The 50-μl PCR reactions using Q5 (New England Biolabs (NEB)) were cycled: at 98 °C for 1 min; 35 cycles at 98 °C for 15 s, 60 °C for 30 s and 72 °C for 30 s; and a final extension of 72 °C for 3 min. PCR reactions were loaded on a 1% agarose gel, expected bands were cut and PCR products were extracted using a GeneJet Gel Extraction Kit (Thermo Fisher Scientific) and eluted in 40 μl of water. Cloning was performed using the FX system as described previously[31]. Briefly, each eluted insert was mixed with 50 ng of pDXinit-PAC in a molar ratio of 1:5 (vector:insert) in 10-μl reactions and digested with 0.5 μl of SapI (NEB) for 1 h at 37 °C. Reactions were incubated for 20 min at 65 °C to heat inactivate SapI and cooled down to room temperature, and constructs were ligated by addition of 1.1 μl of 10× T4 ligase buffer (NEB) and 0.25 μl of T4 ligase (NEB) and incubated for 1 h at 25 °C. Ligation was stopped by heat inactivation for 20 min at 65 °C, followed by cooling to room temperature. Ligation reactions, 2 μl, were transformed into chemically competent *E. coli* (NEB Stable Competent) and plated on 5% sucrose-containing LB-Amp, yielding pDXinit-CD4Nb-PAC, pDXinit-CD8Nb-PAC, pDXinit-CD16Nb-PAC, pDXinit-CEACAM4Nb-PAC and all 28 pDXinit-SARS2-SNb-PAC constructs. For cloning of CD8 hashtag phagemids, 20 ng of pDXinit-CD8Nb-PAC was used as a template for site-directed mutagenesis (as described earlier), using primers EF156 and EF157 to generate pDXinit-CD8Nb(PH-A)-PAC, primers EF158 and EF159 for pDXinit-CD8Nb(PH-B)-PAC, primers EF164 and EF165 for pDXinit-CD8Nb(PH-C)-PAC, and primers EF166 and EF167 for pDXinit-CD8Nb(PH-D)-PAC. For cloning of EGFP Nb-displaying phagemids, the EGFP Nb sequence from pOPINE GFP nanobody (Addgene, catalog no. 49172) was amplified in 50-μl PCR reactions with Q5 (NEB) using 25 ng of the plasmid template and EF05 and EF06 primers. The EGFP Nb insert was cloned into pDXinit using FX cloning (described earlier), yielding pDXinit-EGFPNb. EGFP Nb-displaying phagemids containing RD1 in different orientations were cloned by using pDXinit-EGFPNb and performing site-directed mutagenesis (described earlier) with EF73 and EF74 to obtain pDXinit-EGFPNb-PAC or using EF75 and EF76 to yield pDXinit-EGFPNb-RD1(5-3). For introduction of a PCR handle required for PDT library amplification, pDXinit-EGFPNb-PAC was subjected to site-directed mutagenesis (as described earlier) using primers EF78 and EF79, yielding pDXinit-EGFPNb(handle)-PAC. For cloning of mCherry Nb-displaying phagemids, the mCherry Nb sequence from pGex6P1 mCherry Nb (Addgene, catalog no. 70696) was amplified in 50-μl PCR reactions with Q5 (NEB) using 25 ng of the plasmid template and EF07 and EF08 primers. The mCherry Nb insert was cloned into pDXinit using FX cloning (as described earlier), yielding pDXinit-mCherryNb. All constructs are listed in Supplementary Table 2.

Analysis of RD1-mediated phagemid amplification using RD1-containing primers. Of pDXinit-EGFPNb, pDXinit-EGFPNb-PAC or pDXinit-EGFPNb-RD1(5–3), 5 ng was subjected to linear PCR (10-μl reaction volume) using primer EF170 and 5 μl of 2× KAPA HiFi HotStart ReadyMix (Roche) and cycling conditions of: 98 °C for 2 min; 12 cycles at 98 °C for 10 s, 59 °C for 30 s and 72 °C for 1 min; and a final extension of 72 °C for 5 min. After completion, 0.625 μl of each primer EF147 and EF57, 1.25 μl of water and 12.5 μl of 2× KAPA were added. Nb-specific PCR was performed using: 98 °C for 3 min; 30 cycles at 98 °C for 15 s, 65 °C for 20 s and 72 °C for 1 min; and a final extension of 72 °C for 5 min. PCR using primers EF57 and EF58 and indicated plasmid templates was used as an amplification control.

Phage production. Phagemid-containing SS320 (Lucigen) cultures were incubated overnight in 2YT/2%/A/T at 37 °C and 240 r.p.m. Cultures were diluted 1:50 in 2YT/2%/A/T and grown for 2–3 h at 37 °C and 240 r.p.m. until the optical density at 600 nm (OD_{600}) = 0.4–0.5. Bacteria, 5 ml, were then infected with 200 μl of M13K07 helper phage (NEB) and incubated for 60 min at 37 °C. Bacteria were collected by centrifugation and resuspended in 50 ml of 2YT containing 50 μg ml^{-1} of ampicillin and 25 μg ml^{-1} of kanamycin (2YT/A/K). Phages were produced overnight by incubation at 37 °C and 240 r.p.m. Cultures were centrifuged and phages were precipitated from supernatants by addition of a quarter volume of 20% poly(ethylene glycol)-6000/2.5 M NaCl solution and incubation on ice for 75 min. Phages were collected by centrifugation (17 min, 12,500 g, 4 °C). Phage pellets were resuspended in 1.2 ml of phosphate-buffered saline (PBS), suspensions were cleared (5 min, 12,500 g, 25 °C) and supernatants containing phages were stored.

Cell culture. NIH3T3 (American Type Culture Collection (ATCC), catalog no. CRL-1658) and HEK293T cells (ATCC, catalog no. CRL-3216) were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum (FBS), 2 mM L-glutamine and 100 U ml^{-1} of penicillin–streptomycin (Thermo Fisher Scientific), and cultured at 37 °C and 5% CO_2. For subculturing, medium was aspirated, cells were washed with PBS and detached with trypsin–ethylenediaminetetraacetic acid (EDTA) 0.25% (Thermo Fisher Scientific). Detachment reactions were stopped with culture medium and cells were seeded at the desired densities. Cell stocks were prepared by resuspending cell aliquots in FBS with 10% dimethyl sulfoxide and freezing them slowly at −80 °C. Frozen aliquots were then moved to liquid nitrogen for long-term storage. All cell lines were regularly tested for *Mycoplasma* contamination.

Plasmid transfection of HEK293T cells. A day before transfection, $2 × 10^6$ HEK293T cells were seeded in 10-cm dishes (Corning) in complete culture medium (as described in Cell culture). Transfection was performed using GeneJuice reagent (Thermo Fisher Scientific). Then, 600 μl of Opti-MEM and 12 μl of GeneJuice were mixed in 1.5-ml tubes, vortexed shortly and spun down. Plasmid DNA, 4 μg (pCAG (Addgene, catalog no. 11160), pCAC-EGFP (Addgene, catalog no. 89684), pCAC-EGFP-GPI (Addgene, catalog no. 32601) or pHDM-SARS2 Spike-delta21 (Addgene, catalog no. 155130)), was added, and tubes were vortexed shortly and spun down. The transfection mix was added dropwise to HEK293T cells. Cells were grown for 24 h at 37 °C and 5% CO_2 to allow transgene expression. Successful transfection was assessed by fluorescence microscopy on an EVOS M5000.

Flow cytometry for detection of phage binding. Harvested antigen-expressing cell lines or thawed PBMCs (for the harvest and thawing protocol, see PHAGE-ATAC workflow) were resuspended in cold flow cytometry buffer (FC buffer: PBS containing 2% FBS) and incubated with respective pNbs for 20 min on a rotator at 4 °C. Cells were centrifuged and washed with cold FC buffer twice to remove unbound phages (all centrifugation steps were 350 g, 4 min, 4 °C). For optimization of fixation and lysis conditions, cells were fixed using either 0.1% or 1% formaldehyde (Thermo Fisher Scientific) and permeabilized with lysis buffers containing varying concentrations of NP-40, digitonin or Tween-20. Cells were resuspended in FC buffer and anti-M13 antibody (Sino Biological, catalog no. 11973-MM05T-50) was added at 1:500 dilution. After 10 min on ice, cells were washed twice in FC buffer and anti-mouse Fc Alexa Fluor-647-conjugated secondary antibody (Thermo Fisher Scientific, catalog no. A-21236) was added at 1:500 dilution. Cells were incubated for 10 min on ice, washed twice in FC buffer and resuspended in Sytox Blue (Thermo Fisher Scientific) containing FC buffer for live/dead discrimination according to the manufacturer's instructions. In indicated cases, cells were stained with anti-CD4-FITC (clone OKT4, BioLegend) at 1:500 dilution; thereby no anti-M13 and anti-mouse Fc antibodies were used. Stained cells were analyzed using a CytoFLEX LX Flow Cytometer (Beckman Coulter) at the Broad Institute Flow Cytometry Facility. Flow cytometry data were analyzed using FlowJo software v.10.6.1.

CEACAM4 Nb selection and validation. CEACAM4 Nbs were selected by biopanning with phage display using a previously described Nb library[32]. Selected Nbs were expressed as Fc-fusion proteins and assessed for binding to recombinant CEACAM4 (Enquire Bio, catalog no. QP5812-ec) by ELISA. Briefly, 96-well MaxiSorp plates (Thermo Fisher Scientific, catalog no. 442404) were coated with 50 μl per well of recombinant CEACAM4 protein or bovine serum albumin (BSA; Thermo Fisher Scientific, catalog no. BP1600100) at 5 μg ml^{-1} in PBS and incubated overnight at 4 °C. After coating, plates were washed four times with buffer PT (PBS with 0.05% Tween 20), 200 μl of blocking solution (PBS with 1% casein) was added, plates were incubated for 1 h at room temperature and then washed again four times. Nbs were first diluted to 0.5 μM and then serially diluted by half logs in blocking solution. Diluted Nb, 50 μl, was added for 1 h at room temperature. Plates were washed four times and 50 μl of horseradish peroxidase anti-human immunoglobulin G antibody (BioLegend, catalog no. 410603, 1:5,000) diluted in blocking solution was added to each well. After 30 min of incubation at room temperature, plates were washed six times with PT and once with PBS. Plates were developed with 100 μl of TMB Substrate Reagent Set (BD Biosciences, catalog no.

555214) and the reaction was stopped after 5 min by the addition of 100 µl of 1 M sulfuric acid. Plates were then read at wavelengths of 450 nm and 570 nm.

PHAGE-ATAC workflow. For the cell line 'species-mixing' experiment, culture medium was aspirated, cell lines were washed with PBS, harvested using trypsin–EDTA 0.25% (Thermo Fisher Scientific), resuspended in DMEM containing 10% FBS, centrifuged, washed with PBS and resuspended in FC buffer (above). For PBMC and CD8 T-cell experiments, cryopreserved PBMCs or CD8 T cells (AllCells) were thawed, washed in PBS and resuspended in cold FC buffer. All centrifugation steps were carried out at 350g for 4 min and 4 °C unless stated otherwise.

