BRIEF COMMUNICATIONS

Multiplexed quantification of proteins and transcripts in single cells

Vanessa M Peterson[1,5], Kelvin Xi Zhang[2,5], Namit Kumar[1], Jerelyn Wong[3], Lixia Li[1], Douglas C Wilson[3], Renee Moore[4], Terrill K McClanahan[3], Svetlana Sadekova[3] & Joel A Klappenbach[1]

We present a tool to measure gene and protein expression levels in single cells with DNA-labeled antibodies and droplet microfluidics. Using the RNA expression and protein sequencing assay (REAP-seq), we quantified proteins with 82 barcoded antibodies and >20,000 genes in a single workflow. We used REAP-seq to assess the costimulatory effects of a CD27 agonist on human CD8[+] lymphocytes and to identify and characterize an unknown cell type.

Recent increases in the throughput of single-cell (sc) RNA-seq[1,2] experimentation has enabled its use in the identification and characterization of novel or rare cell types[3], in addition to providing insights into the underlying mechanisms of cellular development[4] and the response to therapeutic interventions[5]. However, proteins, not mRNAs, are the primary targets of drugs, and protein abundance cannot necessarily be inferred directly from mRNA abundance[6–9]. An unbiased view of proteins is thus necessary to model cellular dynamics and response to environmental and therapeutic perturbations.

REAP-seq enables simultaneous measurement of proteins and mRNAs in single cells. Cells are labeled via methods similar to standard flow cytometry methods but with antibodies conjugated to DNA barcodes instead of fluorophores. This removes the limitations imposed by spectral overlap of fluorescent labels (~17) (ref. 10) or the available number of stable isotopes (~40) (ref. 11), in flow and mass cytometry. Using sequencing as a readout instead of qPCR[12], a DNA barcode of eight nucleotides provides up to 65,536 unique indices (B^n, where B = any of the four bases GATC, and n = length of the nucleotide sequence). In addition to the unique 8-bp barcode, the antibody DNA label consists of a poly (dA) sequence for priming to the cell barcode and a universal sequence for amplification (**Supplementary Figs. 1–3** and **Supplementary Discussion**). Excess unbound antibody barcodes (AbBs) are washed from the labeled cells before they are processed using

the standard 10x Genomics single-cell (sc)RNA-seq platform[3], which is a droplet-based system designed for 3′ digital counting of mRNA in thousands of single cells.

REAP-seq leverages the DNA polymerase activity of the reverse transcriptase to simultaneously extend the primed AbB with the poly(dT) cell barcode and synthesize complementary DNA from mRNA in the same reaction. Exonuclease I is then used to degrade any excess unbound single-stranded oligonucleotides from the protein double-stranded (ds) DNA (~155 bp) products to prevent crosstalk between AbBs and cell barcodes from different cells (**Supplementary Fig. 4**). Dextran sulfate was added to AbB labeling buffer to reduce non-specific binding of negatively charged DNA barcodes to the cell surface and isotype controls (Mouse IgG1, Mouse IgG2a, Mouse IgG2b, Rat IgG1, Rat IgG2a) were used to determine the threshold of background noise (**Supplementary Figs. 5** and **6**).

To initially test REAP-seq, we stained peripheral blood mononuclear cells (PBMCs) with a mixture of 45 AbBs (**Fig. 1** and **Supplementary Tables 1** and **2**) and then magnetically enriched for three populations of cells: CD3[+] T cells, CD11b[+] myeloid cells, and CD19[+] B cells (**Supplementary Fig. 7**). Cell barcodes identified in both gene and protein expression matrices were filtered for cells with a mitochondrial read rate of <20% and >250 genes expressed (3,797 CD3[+], 2,883 CD11b[+], 1,533 CD19[+] cells, and 7,271 PBMCs). We used the nonlinear dimensionality reduction method 't-distributed stochastic neighbor embedding' (t-SNE) to visualize the principal component analysis (PCA)-reduced data set in two-dimensional space[13] where the cells were color-coded by cluster (**Fig. 1a** and **Supplementary Fig. 7a**). The cells were also colored by the magnetic beads used for isolation (CD3[+], CD19[+], CD11b[+]) (**Supplementary Fig. 7b**), which showed three easily discernible purified populations of cells, and was used as a positive control to assess the sensitivity and specificity of REAP-seq mRNA and protein measurements for canonical markers of these cell types (**Supplementary Fig. 7c**). Also as a control, scRNA-seq alone was run on PBMCs to ensure the protein assay has no effect on mRNA measurements (**Supplementary Figs. 8** and **9**).