Cells were incubated with phages on a rotating wheel for 20 min at 4 °C. After three washes in FC buffer, cells were fixed in PBS containing 1% formaldehyde (Thermo Scientific) for 10 min at room temperature. Fixation was quenched by addition of 2.5 M glycine to a final concentration of 0.125 M. Cells were washed twice in FC buffer and permeabilized using lysis buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40, 1% BSA) for 3 min on ice. This buffer was used because we found that standard 10× Genomics scATAC lysis buffer results in loss of pNb cell staining (Extended Data Fig. 4). After lysis, cells were washed by the addition of 1 ml of cold wash buffer (lysis buffer without NP-40), inverted and centrifuged (5 min, 500g, 4 °C). Supernatant was aspirated and the cell pellet was resuspended in 1× Nuclei Dilution Buffer (10× Genomics). Cell aliquots were mixed with Trypan Blue and counting was performed using a Countess II FL Automated Cell Counter. Processing of cells for tagmentation, loading of 10× Genomics chips and droplet encapsulation via the 10× Genomics Chromium controller microfluidics instrument were performed according to the Chromium Single Cell ATAC Solution protocol.

For PHAGE-ATAC detection of intracellular EGFP, harvested cells were resuspended in cold FC buffer and immediately fixed in PBS containing 1% formaldehyde (Thermo Fisher Scientific) for 10 min at room temperature. Fixation was quenched by the addition of 2.5 M glycine to a final concentration of 0.125 M. Cells were washed twice in PBS and permeabilized using lysis buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40, 1% BSA) for 5 min on ice. After lysis, cells were washed by the addition of 1 ml of cold wash buffer (lysis buffer without NP-40), inverted and centrifuged (5 min, 500g, 4 °C). Cells were resuspended in FC buffer and incubated with anti-EGFP phage on a rotating wheel for 20 min at 4 °C. After three washes in FC buffer, the supernatant was aspirated and the cell pellet was resuspended in 1× Nuclei Dilution Buffer (10× Genomics). Downstream processing of cells for PHAGE-ATAC was as described above.

For species mixing, a single 10× channel was 'super-loaded' with 20,000 cells. Linear amplification and droplet-based indexing were performed as described in the 10× ATAC protocol on a C1000 Touch Thermal cycler with 96-Deep Well Reaction Module (BioRad). After linear PCR, droplet emulsions were broken, barcoded products were purified using MyONE silane bead cleanup and eluted in 40 µl of elution buffer I (the Chromium Single Cell ATAC Solution protocol). At this point eluates were split for PDT and ATAC library preparation. Whereas 5 µl of eluate was used for PDT library preparation as described below, the remaining 35 µl of eluate was used for scATAC library generation (according to the Chromium Single Cell ATAC Solution protocol). Splitting samples at this point is not expected to result in a loss of library complexity because PDTs and ATAC fragments already underwent amplification via linear PCR.

The aliquot for PDT library preparation was used for PDT-specific PCR in a 100-µl reaction using 2× KAPA polymerase and primers EF147 and EF91; cycling conditions were: 95 °C for 3 min, 20 cycles at 95 °C for 20 s, 60 °C for 30 s and 72 °C for 20 s; and a final extension of 72 °C for 5 min. Amplified PDT products were purified by addition of 65 µl of SPRIselect beads (Beckman Coulter); 160 µl of supernatants were saved and incubated with 192 µl of SPRIselect. Beads were washed twice with 800 µl of 80% ethanol and the PDT library was eluted in 40 µl of buffer EB (QIAGEN).

The concentration of PDT libraries was determined and 15 ng was used for 100-µl indexing PCR reactions using 50 µl of Amp-Mix (10× Genomics), 7.5 µl of SI-PCR Primer B (10× Genomics) and 2.5 µl of i7 sample index-containing primers (10× Genomics); cycling conditions were: 98 °C for 45 s; 6 cycles at 98 °C for 20 s, 67 °C for 30 s and 72 °C for 20 s; and a final extension of 72 °C for 1 min. Indexed PDT libraries were purified by the addition of 120 µl of SPRIselect and eluted in 40 µl of buffer EB. The concentration of the final libraries was determined using a Qubit dsDNA HS Assay kit (Invitrogen) and size distribution was examined by running a High Sensitivity DNA chip on a Bioanalyzer 2100 system (Agilent).

PDT and ATAC libraries were pooled and paired-end sequenced (2 × 34 cycles) using Nextseq High Output Cartridge kits on a Nextseq 550 sequencer (Illumina). Raw sequencing data were demultiplexed with CellRanger-ATAC mkfastq. ATAC fastqs were used for alignment to the GRCh38 or mm10 reference genomes using CellRanger-ATAC count v.1.0.

ASAP-seq and PHAGE-ASAP of PBMCs. For ASAP-seq and PHAGE-ASAP, PBMCs were resuspended in cold FC buffer and first blocked with Human TruStain FcX (BioLegend) for 10 min at 4 °C. Cells were then either stained with TotalSeq-A antibodies (BioLegend; listed in Supplementary Table 5) or costained with both pNbs and TotalSeq-A antibodies by incubation on a rotating wheel

for 20 min at 4 °C. Cells were then washed and processed as outlined above for PHAGE-ATAC with the following modifications: before droplet encapsulation and barcoding, 0.5 µl of 1 µM TotalSeq-A bridge oligo EF369 was added per 65 µl of Chromium ATAC barcoding master mix, as described previously[9]. After MyONE silane bead cleanup, samples were eluted in 43 µl of elution buffer I (the Chromium Single Cell ATAC Solution protocol) and 3 µl of eluate was used for ADT library preparation by PCR using primers EF147, and EF370, EF371 or EF372, to obtain indexed ADT libraries as reported[9]. For PHAGE-ASAP, 5 µl of eluate was used for PDT library preparation as described above.

Computational workflow for generation of PDT count matrices. PDT fastqs were obtained by running CellRanger-ATAC mkfastq on raw sequencing data and customized UNIX code was used to derive PDT-cell barcode count tables. Customized Python code ('phage_to_kite-R3.py') was used to reformat the 10× scATAC R1/R2/R3 file conventions into a paired-end read file format compatible with kallisto|bustools for quantification. Using a kmer length of 13 for CDR3 regions, PDT libraries were pseudo-aligned to a user-specific reference and per-cell counts were determined using error-corrected barcodes and bustools (up to one mismatch for both cell barcodes and CDR3 barcodes). Notably, as phages do not have unique molecular identifiers, we used a dummy poly(A) sequence for compatible processing.

Analysis of species-mixing PHAGE-ATAC experiment. PHAGE-ATAC-seq data from the species-mixing experiment were demultiplexed using CellRanger-ATAC mkfastq, and generated ATAC fastqs were processed with CellRanger-ATAC count to filter reads, trim adapters, align reads to both GRCh38 and mm10 reference genomes, count barcodes, identify transposase cut sites, detect accessible chromatin peaks and identify cutoffs for cell barcode calling. The 'force-cells' parameter was not set. Barcodes were classified as human or mouse if >90% of barcode-associated fragments aligned to GRCh38 or mm10, respectively. Cutoffs for cell barcode calling were >3,000 ATAC fragments overlapping peaks for human and >10,000 for mouse barcodes (based on empirical density). Doublet barcodes were defined as containing >10% ATAC fragments aligning to both GRCh38 and mm10 reference genomes. The EGFP PDT count table was generated as described above by searching PDT fastqs for the corresponding phage barcode (Supplementary Table 4) and deriving PDT-associated cell barcodes via filtering using the entire list of called cell barcodes (human and mouse).

After flow cytometry measurement of HEK293T-EGFP-GPI (EGFP+) and HEK293T cells (EGFP−), FCS files were exported using CytExpert Software (Beckman Coulter). Values for forward scatter (FSC area) and EGFP fluorescence (FITC area) were derived from FCS files. Human EGFP+ and EGFP− cells were defined based on the distribution of EGFP PDT counts (for PHAGE-ATAC) or EGFP fluorescence represented by FITC-area values (for flow cytometry) by setting a gate at the minimum value in between both populations.

Analysis of PBMC PHAGE-ATAC, ASAP-seq and PHAGE-ASAP experiments. Sequencing data from PHAGE-ATAC, ASAP-seq and PHAGE-ASAP libraries of PBMCs were processed using CellRanger-ATAC count to the GRChg38 reference genome using all default parameters, yielding 1,408 (PHAGE-ATAC), 5,654 (PHAGE-ASAP) and 4,806 (ASAP-seq) high-quality PBMCs, respectively (no filtering was applied beyond the CellRanger-ATAC knee call). Per-library ADTs[9] and PDTs were computed using the processing pipelines described above. We further downloaded processed CITE-seq PBMC data[8] from the Gene Expression Omnibus (GEO, accession no. GSE100866), which resulted in recovery of 7,660 PBMCs after removal of spiked-in mouse cells. This published dataset was jointly analyzed with the newly generated datasets described above. We performed data integration using canonical correlation analysis[20] and the 2,000 most variable RNA genes is the default in Seurat. Next, we performed RNA imputation for the ATAC-seq data using Seurat v.3 with the default settings[33]. Reduced dimensions and cell clusters were inferred using this merged object via the first 25 canonical correlation components, with the default Louvain clustering in Seurat v.3. Centered log(ratio) (CLR)-normalized PDTs were visualized in the reduced dimension space and a per-tag, per-cluster mean was computed to further access staining correlation between the modalities (Fig. 2d).

Cell annotations were derived based on well-established marker genes for PBMCs (Extended Data Fig. 6h). For protein-based clustering and analyses, we identified T-cell clusters from the integrated embedding (using the chromatin/RNA data) and then further stratified them into subpopulations based on CD4 and CD8 PDT CLR (Extended Data Fig. 6d,f). Differential gene activity scores between these populations were then computed using the default functionality in Seurat/Signac (Wilcoxon's rank-sum test). To compare the protein quantification of each modality, we utilized the labeled clusters of CD4 and CD8 T cells (computed using only accessible chromatin and RNA abundances) as positive labels and other cell types as negatives (thus, the labels are a function of clustering and are imperfect). Utilizing these per-cell positive and negative annotations, we determined the receiver operating curves for each protein in each modality (Fig. 2e).

To verify the high-quality capture of somatic mtDNA mutations in this experiment, sequencing reads aligning to chrM were processed using mgatk, as previously described[11]. A total of 518 high-quality variants were identified in at

least one cell using the standard variant thresholds (variance mean ratio >−2; strand correlation >0.65), and the enrichment of nucleotide substitutions matched our previously identified patterns of strand-specific transitions[11].

Analysis of cell hashing PHAGE-ATAC experiment. One channel of sequencing data from the hashed, combined, CD8-enriched T cells was processed using CellRanger-ATAC count via the GRCh38 reference genome and all default parameters, yielding 8,366 high-quality PBMCs (no filtering was applied beyond the CellRanger-ATAC knee call). As we suspected the presence of contaminating B cells, we first characterized cell states using latent semantic indexing (LSI)-based clustering and dimensionality reduction using Signac and Seurat[33]. Specifically, all detected peaks were used as input into LSI. The first 20 LSI components (except for the first component, which was found to be correlated with the per-cell sequencing depth) were used to define cell clusters using the default Louvain clustering algorithm in Seurat. Per-cluster chromatin accessibility tracks were computed using a per-million fragments abundance for each cluster, as previously implemented[11]. These chromatin accessibility tracks were used to annotate cell clusters based on promoter accessibility of known marker genes.

To assign hash identities to cell barcodes, we utilized the HTODemux function from Seurat[23] with the positive.quantile parameter set at 0.98. This yielded 703 doublets, 1,225 negatives and 6,438 singlets based on the abundance and distribution of CD8 hashtag PDTs.

To verify PHAGE-ATAC hashtag-based assignments, we performed mtDNA genotyping using mgatk[11], and nuclear genotyping and donor assignment using souporcell[24] with '--min_alt 8 --min_ref 8 --no_umi True -k 4 --skip_remap True --ignore True' options, which resulted in 92.9% accuracy (99.3% singlet accuracy, 74% overlap in called doublets), confirming the concordance of our hashing design.

Analysis of PBMC–HEK293T mixture PHAGE-ATAC experiment. Due to the low cell knee call by the default CellRanger-ATAC knee call (probably due to the mixture of PBMCs and HEK293T cells), we manually identified high confidence cells that had a TSS score >4 and at least 500 accessible chromatin fragments in peaks, yielding 4,690 cells. Using components 2–30 from LSI, we produced a dimensionality reduction and clustering with Signac[33]. PDTs were quantified using kallisto|bustools as described above for all phages used in the library.

Cloning of PANL, a synthetic high-complexity pNb library. To generate randomized library inserts, three separate primer mixes (for long CDR3, medium CDR3 and short CDR3 inserts) were used for PCR-mediated assembly. For short CDR3 inserts, the primer mix contained 0.5 µl of each of polyacrylamide gel electrophoresis-purified EF42, EF43, EF64, EF44, EF65, EF45, EF46, EF47, EF66 and EF48 (each 100 µM) (EllaBiotech). For medium CDR3 inserts, EF67 was used instead of EF66. For long CDR3 inserts, EF68 was used instead of EF66. Primer mixes were diluted 1:25 and 1 µl of each mix was used for overlap-extension PCR using Phusion (NEB). Four 50-µl reactions for each mix were performed using the following cycling conditions: 98 °C for 1 min; 20 cycles at 98 °C for 15 s, 60 °C for 30 s and 72° for 30 s; and a final extension of 72 °C for 5 min. PCR reactions of the same mix were pooled and purified by addition of 280 µl of AMPure XP beads (Beckman Coulter). The beads were washed twice with 800 µl of 80% ethanol and assembled inserts were eluted in 100 µl of water. Concentrations of each insert (long, medium, short) were determined and pooled in a 1:2:1 molar ratio. Five identical 50-µl PCR reactions with pooled inserts and primers EF40 and EF41 were performed using Phusion (NEB), with the following cycling conditions: 98 °C for 1 min; 30 cycles at 98 °C for 15 s, 62 °C for 30 s and 72 °C for 30 s; and a final extension of 72 °C for 5 min. The amplified library insert was pooled and purified by adding 350 µl of AMPure XP beads (Beckman Coulter). Beads were washed twice with 1 ml of 80% ethanol and the library insert was eluted in 60 µl of water. Five identical 60-µl restriction digest reactions for the digest of 7.5 µg of library vector pDXinit-PAC with 2.5 µl of SapI were performed. Library insert (4.8 µg) was digested in a 30-µl reaction using 2.5 µl of SapI. Digests were incubated for 4 h at 37 °C and loaded on to 1% agarose gels. Bands corresponding to digested library vector and insert were cut and products were extracted using a GeneJet Gel Extraction Kit (Thermo Fisher Scientific) and eluted in 40 µl of water. Five identical 100-µl ligation reactions were performed, each containing 1.25 µg of digested pDXinit-PAC, 450 ng of digested insert and 0.5 µl of T4 ligase (NEB). Ligations were incubated for 16 h at 16 °C, heat inactivated for 20 min at 65 °C and cooled to room temperature. Then, 100 µl of AMPure XP beads was added to each ligation reaction, the beads were washed twice using 300 µl of 80% ethanol and ligation products were eluted in 15 µl of water, and pooled. Five electroporations in 2-mm cuvettes (BioRad) were performed, each using 90 µl of electrocompetent SS320 *E. coli* (Lucigen) and 12 µl of ligation product. Pulsing was performed on a GenePulserXcell instrument (BioRad) with parameters 2.5 kV, 200 Ω and 25 µF. After electroporation, bacterial suspensions were added to 120 ml of prewarmed super optimal broth with catabolite repression (SOC) and incubated for 30 min at 37 °C and 225 r.p.m. An aliquot of library-carrying bacteria was saved at this point and used to prepare a dilution series. Each dilution was plated on LB-Amp plates. After overnight incubation at 37 °C, colonies were counted, transformation efficiency was determined and library complexity was estimated. The remaining 120 ml of library-containing culture were added to 1.125 l of

2YT medium containing 2% glucose, 50 µg ml−1 of ampicillin and 10 µg ml−1 of tetracycline (2YT/2%/A/T) and incubated overnight at 37 °C and 240 r.p.m. The library-containing culture was harvested, glycerol stocks were prepared and library aliquots were stored.

Analysis of picked PANL clones using PCR and Sanger sequencing. Library-containing bacteria were plated on LB-Amp, incubated overnight, and colonies were picked and inoculated in 8 ml of LB-Amp. Cultures were incubated for at least 8 h at 37 °C and 240 r.p.m. Bacteria were harvested and plasmids isolated using GeneJet Plasmid Miniprep kit (Thermo Fisher Scientific). PCR was performed to evaluate clone inserts; 10-µl PCR reactions were set up that contained 10 ng of isolated plasmid, 0.5 µl each of primers EF52 and EF53, and 4.5 µl of 2× OneTaq Quick Load Master Mix (NEB). The cycling conditions were: 94 °C for 4 min; 28 cycles at 94 °C for 15 s, 62 °C for 15 s and 68 °C for 30 s; and a final extension at 68 °C for 5 min. PCR reactions were analyzed on 2% agarose gels. Selected clones were analyzed by Sanger sequencing using primer EF17.

Phage Nb library production. A PANL aliquot corresponding to 3×10^{10} bacterial cells (around 5× coverage of the library) was transferred to 200 ml of 2YT/2%/A/T and cultures were grown until $OD_{600} = 0.5$ was reached (~2 h). Cultures were infected with 8 ml of M13K07 helper (NEB) for 60 min at 37 °C. They were then harvested, supernatants discarded and bacterial pellets resuspended in 1 l of 2YT/A/K. Cultures were incubated overnight at 37 °C and 250 r.p.m. for production of the input library of pNb particles. Bacterial cultures were harvested, supernatants collected and phages precipitated using poly(ethylene glycol)/NaCl as described earlier. Final phage pellets were resuspended in a total of 20 ml of PBS and stored. Phage titers were determined by infecting a log(phase culture) of SS320 with a dilution series of the produced phage library and plating bacteria on LB-Amp. Colonies were counted and titers calculated. Produced phage libraries were characterized by titers $>4 \times 10^{11}$ plaque-forming units (p.f.u.) ml−1.

Phage display selection. HEK293T cells were transiently transfected with either pCAG or pCAG-EGFP-GPI for 24 h as described above. Cells were harvested, 10^7 pCAG-transfected cells were resuspended in 1 ml of PBS containing 2% BSA (PBS–BSA) and 8 ml of the PANL library (1.6×10^{12} p.f.u.) in PBS–BSA was added for counterselection. Samples were incubated for 1 h on a rotating wheel at 4 °C and then centrifuged at 350*g* for 5 min at 4 °C. Supernatants containing phages were added to 10^7 pCAG-EGFP-GPI-expressing cells for positive selection. After 1 h on a rotating wheel at 4 °C, samples were centrifuged (350*g*, 5 min and 4 °C) and washed six times with PBS–BSA to remove unbound phages. Cells were washed once in PBS and centrifuged, and cell pellets were resuspended in 500 µl of trypsin solution (1 mg ml−1 of trypsin (Sigma-Aldrich) in PBS) to elute bound phages. Cells were incubated for 30 min on a rotating wheel at room temperature and digests were stopped by the addition of AEBSF protease inhibitor (Sigma-Aldrich) to a final concentration of 0.5 mg ml−1. Samples were centrifuged (400*g* and 4 min at room temperature) and the supernatant containing eluted phages was used to infect 10 ml of log(phase culture) SS320 ($OD_{600} = 0.4$). After infection for 40 min at 37 °C, cultures were added to 90 ml of 2YT/2%/A/T and incubated overnight at 37 °C and 250 r.p.m. Cultures containing output libraries were aliquoted and glycerol stocks were prepared. Output library phage particles were prepared as described earlier for PANL and used in subsequent selection rounds using the same protocol described here.

Sequencing of PANL and selection output libraries. Bacterial cultures harboring phagemid libraries were grown overnight at 37 °C and 240 r.p.m. in 50 ml of LB containing 2% glucose and 50 µg ml−1 of ampicillin. Bacteria were harvested and plasmids isolated using ZymoPURE II Plasmid Midiprep Kit (Zymo Research). A first PCR was performed to amplify Nb inserts; 100-µl PCR reactions were set up that contained 100 ng of isolated plasmid library, 2.5 µl of primer mix EF235–EF241 and 2.5 µl of mix EF249–EF255, and 50 µl of 2× KAPA HiFi HotStart ReadyMix (Roche). The cycling conditions were: 95 °C for 3 min; 16 cycles at 95 °C for 20 s, 60 °C for 30 s and 72 °C for 20 s; and a final extension of 72 °C for 5 min. Nb amplicons were purified by addition of 120 µl of SPRIselect beads (Beckman Coulter), beads were washed twice with 200 µl of 80% ethanol and Nb product libraries were eluted in 40 µl of buffer EB (QIAGEN).

Concentration of amplicon libraries was determined and 20 ng was used for 100-µl indexing PCR reactions using 50 µl of 2× KAPA HiFi HotStart ReadyMix (Roche), 2.5 µl of primer EF242 and 2.5 µl of primer EF256; cycling conditions were: 95 °C for 45 s; 6 cycles at 95 °C for 20 s, 67 °C for 30 s and 72° for 20 s; and a final extension of 72 °C for 1 min. Indexed amplicon libraries were purified by the addition of 120 µl of SPRIselect and eluted in 40 µl of buffer EB. The concentration of the final libraries was determined using a Qubit dsDNA HS Assay kit (Invitrogen) and size distribution was examined by running a High Sensitivity DNA chip on a Bioanalyzer 2100 system (Agilent). Amplicon libraries were pooled and paired-end sequenced (read 1: 96 cycles, read 2: 184 cycles) on a MiSeq sequencer (Illumina).