Protein and mRNA expression of canonical markers for monocytes (CD11b, CD14, CD33), B cells (CD20, CD19), T cells (CD3, CD4, CD8), and natural killer (NK) cells (CD56, CD158e1) were projected on the mRNA t-SNE plot to visualize expression across all PBMCs, and to assess the specificity and sensitivity of the protein and mRNA assays (**Fig. 1b**). For each marker, the Pearson correlation coefficient between mRNA and protein expression was calculated. The markers most highly correlated were HLA-DR ($R = 0.69$), CD20 ($R = 0.46$), and CD14 ($R = 0.51$), and these markers also had the highest levels of transcriptional expression (**Supplementary Table 3**). For CD4, the correlation between mRNA and protein was low, and we found it expressed both in monocytes and T cells, a finding we confirmed by flow cytometry, ruling out non-specific

[1]Genetics & Pharmacogenomics, Department of Translational Medicine, Merck & Co., Inc., Boston, Massachusetts, USA. [2]Informatics IT, Merck & Co., Inc., Boston, Massachusetts, USA. [3]Department of Profiling and Expression, Merck & Co., Inc., Palo Alto, California, USA. [4]Protein Sciences, Department of Biologics, Merck & Co., Inc., Boston, Massachusetts, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to V.M.P. (vanessa.peterson1@merck.com).
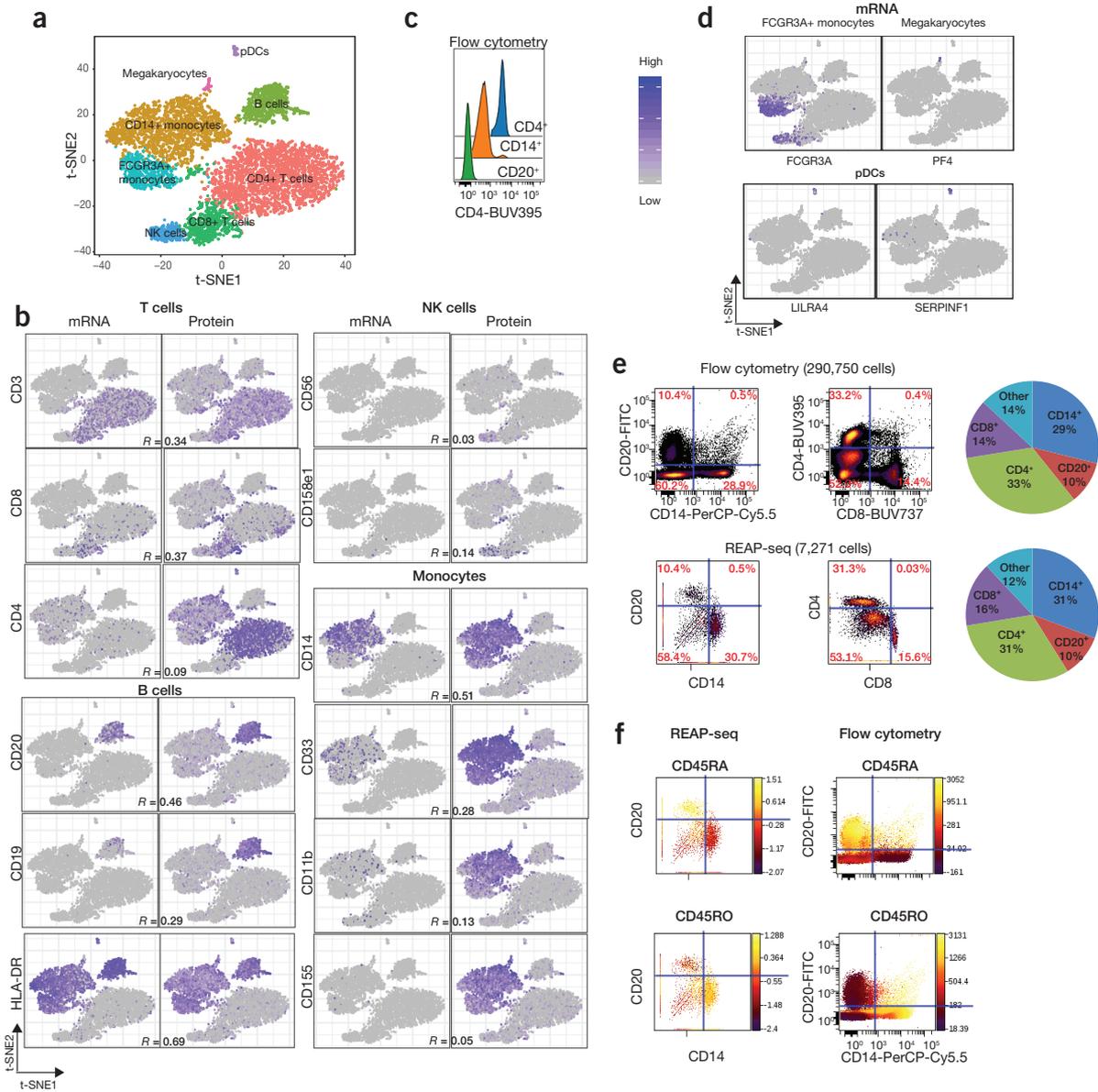
Figure 1 Benchmarking of REAP-seq on PBMCs. (**a**) PBMCs ($n = 7,271$) were processed with REAP-seq. t-SNE visualization of eight clusters identified using the top ten significant principal components across 1,664 variable genes. Cells are colored by cluster. (**b**) mRNA and protein signal for canonical markers expressed in monocytes (CD11b, CD33, CD14, CD155), B cells (CD19, CD20), T cells (CD3, CD4, CD8), and NK cells (CD56, CD158e1) were projected on the t-SNE plot from **a**. For each marker, the Pearson correlation coefficient (R) between mRNA and protein expression across 7,271 single cells is displayed. (**c**) Flow cytometry staining confirms that CD14+ monocytes have low level of CD4+ expression, which was seen in the REAP-seq t-SNE plot showing CD4 protein expression (**b**, upper left cluster in CD4 protein panel). (**d**) mRNA signal for markers expressed in plasmacytoid dendritic cells (pDCs; *LILRA4*, *SERPINF1*), FCGR3A+ monocytes (*FCGR3A*), and megakaryocytes (*PF4*) were projected on the t-SNE plot from **a**. (**e**) REAP-seq conducted on 7,271 PBMCs compared to flow cytometry on 290,750 PBMCs (from the same sample). REAP-seq protein counts were log normalized, first scaling each cell to a total of 1×10^4 molecules and then analyzed in Cytobank along with flow cytometry data. (**f**) REAP-seq protein assay enables measurement of CD45RO and CD45RA isoforms that cannot be measured with 3' RNA-seq methods. Flow cytometry data confirmed the high expression of CD45RA in CD20+ cells and CD45RO in CD14+ cells. The color of each dot in the bivariate plot represents the level of CD45RA and CD45RO expression. Yellow indicates high expression and dark red indicates low expression.