Analysis of phagemid library sequencing experiments. Customized Python code ('process_phage_library_construct.py') was written to parse out the variable CDR1, CDR2 and CDR3 sequences using a positional sequence logic relative to constant regions in the PANL library design. Sequencing reads where constant regions could not be identified (up to two mismatches) were discarded, noting that all libraries had at least 90% parsing efficiency (range: 90–94%). Library complexity was estimated from annotated duplicate reads with identical variable CDR sequences using the Lander–Waterman method[34]. Nucleotide sequences were converted to amino acid sequences using a standard codon dictionary. To account for sequencing errors in annotating clones, we determined clones based on a rank ordering of sequences and collapsed any sequencing within a Hamming distance of 2 (based on nucleotide identity of the variable CDR1, CDR2 and CDR3 sequences). Per-position amino acid frequencies were estimated using the top 1,000 collapsed clones.

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

Data availability
Data associated with this work are available at GEO accession no. GSE157486. Source data are provided with this paper.

Code availability
Code to reproduce custom analyses and figures associated with this work is available at https://github.com/evgenijfiskin/phage-atac.

References
31. Geertsma, E. R. & Dutzler, R. A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry* **50**, 3272–3278 (2011).
32. McMahon, C. et al. Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat. Struct. Mol. Biol.* **25**, 289–296 (2018).
33. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 e1821 (2019).
34. Lander, E. S. & Waterman, M. S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231–239 (1988).
35. Kubala, M. H., Kovtun, O., Alexandrov, K. & Collins, B. M. Structural and thermodynamic analysis of the GFP:GFP–nanobody complex. *Protein Sci.* **19**, 2389–2401 (2010).

Author contributions
E.F. conceived and designed the project with guidance from A.R. E.F. designed and performed experiments. E.F. developed the PHAGE-ATAC computational workflow with input from C.A.L. E.F. developed the PHAGE-ATAC experimental protocol with input from L.S.L. E.F. and C.A.L analyzed the data. G.E. contributed to data analysis. F.L. and A.M.R. selected CEACAM4 Nb and validated CEACAM4 Nb. A.R. and R.J.X. provided project oversight. A.R. acquired funding. E.F. and A.R. wrote the manuscript with input from all authors.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41587-021-01065-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-021-01065-5.

Correspondence and requests for materials should be addressed to Evgenij Fiskin or Aviv Regev.