binding of the CD4 AbB (**Fig. 1c**). Markers such as CD56, CD158e1, CD19, CD33, CD11b, and CD155 all showed higher sensitivity at the protein level than mRNA. However, measuring the entire transcriptome enabled us to identify plasmacytoid dendritic cells (*LILRA4*, *SERPINF1*), megakaryoctyes (*PF4*), and FCGR3A+ monocytes (*FCGR3A*), as these markers were not included in our AbB panel (**Fig. 1d**).

We compared REAP-seq protein measurements to flow cytometry and found agreement in the relative abundances of the four major cell types identified (**Fig. 1e**). REAP-seq protein measurements of CD45RO and CD45RA (alternatively spliced isoforms of CD45) were also consistent with flow cytometry data showing high CD45RA expression in CD20+ B cells and high CD45RO expression in CD14+ monocytes (**Fig. 1f**). These isoforms are not measured by 3' mRNA sequencing, demonstrating the utility of this method to measure proteins and transcripts in a single assay.

Agonist monoclonal antibodies to CD27 (aCD27) have been shown to be useful in effectively modulating immune responses including antitumor immunity in preclinical models[14,15]. We used REAP-seq to characterize the *in vitro* activation of naive CD8+ T cells with
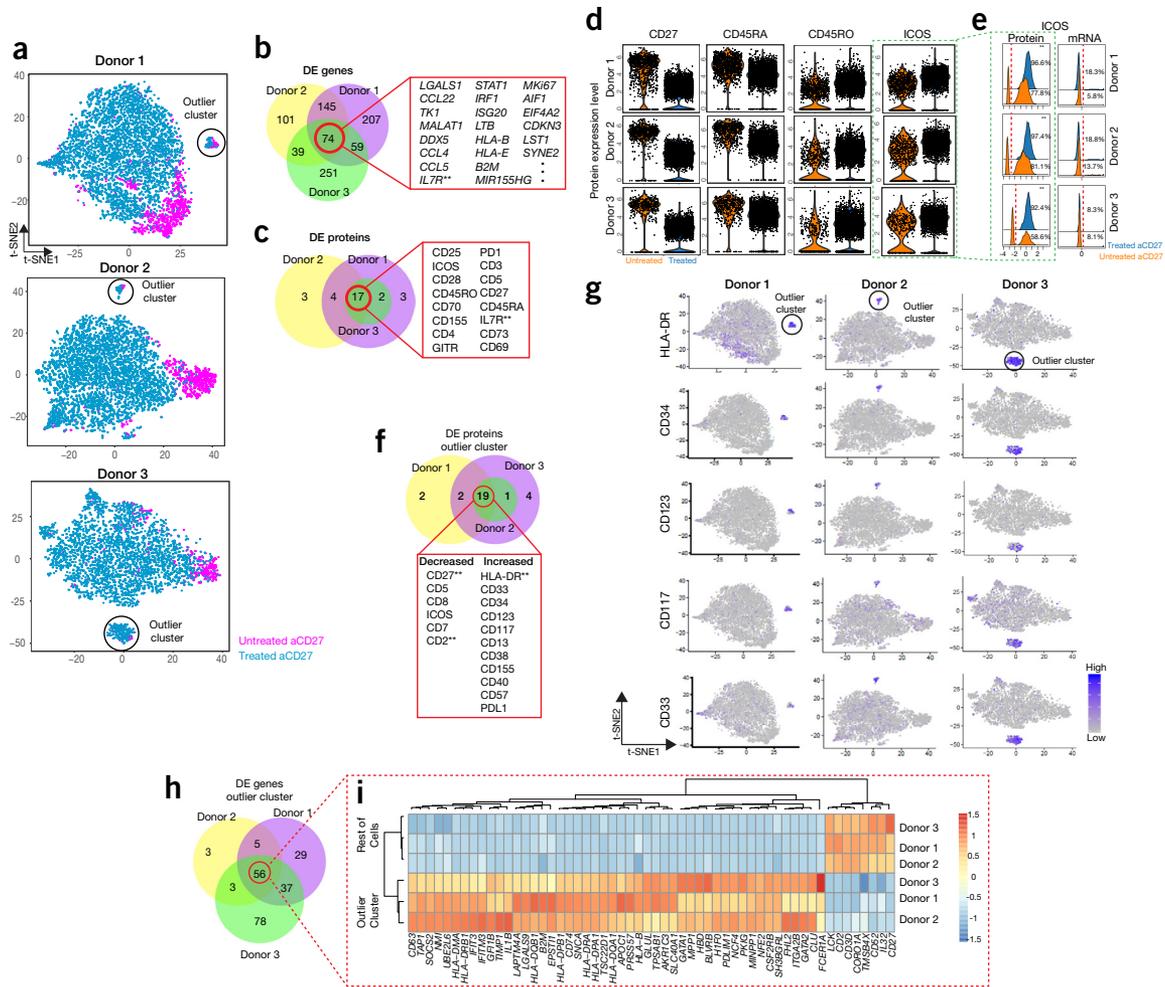
Figure 2 REAP-seq characterization of *ex vivo*-activated naive CD8[+] T cells with aCD27. (**a**) t-SNE visualization plots based on protein expression for each of the three donors. Blue dots indicate cells treated with aCD27 (donor 1: 4,246; donor 2: 4,044; donor 3: 3,550 cells) and magenta indicates cells not treated with aCD27 (donor 1: 950; donor 2: 622; donor 3: 406 cells). (**b,c**) Venn diagram showing the number of differentially expressed (DE) genes (*n* = 74) or proteins (*n* = 16, 17 AbBs, where two were different CD4 AbBs) shared across the three individual donors. Differentially expressed genes and proteins had adjusted *P*-values < 0.01 (corrected for multiple testing using the Bonferroni correction) and fold changes greater than 1.3 (threshold used for differential expression). (**d**) Violin plots showing expression distribution of CD27, CD45RO, CD45RA, and ICOS for untreated or treated aCD27 cells for each donor. Each dot represents an individual cell. Expression levels are log transformed, first scaling each cell to a total of 1×10^4 molecules. (**e**) Histograms showing ICOS protein and gene expression distribution in aCD27-treated (blue) and untreated (orange) cells. On the histogram the red dotted line shows the separation between signal and background noise and cells to the right of the line express ICOS. (**f**) Venn diagram showing the overlap of 17 differentially expressed proteins (19 AbB where three were different CD8 AbBs) in the outlier cluster compared to all other cells across all three donors (fold change > 1.5 and adjusted *P* < 0.01, corrected for multiple testing using the Bonferroni correction). (**g**) Expression of five selected upregulated proteins (HLA-DR, CD34, CD123, CD117, CD33) in the outlier cluster was projected on the t-SNE visualization plots from **a**. The outlier cluster consisted of 116 cells in donor 1 (2.2% total cells), 78 cells in donor 2 (1.7% of total cells), and 211 cells in donor 3 (5.3% total cells) (**h**) Venn diagram showing the overlap of differentially expressed genes in the outlier cluster compared to all other cells for each donor (fold ratio > 1.5, adjusted *P*-value < 0.01 corrected for multiple testing using the Bonferroni correction). (**i**) Heatmap showing the average scaled expression across all cells in the outlier cluster versus the rest of the cells for the 56 differentially expressed genes. Red indicates genes upregulated, blue indicates genes downregulated. ** indicates markers differentially expressed at both protein and mRNA level.