Peer review information *Nature Biotechnology* thanks Andrew Adey, Dan Xie 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 | Barcoding strategies for epitope quantification by PHAGE-ATAC vs. CITE-seq. a, Nanobody-displaying phages for PHAGE-ATAC. The phagemid contained within a particular phage particle encodes the protein displayed on that same phage, and PHAGE-ATAC leverages the hypervariable nanobody CDR3 sequences as unique genetic barcode identifiers for each phage. **b**, Oligonucleotide-conjugated antibodies for CITE-seq. Each antibody is separately conjugated with a unique DNA-barcode. **c**, Schematic of gel bead oligos showing Illumina P5 sequence (P5), random bead barcode (BC) and the first 14 bp of RD1 used for hybridization with RD1-containing Tn5-derived chromatin fragments and engineered PHAGE-ATAC phagemids. **d**, Nanobody-encoding phagemid constructs for RD1-mediated CDR3 barcode capture by 10x Genomics primers. The top strand is the coding strand. Orientation (arrows and shaded boxes), nucleotide sequence and translation product of RD1-containing constructs are shown. To avoid generating a stop codon by introduction of RD1 into the nanobody-p3 reading frame additional codons are introduced to maintain the reading frame across RD1, thus establishing the PAC tag. **e**, Agarose gel after two-step PCR consisting of linear amplification using the 10x ATAC primer followed by exponential PCR using P5 and Illumina Read 2 (RD2)-containing nanobody-specific primers. PDTs were only obtained for PAC-tagged phagemids with RD1 located on the non-coding strand (3'-5' orientation relative to nanobody). Abbreviations as in **c**. Control PCR was performed using two primers hybridizing within the nanobody sequence (Methods). A representative gel image from three independent experiments is shown. **f**, Schematic of post barcoding steps for the generation of ATAC and PDT sequencing libraries (Methods). After breaking emulsions, barcoded linear amplification products are purified and samples are split. ATAC fragment libraries are immediately processed for sample index PCR. PDT libraries are first amplified in a PDT-specific PCR using a CDR3 flanking constant nanobody sequence as PCR handle. PDT amplification allows RD2 adapter introduction required for final sample indexing. P5 and P7, Illumina P5 and P7 sequences. CBC, random 10x bead cell barcode. i7, sample index.
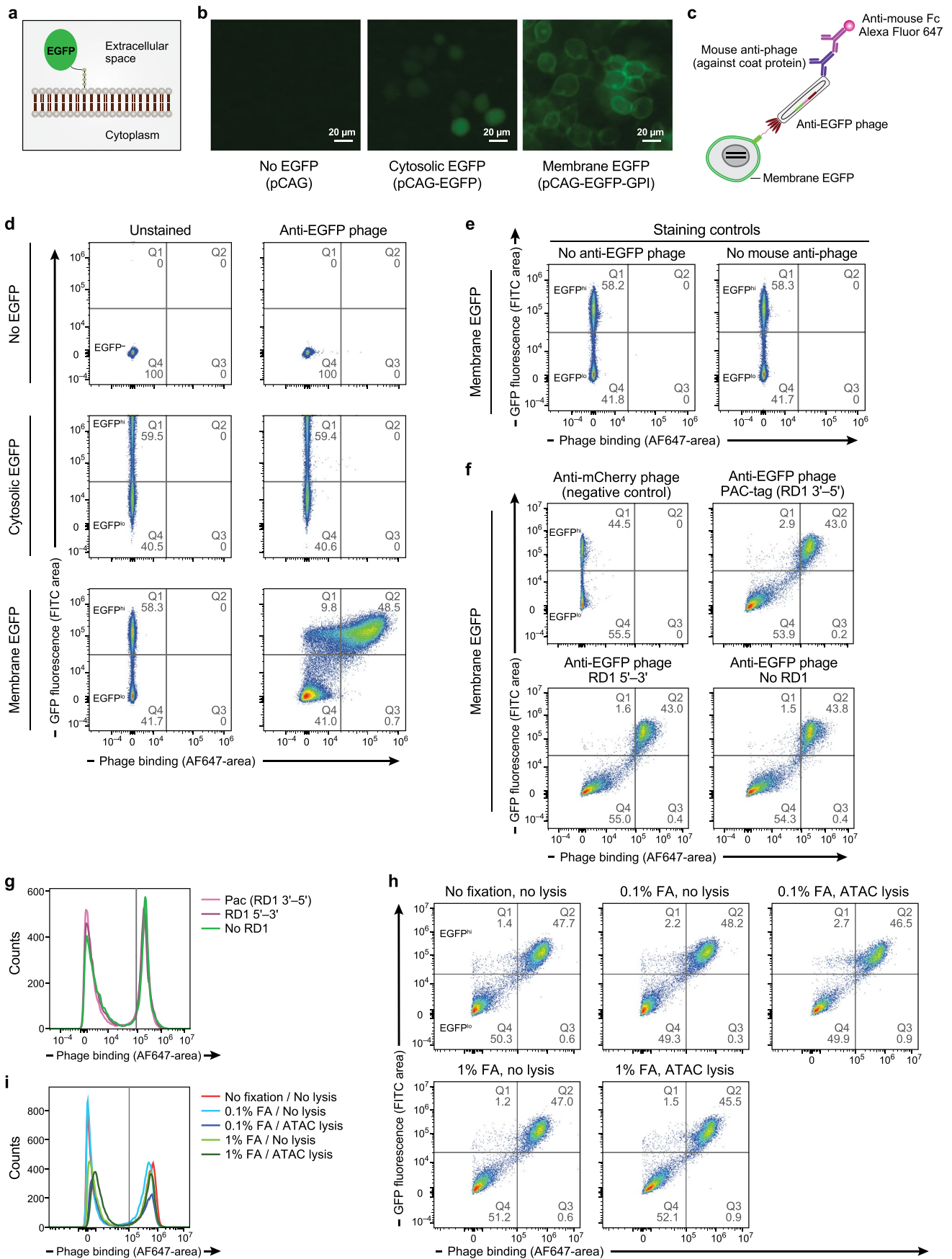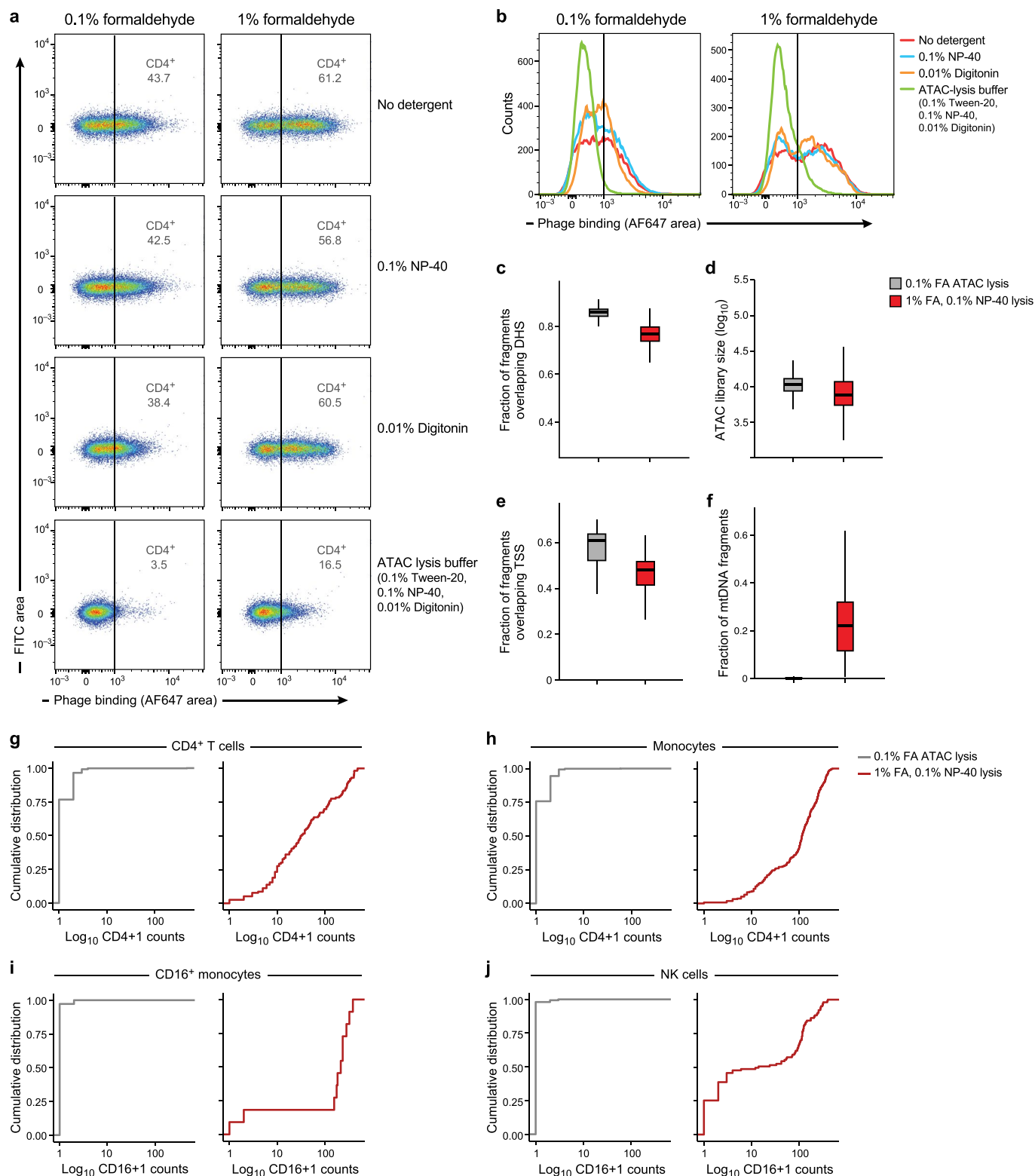
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Detection of membrane-localized EGFP via anti-EGFP nanobody-displaying phages and optimization of fixation and lysis conditions. a,b, Membrane expressed EGFP. **a,** Schematic of surface-exposed GPI-anchored EGFP. **b,** Microscopy images of HEK293T cells expressing indicated constructs, showing differential localization of untagged cytosolic EGFP (pCAG-EGFP, middle) and GPI-anchored membrane-localized EGFP (pCAG-EGFP-GPI, right, Methods), a representative micrograph from 4 independent experiments is shown. **c,** Schematic for the detection of phages via flow cytometry. Phage-stained cells are incubated with mouse anti-M13 coat protein antibodies followed by detection by Alexa Fluor 647-conjugated anti-mouse secondary antibodies. Phage binding is thus reflected by Alexa Fluor 647 signal. **d,** Flow cytometry analysis of anti-EGFP phage nanobody binding to EGFP-expressing HEK293T cells. EGFP fluorescence (y axis) and phage binding (x axis, Alexa Fluor 647) in each of the HEK293T cell populations as in **b,** either unstained (left) or stained with an anti-EGFP phage (right). EGFP-expressing cells were always characterized by the presence of both EGFP[hi] and EGFP[lo] populations. **e,** Specificity of detection. As in **d** but using the indicated staining controls for specific staining of membrane-EGFP-expressing cells. **f,g,** PAC-tag does not impact nanobody display and antigen interaction. EGFP fluorescence (**f**, y axis) and phage binding (**f**, x axis, Alexa Fluor 647) and distribution of level of phage binding (**g**) for phage-stained EGFP-GPI expressing cells using indicated phage nanobodies (for RD1 sequences see Extended Data Fig. 1d). **h,i,** EGFP fluorescence (**h**, y axis) and phage binding (**h**, x axis, Alexa Fluor 647) and histograms reflecting phage binding (**i**) for EGFP-GPI expressing cells stained with PAC-tagged anti-EGFP-Nb displaying phages using indicated fixation and permeabilization conditions.
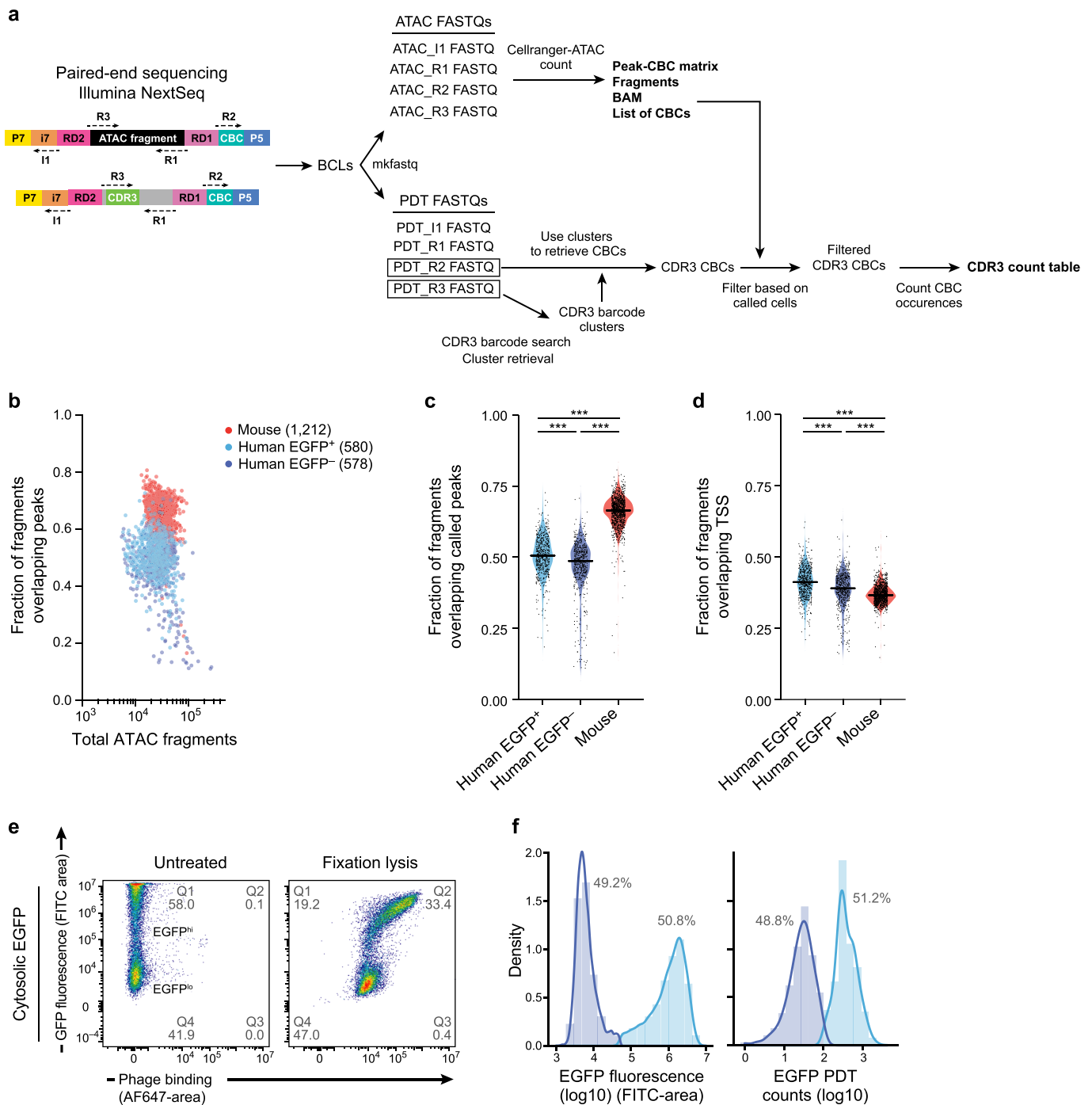
Extended Data Fig. 3 | PHAGE-ATAC quality metrics for human-mouse species-mixing experiment and intracellular EGFP detection via flow cytometry and PHAGE-ATAC. a, Paired-end sequencing output is demultiplexed using sample index information (left) to recover ATAC and PDT fastqs. ATAC fastqs are processed using CellRanger-ATAC count for fragment alignment, assignment of cell barcodes and generation of peak-cell barcode matrices. CDR3 barcode sequences are used to search PDT_R3 fastqs and identify CDR3-containing sequencing clusters. Matching of cluster identifiers is used to derive corresponding cell barcodes from PDT_R2 fastqs. Recovered PDT cell barcode lists are filtered using cell barcodes called by CellRanger. Cell barcode occurrences are counted to generate PDT-cell barcode count matrices (Methods). **b**, Fraction (y axis) and number (x axis, $\log_{10}$ scale) of unique chromatin fragments overlapping peaks for each barcode (dot) colored by populations (color legend). **c,d**, Distribution of fraction of unique ATAC fragments overlapping peaks (**c**, y axis) or TSS (**d**, y axis) in each of the three cell populations (x axis,, human EGFP+ cells n = 580, human EGFP- cells n = 578, mouse cells n = 1,212 from one experiment) (Mann-Whitney two-tailed, ***p < 10^{-4}, NS = not significant, in **c** human EGFP+ vs human EGFP- p = 2.155 ×10^{-10}, human EGFP- vs mouse p = 5.367 ×10^{-239}, human EGFP + vs mouse p = 1.711 ×10^{-225}, in **d** human EGFP + vs human EGFP- p = 3.311 ×10^{-12}, human EGFP- vs mouse p = 7.008 ×10^{-26}, human EGFP + vs mouse p = 6.099 ×10^{-77}). Line: median. **e,f**, Intracellular detection with PHAGE-ATAC. **e**, Flow cytometry analysis of anti-EGFP phage nanobody binding to cytosolic EGFP-expressing HEK293T cells. EGFP fluorescence (y axis) and phage binding (x axis, Alexa Fluor 647) in either untreated cells (left) or in cells post fixation and lysis (right) (Methods), both incubated with anti-EGFP pNb. Cytosolic EGFP-expressing cells were always characterized by the presence of both EGFP^hi and EGFP^lo populations. **f**, PDT quantification is consistent with flow cytometry results. Cytosolic EGFP fluorescence distribution (left, x axis) and distribution of EGFP PDT (right, x axis) in cytosolic EGFP+ (light blue) and EGFP− (dark blue) HEK293T cells.
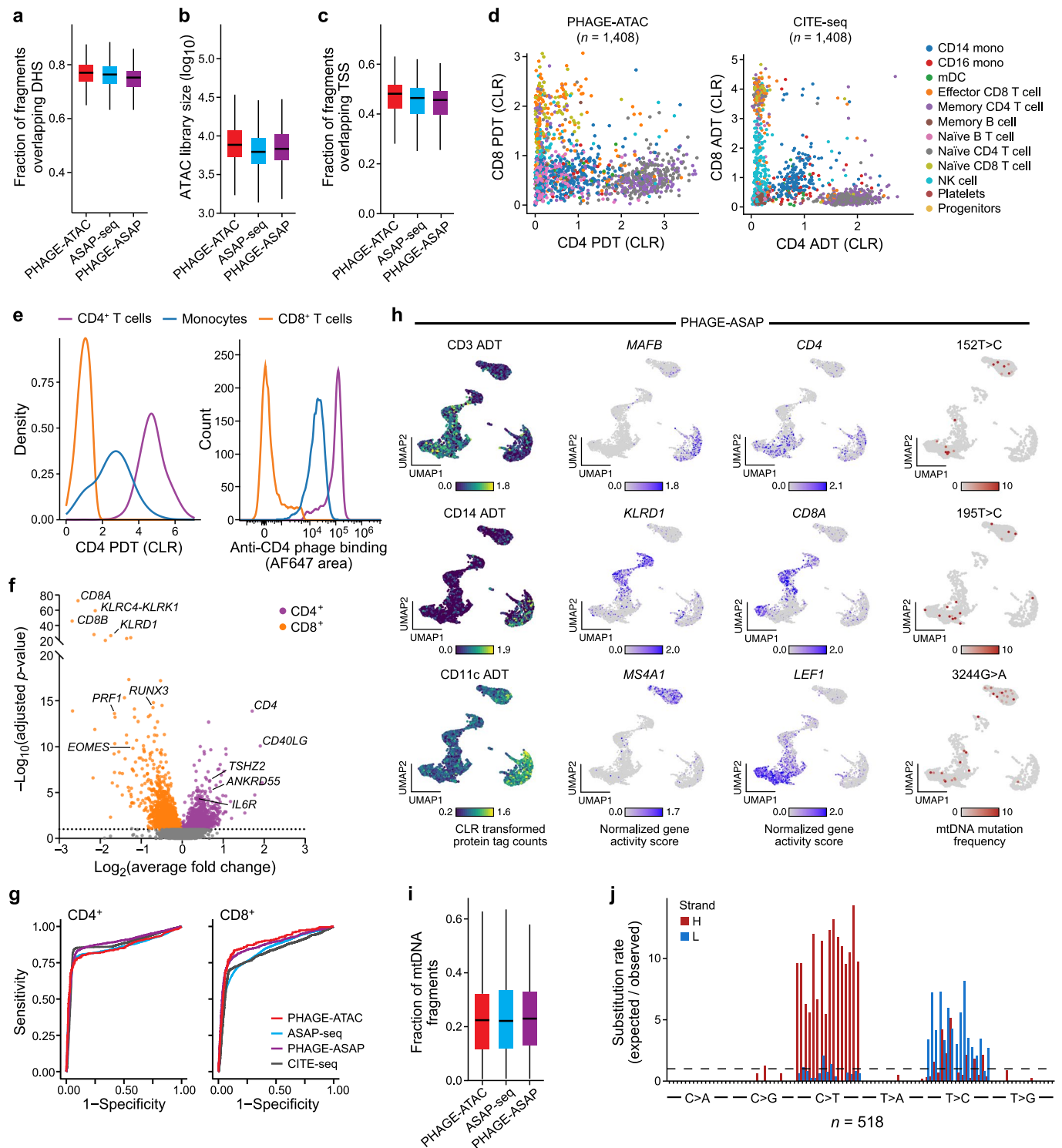
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Optimization of fixation and lysis conditions for PHAGE-ATAC using PBMCs. a, Binding of generated anti-CD4 phage nanobodies to PBMCs under indicated conditions. Two different formaldehyde concentrations (0.1% or 1%) as well as indicated lysis conditions were used. Phage binding is reflected by Alexa Fluor 647 fluorescent signal intensity. **b**, Histogram of data in **a**. **c-e**, Comparison of quality metrics of ATAC libraries under different fixation and lysis conditions. Distribution of the fraction of fragments in DNase hypersensitive sites (**c**, *y* axis), ATAC library size (**d**, *y* axis) or fragment fraction overlapping TSS (**e**, *y* axis) for each cell barcode after 0.1% formaldehyde fixation and ATAC lysis *vs.* 1% formaldehyde fixation and 0.1% NP-40 lysis (*x* axis, 0.1% fixation cells n = 2,153, 1% fixation cells n = 1,408, each from one experiment). **f**, Efficient capture of mtDNA requires optimized fixation and lysis conditions. Distribution of the fraction of mtDNA-derived fragments (*y* axis) for each cell barcode after 0.1% formaldehyde fixation and ATAC lysis *vs.* 1% formaldehyde fixation and 0.1% NP-40 lysis (*x* axis). **g-j**, Detection of PBMC-bound phages in PHAGE-ATAC is enabled by optimized fixation and lysis conditions. Cumulative distribution (**g-j**, *y* axis) of CD4 PDTs in CD4[+] T cells (**g**, *x* axis) or monocytes (**h**, *x* axis) and CD16 PDTs in CD16[+] monocytes (**i**, *x* axis) or NK cells (**j**, *x* axis) in PHAGE-ATAC data after processing cells using 0.1% formaldehyde fixation and ATAC lysis *vs.* 1% formaldehyde fixation and 0.1% NP-40 lysis. Boxplots for **c-f**: center line, median; box limits, first and third quartiles; whiskers, 1.5× interquartile range.
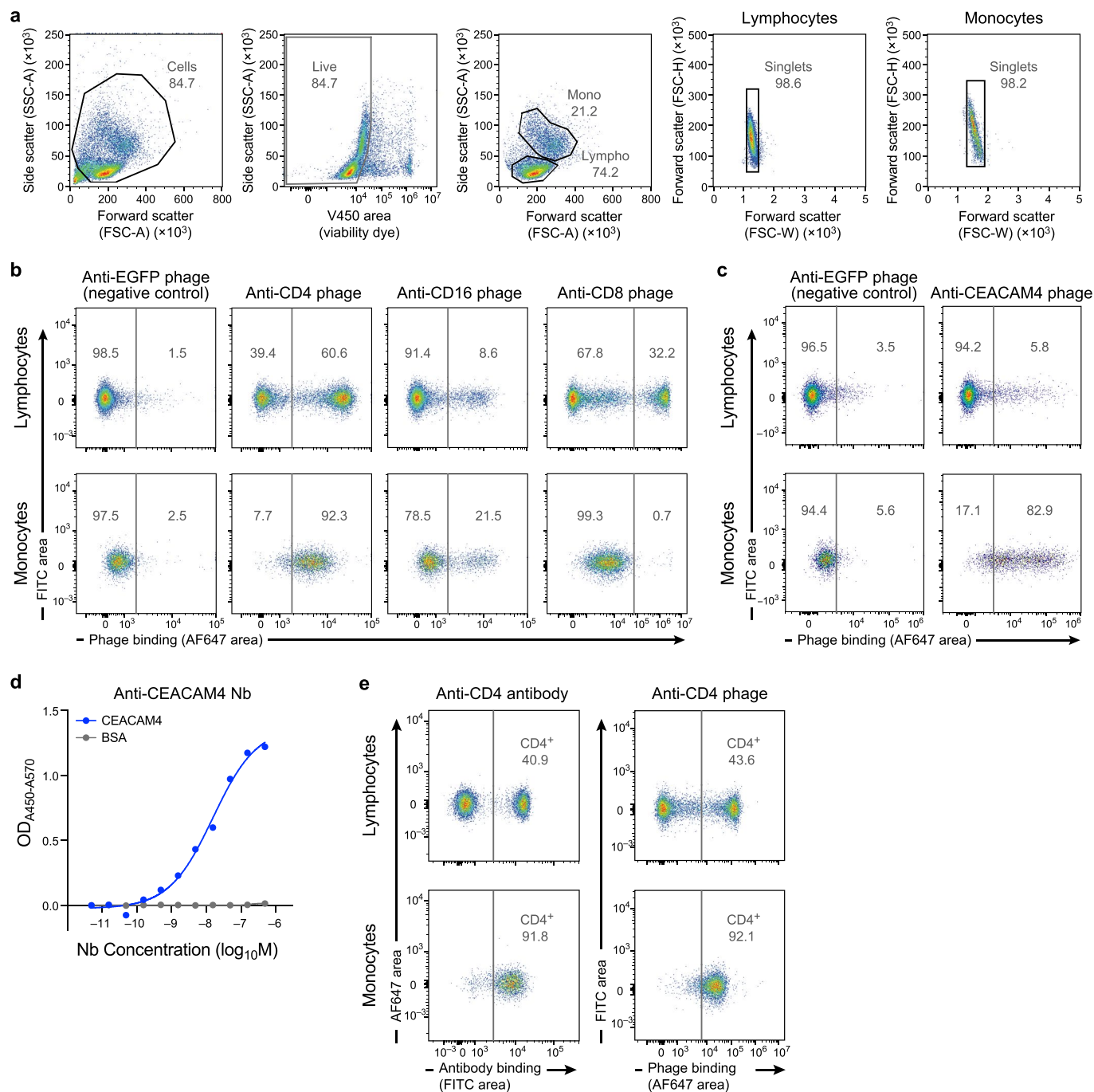
Extended Data Fig. 5 | Flow cytometry-based validation of PAC-tagged anti-CD4, anti-CD8, anti-CD16 and anti-CEACAM4 nanobody-displaying phages. a, Flow cytometry gating strategy for analyzed phage-stained PBMCs. **b-c**, Flow cytometry-based binding assessment of indicated surface marker-recognizing phage nanobodies to gated lymphocyte (**b**, CD4+ T cells, CD8+ T cells, CD16+ NK cells) and monocyte (**c**, CD4+ and CEACAM4+) populations, anti-EGFP pNb was used as negative control. **d**, Analysis of anti-CEACAM4 nanobody binding to purified CEACAM4. Binding of anti-CEACAM4 nanobody Fc-fusion protein to purified CEACAM4 (*y* axis, optical density 450nm-570nm) across varying concentrations of nanobody (*x* axis), purified BSA was used as negative control. **e**, Comparison of PBMCs stained with a well-characterized anti-CD4 antibody or generated anti-CD4 phage nanobody. Phage binding is reflected by Alexa Fluor 647 fluorescent signal intensity.