aCD27 and TCR stimulation. Naive CD8[+] T cells from the blood of three individual donors were treated with either anti-CD3 (aCD3) and anti-CD28 (aCD28) antibodies or aCD27, aCD3, and aCD28 antibodies (**Supplementary Fig. 10a**) and labeled with a panel of 80 AbBs (**Supplementary Table 1**). After sorting for live cells by flow cytometry, we loaded the cells onto the 10x Genomics Chromium Controller instrument.

We used REAP-seq to analyze cells treated with aCD27 (**Supplementary Table 4**). For each donor, the aCD27-treated and untreated samples were merged into a digital gene or protein expression matrix and unsupervised clustering was performed (**Supplementary**

Fig. 10b). Cells were color-coded by the different treatment conditions to visualize co-localization of cells within the t-SNE plots (**Fig. 2a** and **Supplementary Fig. 11**). Although the number of protein markers was far fewer than the number of genes measured, there was clear visual separation between cells from the different treatment conditions when only protein expression data were used for clustering. When mRNA was clustered using the smaller set of protein markers there was no visual separation between aCD27-treated and untreated cells in donors 2 and 3 (**Supplementary Fig. 11b**).

Genes and proteins differentially expressed in the aCD27-treated and untreated naive CD8[+] T cells for each donor were identified

as those having adjusted *P*-values < 0.01 and fold changes > 1.3 (**Supplementary Fig. 12** and **Supplementary Discussion**). Among all three donors, there were 74 overlapping differentially expressed genes (**Fig. 2b** and **Supplementary Fig. 13**). In this overlapping gene set, the cell proliferation marker *MKI67* was upregulated in aCD27-treated cells, agreeing with previous findings where T cells co-activated by TCR and CD27 were shown to induce a remarkable level of proliferation[16]. This unbiased scRNA-seq approach enabled us to see differential expression in unexpected genes such as *MALAT1*, *SYNE2*, *LST1*, and *TK1* which would not have been included in a pre-specified antibody panel. Also, *MIR155HG*, a non-coding RNA that cannot be detected at the protein level, showed decreased expression levels in cells treated with aCD27.

REAP-seq protein analysis identified 16 differentially expressed proteins overlapping across the three donors (**Fig. 2c** and **Supplementary Figs. 14** and **15a**) when cells were treated with aCD27. Interleukin (IL)-7R was consistently downregulated at both the protein and mRNA level (**Supplementary Figs. 16**, **17** and **18a**) and there was a decrease in CD27 expression, which was confirmed by flow cytometry (**Supplementary Fig. 19**). We observed the loss of naive T-cell marker CD45RA expression and gain-of-effector memory T-cell marker CD45RO expression (**Fig. 2d**). Also, we found that a substantial number of CD8+ T cells treated with aCD27 cells expressed CD4 and CD25, a finding we confirmed by flow cytometry (**Supplementary Figs. 14b**, **18b**, and **20**).