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

Extended Data Fig. 6 | Benchmarking of PHAGE-ATAC, PHAGE-ASAP, ASAP-seq and CITE-seq for human PBMCs profiling. a-c, Comparison of ATAC quality metrics across different assays. Fragment fraction in DNase hypersensitive sites (**a**, *y* axis), ATAC library size (**b**, *y* axis) or fragment fraction overlapping TSS (**c**, *y* axis) for each cell barcode in PHAGE-ATAC (cells n = 1,408), ASAP-seq (cells n = 4,806) or PHAGE-ASAP (cells n = 5,654) data (*x* axis, one experiment). **d**, Agreement in protein level measurements in PHAGE-ATAC and CITE-seq. PDT (left, PHAGE-ATAC) or ADT (right, CITE-seq[8]) counts (CLR transformed) of CD8 (*y* axis) and CD4 (*x* axis) in each cell (dots), colored by cell type annotation derived from sc-ATAC-seq (PHAGE-ATAC) or scRNA-seq (CITE-seq) profile based clustering (as in Fig. 2c). Cell barcodes in CITE-seq data were downsampled to match the PHAGE-ATAC experiment. **e**, CD4 PDT quantification is consistent with flow cytometry. Distribution (**e**, *y* axis) of CD4 PDT counts (CLR transformed) (**e** (left), *x* axis) or anti-CD4 phage binding (**e** (right), Alexa Fluor 647 fluorescent signal, *x* axis) in CD4[+] T cells, CD8[+] T cells and monocytes. **f,g**, PDT and ADT count-based classification of CD4[+] and CD8[+] T cells. **f**, Mean log$_2$(fold change) (*x* axis) and significance (*y* axis, -log$_{10}$(P-value)) for each gene activity score comparing PDT-classified CD4[+] and CD8[+] T cells in PHAGE-ATAC (Methods). *Bona fide* markers are labeled. **g**, AUROC curves of sensitivity (*y* axis) and specificity (*x* axis) for CD4[+] (left) and CD8[+] (right) T cell classification based on PDTs (PHAGE-ATAC and PHAGE-ASAP) or ADTs (ASAP-seq and CITE-seq) (Methods). **h**, Diverse modalities captured by PHAGE-ASAP. 2D embedding of PHAGE-ASAP scATAC-seq data of PBMCs, colored by protein ADTs or PDTs (left), normalized gene activity score (middle) or mtDNA mutation heteroplasmy (right). **i**, Comparable capture of mtDNA. Distribution of mtDNA-derived fragment fraction (*y* axis) for each cell barcode in PHAGE-ATAC, ASAP-seq or PHAGE-ASAP (*x* axis, cell numbers as in **a-c**). **j**, Nucleotide substitution rate for 518 mgatk-called mutations (*y* axis) across each group of mono- and trinucleotide changes (*x* axis) in heavy (H) and light strands (L) of mtDNA (color). Boxplots for **a, b, c and i**: center line, median; box limits, first and third quartiles; whiskers, 1.5× interquartile range.
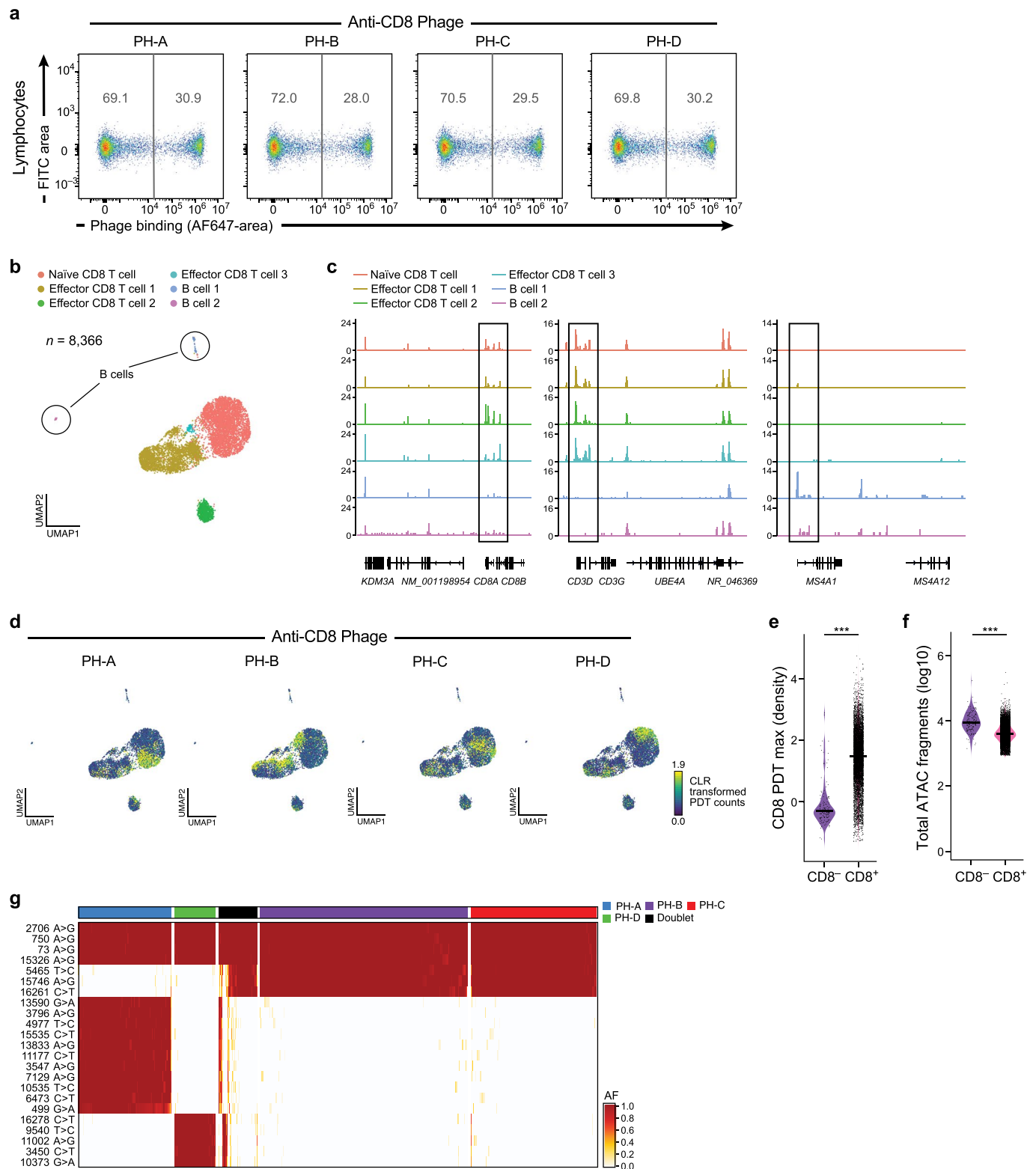
**Extended Data Fig. 7 | Sample multiplexing using hashtag phages. a**, Validation of phage hashtag binding. Flow cytometry of anti-CD8 hashtag phages bound (Alexa Fluor 647 fluorescent signal, *x* axis) to lymphocytes gated via flow cytometry of phage-stained PBMCs (as shown in Extended Data Fig. 5a). **b**, Cell type identification. Two-dimensional embedding of hashed CD8+ T cells analyzed by PHAGE-ATAC, colored by cell type annotation. **c**, Pseudobulk chromatin accessibility track plots for *CD8*, *CD3* and *MS4A1* (*CD20*) loci across identified cell types. **d**, Embedding as in **b** with cells colored by CD8 hashtag PDTs. **e,f**, Distribution of maximal CD8 PDT density (**e**, *y* axis) or unique chromatin fragments (**f**, *y* axis) for each cell barcode in CD8$^-$ (B cell 1 and B cell 2) and CD8$^+$ (non-B cell) cells (*x* axis, CD8$^-$ cells n = 137, CD8$^+$ cells n = 8,229, from one experiment) (Mann-Whitney two-tailed, ***p < 10^{-4}, in **e** p = 3.138 × 10^{-57}, in **f** p = 2.171 × 10^{-33}). **g**, Concordance between hashtag-based classification of barcodes and identified mtDNA SNPs. Heteroplasmy (allele frequency percentage; color bar) of different mtDNA variants (rows) in each cell (column), labeled by hashtag assignment (vertical top color bar).
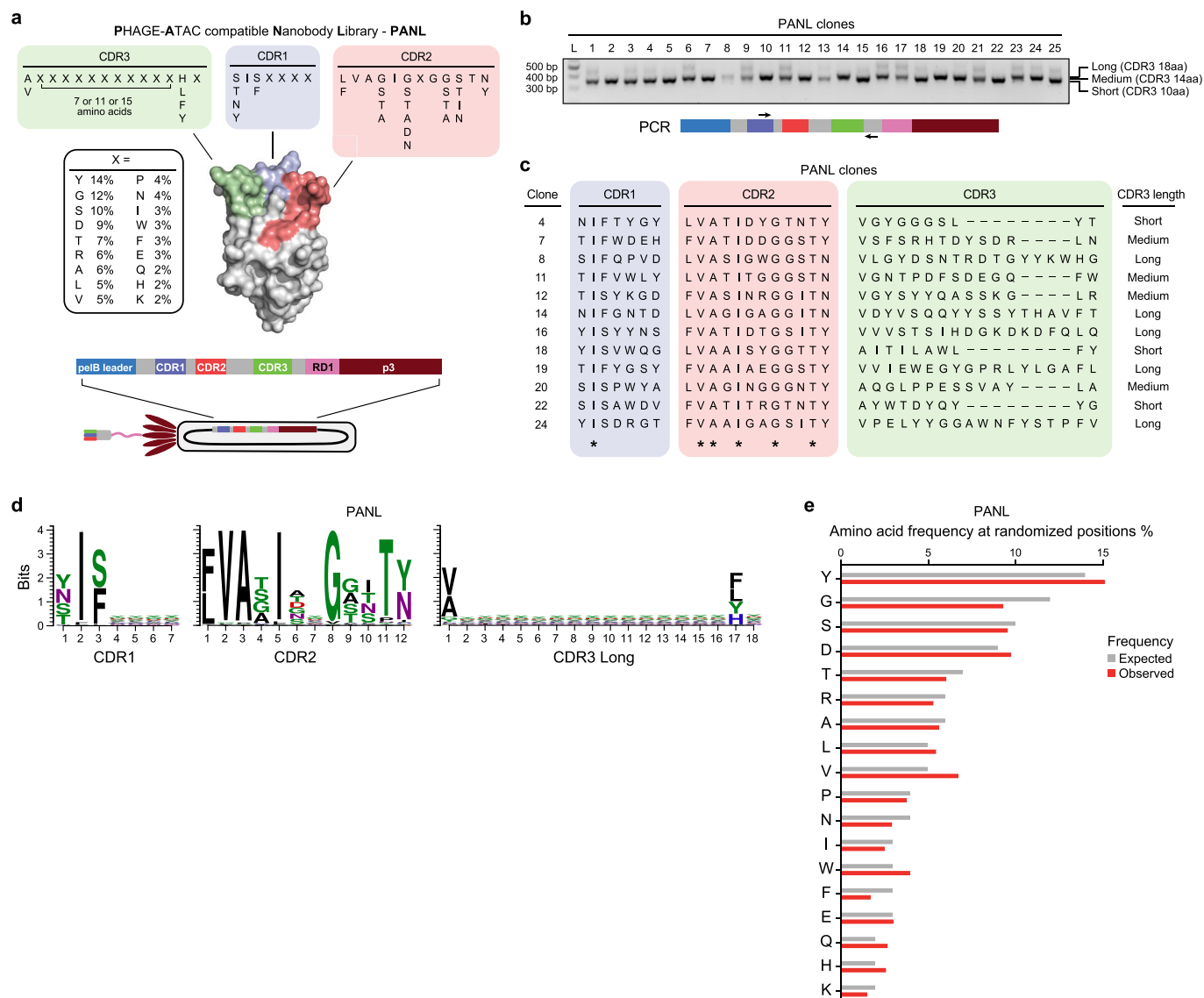
Extended Data Fig. 8 | Establishment of PANL, a fully synthetic high-complexity PAC-tagged phage nanobody library. a, Schematic of PANL library design and library phagemid. CDR3 sequence diversification and nanobody framework (grey) in PANL are based on a previously reported nanobody randomization strategy[32]. White box: expected frequency of amino acids at each hypervariable position (denoted by X), adjusted by using a custom randomized primer mix for library generation (Methods). CDR3 loops contained either 7, 11 or 15 hypervariable positions, resulting in total CDR3 lengths of 10 (short), 14 (medium) or 18 (long) amino acids. Partially randomized positions are depicted as columns, constant positions contain a single amino acid. A deposited structure of anti-EGFP Nb (PDB: 3ogo[35]) with colored CDR3 loops is shown. PANL phagemid is analogous to the one shown in Fig. 1a. **b**, Amplification products of phagemid insert-spanning PCR reactions using depicted primers for 25 randomly picked PANL clones. Product sizes due to presence of long, medium or short CDR3 are shown. **c**, CDR3 sequences of selected clones from **b** obtained by Sanger sequencing, CDR3 length is indicated, * non-randomized constant positions in the PANL library. **d**, Amino acid frequencies expressed as bits (y axis) at each CDR position (x axis) quantified by sequencing of PANL, values for the long CDR3 are shown representatively. **e**, Expected (grey) and observed (red) frequencies (x axis) of amino acids at hypervariable positions (y axis) (Methods).
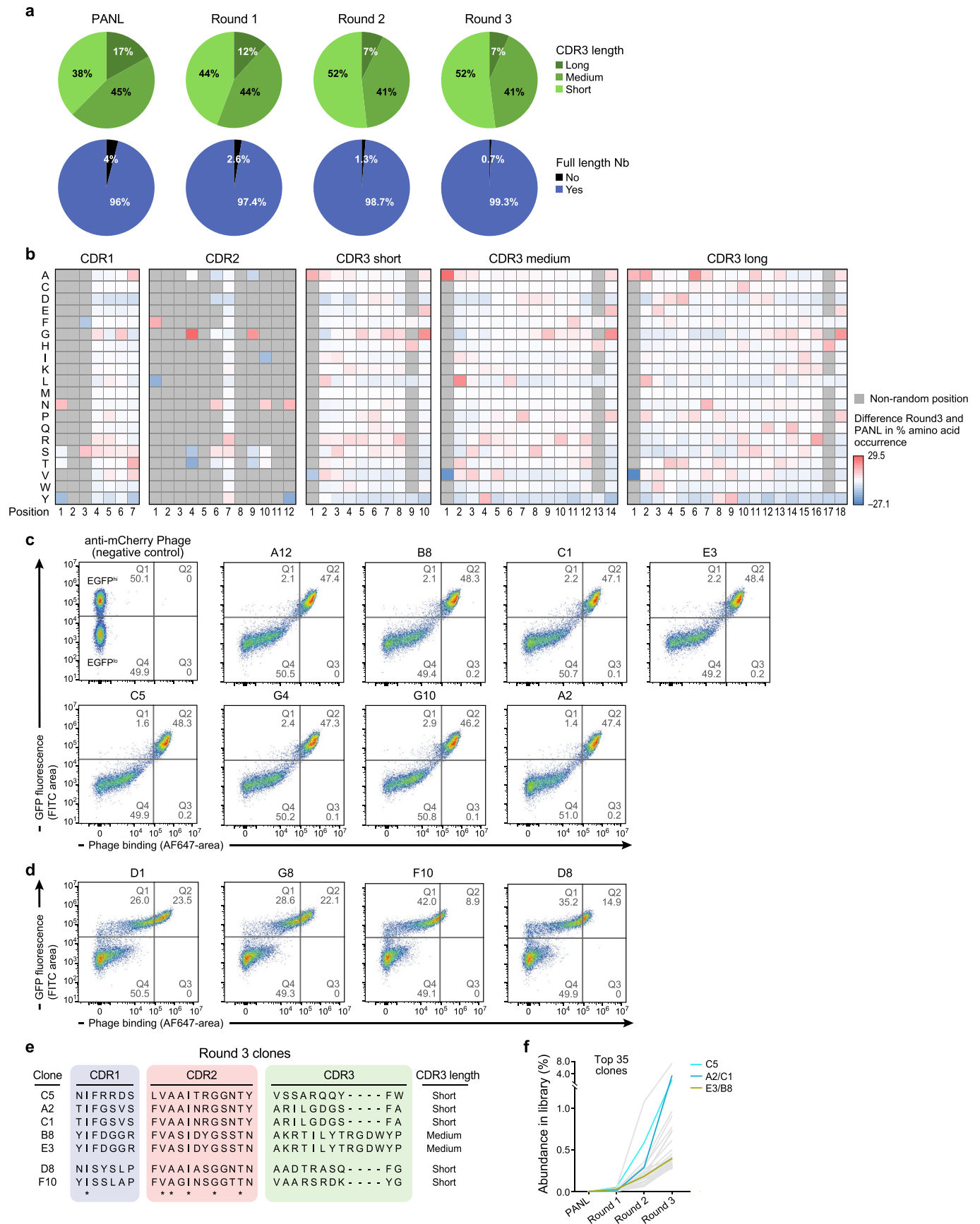
Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Phage display selection of antigen-specific nanobody-displaying phage clones. a-b, Analysis of selection process by sequencing of input and output phage libraries. **a**, Percentage of long, medium and short CDR3-containing pNb clones (top) or of full-length Nb sequences (bottom) within PANL and output libraries after each selection round. **b**, Difference in the frequency (color) of amino acids (rows) at each CDR position (columns) between the final selection output (Round 3) and PANL. **c,d**, Flow cytometry analysis of round 3 phage nanobody clones for binding to EGFP-GPI expressing cells (EGFPhi and EGFPlo populations can be observed) with either strong (**c**) or weaker (**d**) binders. Phage nanobodies against mCherry were used as negative control. Phage binding is reflected by Alexa Fluor 647 signal. **e**, CDR sequences and CDR3 length of selected clones obtained by Sanger sequencing. *non-randomized constant positions in PANL library (as in Extended Data Fig. 8a**). f**, Percentage (*y* axis, by Illumina sequencing) of flow cytometry validated clones (from **e**, color) in PANL and the output libraries after each selection round (see also Fig. 4c).