The levels of mRNA and protein did not always correlate (**Supplementary Figs. 16** and **21**), and protein quantification was more sensitive for markers with lower-abundance mRNA transcripts such as ICOS. ICOS is an immune checkpoint protein that has been shown to increase in cell surface expression upon costimulation with aCD27 and T-cell receptor (TCR) signaling[16]. We found an increase in ICOS expression at the protein level but not at the transcriptional level (**Fig. 2e**). Increased sensitivity in protein measurements may be due to longer protein half-lives and higher absolute protein abundance compared to mRNA[9]. In addition, the protein assay may benefit from signal amplification owing to more than one DNA barcode conjugated to each antibody (an average of three DNA barcodes per antibody).

To demonstrate how REAP-seq can be used to characterize unknown cellular populations, we used REAP-seq to identify a small population of cells present with our enriched naive CD8+ T cells (outlier clusters circled in **Fig. 2a**). In all three donors there were 17 shared differentially expressed proteins between this outlier cluster and the rest of the cells (**Fig. 2f**), with a pattern suggestive of common myeloid progenitors (CMPs)[17] (**Fig. 2g**), having high relative expression of CD34, CD38, CD123, CD117, CD13, CD33, and HLA-DR.

At the mRNA level there were 56 overlapping genes that were differentially expressed in the outlier cluster in all three donors (**Fig. 2h,i**), and three of these (*HLA-DRA*, *CD27*, and *CD2*) were also differentially expressed at the protein level (**Supplementary Fig. 22**). Together, the differentially expressed genes and proteins showed enrichment in the transcriptional regulation of megakaryopoiesis (false-discovery rate, FDR = 1.2×10^{-5}, MetaCore) (**Supplementary Fig. 23**), the process by which mature megakaryocytes (MKs) develop from the common myeloid progenitor.

An approach similar to REAP-seq was recently published that uses DNA-barcoded antibodies together with high-throughput scRNA-seq[18]. A difference between the methods is how the DNA barcode is conjugated to the antibody. REAP-seq minimizes steric hindrance and potential crosstalk by using unidirectional chemistry that creates a small, stable, covalent bond between the antibody and aminated DNA barcode, while the other method, CITE-seq[18] conjugates, on average, two bulky streptavidin (~50 kDa each) to each antibody (~150 kDa) before non-covalently binding to biotinylated DNA barcodes. Minimizing steric hindrance is important in the scalability of the protein assay and in the future extension of this approach to intracellular labeling. Here we demonstrate the unprecedented scalability of REAP-seq by conjugating up to 82 antibodies to unique DNA barcodes.

REAP-seq is readily adaptable to other microfluidic or micro-/nano-well platforms[1,19], and could also be applied to bulk samples to simultaneously measure both protein and RNA. Recent work demonstrates that scRNA-seq on fixed cells is possible[20], and we anticipate REAP-seq being extended to measure intracellular signaling pathways, B- or T-cell receptor sequencing, and other genomic DNA readouts such as point mutations and copy-number variations. REAP-seq could also be coupled with multiple types of perturbations such as small molecules, RNA interference, CRISPR, and other gene editing techniques to provide mechanistic insights into cellular phenotypes in relation to disease and treatment response.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS
V.M.P. and J.A.K. conceived the method. V.M.P., J.A.K, J.W., T.K.M., and S.S. designed experiments. V.M.P, J.W., N.K., D.C.W., L.L., and R.M. conducted experiments. K.X.Z. wrote analysis software. K.X.Z., V.M.P, N.K., J.A.K., R.M., and S.S. analyzed the data. V.M.P., J.A.K., and K.X.Z. wrote the manuscript with input from all authors.