Extended Data Fig. 10 | Generation, screening and use of anti-SARS-CoV-2-S pNbs and PANL-derived pNbs for multiplex PHAGE-ATAC. a, Flow cytometry screening of 28 anti-SARS-CoV-2-S phage nanobody clones for binding to SARS-CoV-2-S expressing cells. Distribution of phage binding signal (Alexa Fluor 647, y axis) for each clone (panel) in SARS-CoV-2-S+ HEK239T cells (red) and control HEK239T with empty vector (blue). Side panel: Phage nanobodies against CD4 used as negative control. **b**, Detection of anti-SARS-CoV-2-S binding by PHAGE-ATAC. Distribution of PDT counts (y axis) in single cells for each anti-SARS-CoV-2-S (S_) and anti-EGFP phage (G_) nanobody (x axis, analyzed cells n = 4,690, from one experiment). Boxplot: center line, median; box limits, first and third quartiles; whiskers, 1.5× interquartile range. **c-g**, Mutually exclusive detection in appropriate cell types. **c**, 2D embedding of PHAGE-ATAC scATAC-seq data of a mixture of PBMCs and HEK293T cells (see also Fig. 4c), colored by normalized gene activity scores (left) or by the level of measured epitope PDTs (right). **d-g**, PDT counts (y and x axes) from pairs of pNbs recognizing distinct antigens across all cells (dots colored as indicated in **g** and Fig. 4c) for pNbs anti-CD4 (**d**, x axis) and anti-SARS-CoV-2-S (S_6) (**d**, y axis) or anti-CD4 (**e**, x axis) and anti-EGFP (clone C5) (**e**, y axis) or anti-CD16 (**f**, x axis) and anti-EGFP (clone C5) (**f**, y axis) or anti-CD4 (**g**, x axis) and anti-CD8 (**g**, y axis). Top left: Pearson's r.

nature research

Corresponding author(s):   Aviv Regev, Evgenij Fiskin

Last updated by author(s):   Aug 10, 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	Sequencing data was acquired using an Illumina NextSeq550 sequencer. Data was processed using CellRanger-ATAC v.1.1.0; Seurat v.3.1; Signac v.0.2; bedtools v.2.25.0; FlowJo software v.10.6.1; mgatk; souporcell; kallisto and bustools
Data analysis	Code to reproduce custom analyses and figures associated with this work will be available at https://github.com/evgenijfiskin/phage-atac

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

Data associated with this work is available at GEO accession GSE157486. Used published CITE-seq data is available at GEO accession GSE100866.

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. Comparative analysis between conditions involved hundreds or thousands of cells providing robust sample sizes in line with similar scRNA-seq comparisons and technologies.
Data exclusions	Common scATAC-seq quality metric cutoffs were applied to identify true cell barcodes as implemented in the CellRanger-ATAC workflow. While metrics for exclusion were determined ahead of time, the exact thresholds were determined empirically based on the density of single cells to determine appropriate, data-specific thresholds.
Replication	We replicated the concordant measurement of high-quality protein and chromatin accessibility data using the outlined workflows during method development and application in 10 successful cell line and primary cell experiments. Results in this study are representative.
Randomization	Not relevant to this study.
Blinding	Blinding is not relevant to our study, as there was no intervention, and 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	FITC-conjugated anti-CD4 antibody (OKT4, 317407, BioLegend) at 1:500 dilution; anti-M13 antibody (Sino Biological, 11973-MM05T-50) at 1:500 dilution; Alexa Fluor 647-conjugated anti-mouse Fc antibody (A-21236, Thermo Scientific) at 1:500 dilution
Validation	Each lot of these antibodies is quality control tested and validated by the respective supplier.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	ATCC: HEK293T cell line, https://www.atcc.org/products/all/crl-3216.aspx#characteristics; NIH-3T3 cell line, https://www.atcc.org/products/all/CRL-1658.aspx#generalinformation; PBMCs and CD8 T cells were purchased from AllCells.
Authentication	Cell lines were not authenticated.
Mycoplasma contamination	Cell lines are routinely tested for mycoplasma contamination. Results were consistently negative.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	For PBMC and CD8 T cell experiments, cryopreserved PBMCs or CD8 T cells (AllCells) were thawed, washed in PBS and resuspended in cold Flow cytometry buffer (FC buffer; PBS containing 2% FBS). For cell line mixing, culture medium was aspirated, cell lines were washed with PBS, harvested using Trypsin-EDTA 0.25% (Thermo Scientific), resuspended in DMEM containing 10% FBS, centrifuged, washed with PBS and resuspended in FC buffer. Cells were incubated with respective phage nanobodies for 20min on a rotator at 4°C. Cells were centrifuged and washed with cold FC buffer twice to remove unbound phages (centrifugation steps all were 350g, 4min, 4°C). Cells were resuspended in FC buffer and anti-M13 antibody (Sino Biological, 11973-MM05T-50) was added at 1:500 dilution. After 10min on ice, cells were washed twice in FC buffer and anti-mouse Fc Alexa Fluor 647-conjugated secondary antibody (Thermo Scientific, A-21236) was added at 1:500 dilution. Cells were incubated for 10min on ice, washed twice in FC buffer and resuspended in Sytox Blue (Thermo Scientific) containing FC buffer for live/dead discrimination according to manufacturer's instructions. In indicated cases, cells were stained with anti-CD4-FITC (clone OKT4, BioLegend) at 1:500 dilution, hereby no anti-M13 and anti-mouse Fc antibodies were used.
Instrument	CytoFLEX LX Flow Cytometer (Beckman Coulter)
Software	FlowJo software v.10.6.1
Cell population abundance	Not relevant to this study as no sorted populations were used.
Gating strategy	Singlets were discriminated and live cells were identified based on live/dead dye Sytox Blue (Thermo Scientific): FSC-A vs SSC-A > FSC-H vs FSC-W > Sytox-. Lymphocytes and Monocytes were gated based on their characteristic FSC-A vs SSC-A profiles. Fluorescent signals were either gated against each other or an unoccupied channel (i.e. FITC-A, for PBMCs). Negative boundaries were determined using a control phage nanobody staining (e.g. a-mCherry phage or a-EGFP phage).

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.