1. Macosko, E.Z. *et al. Cell* **161**, 1202–1214 (2015).
2. Zheng, G.X. *et al. Nat. Commun.* **8**, 14049 (2017).
3. Villani, A.-C. et al. *Science* **356**, eaah4573 (2017).
4. Rizvi, A.H. et al. *Nat. Biotechnol.* **35**, 551–560 (2017).
5. Tirosh, I. *et al. Science* **352**, 189–196 (2016).
6. Liu, Y., Beyer, A. & Aebersold, R. *Cell* **165**, 535–550 (2016).
7. Darmanis, S. *et al. Cell Reports* **14**, 380–389 (2016).
8. Battle, A. *et al. Science* **347**, 664–667 (2015).
9. Schwanhäusser, B. *et al. Nature* **473**, 337–342 (2011).
10. Perfetto, S.P., Chattopadhyay, P.K. & Roederer, M. *Nat. Rev. Immunol.* **4**, 648–655 (2004).
11. Bendall, S.C. *et al. Science* **332**, 687–696 (2011).
12. Genshaft, A.S. *et al. Genome Biol.* **17**, 188 (2016).
13. van der Maaten, L. & Hinton, G. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
14. French, R.R. *et al. Blood* **109**, 4810–4815 (2007).
15. Roberts, D.J. *et al. J. Immunother.* **33**, 769–779 (2010).
16. Ramakrishna, V. *et al. J. Immunother. Cancer* **3**, 37 (2015).
17. Manz, M.G., Miyamoto, T., Akashi, K. & Weissman, I.L. *Proc. Natl. Acad. Sci. USA* **99**, 11872–11877 (2002).
18. Stoeckius, M. *et al. Nat. Methods* http://dx.doi.org/10.1038/nmeth.4380 (2017).
19. Gierahn, T.M. *et al. Nat. Methods* **14**, 395–398 (2017).
20. Alles, J. *et al. BMC Biol.* **15**, 44 (2017).

ONLINE METHODS

Antibody DNA barcode conjugations. Antibodies were conjugated to oligonucleotides (65–66 bp) using the Thunder-Link PLUS Oligo Conjugation System (Innova Biosciences) following manufacturer's protocol to target an average antibody/oligo ratio of 1:3. Before conjugation, antibodies were normalized to ~1 mg/ml. If antibody concentration was <1 mg/ml, it was raised to 1 mg/ml with Amicon Ultra 0.5 ml centrifugal filter units with Ultracel-100 membrane (Millipore) according to manufacturer's instructions. Oligonucleotides were synthesized by Integrated DNA Technologies (IDT) and were 5′ aminated and high-performance liquid chromatography (HPLC)-purified. Each oligonucleotide consisted of three parts: 1) 33-bp Nextera Read 1 sequence that was used as a primer for amplification and sequencing, 2) a unique 8-bp antibody barcode, and 3) 24- to 25-bp poly (dA) sequence that binds to the poly(dT) primer on the bead (10x Genomics) containing the cell barcode (**Supplementary Fig. 2**). A full list of monoclonal antibody clones and DNA barcodes can be found in **Supplementary Table 1**. 45 AbBs were used in the initial PBMC experiment and 80 AbBs were used in the aCD27 *ex vivo* costimulation assay. Before each experiment the antibodies were pooled together before adding to the cells. Isotype controls (Mouse IgG1, Mouse IgG2a, Mouse IgG2b, Rat IgG1, and Rat IgG2a) were included in the AbB panel and used as negative controls to determine the threshold for non-specific binding (**Supplementary Figs. 5 and 6**).

Initial proteomic assay validation experiments with beads. For mixing experiments, Quantum Simply Cellular (QSC) anti-mouse IgG beads (Bangs Laboratories, Inc.) were used to bind to mouse IgG isotype AbBs. For the first test experiment, 2 μl of AbB (CD13 or CD70) was added to 20 μl of beads in separate tubes (**Supplementary Fig. 4a**). For the second experiment, 2 μl of AbB (CD8a, TIGIT, CD9, CD27, CD28, CD40, NT5E, CD127, OX40, and Mouse IgG1 isotype control) was added to 20 μl of beads in separate tubes (**Supplementary Fig. 4b**). Excess unbound AbBs were removed from the beads by washing. The beads were then pooled together before running the proteomic part of the REAP-seq assay as described for cells. Cell barcodes with <50 counts were removed from analysis and counts were normalized by the total number of counts for each bead. A T-distributed stochastic neighbor embedding (t-SNE) algorithm[13] was used to visualize resulting clusters.

REAP-seq PBMC preparation. Cryopreserved PBMCs from a healthy donor (AllCells, PB003F, LotA5051) were thawed at 37 °C for 3 min and resuspended in 10 ml of PBS with 1% Bovine Serum Albumin (BSA) (Miltenyi Biotec) added slowly in a dropwise manner. PBMCs were split for three conditions: 1) PBMCs not labeled with AbBs as a control, 2) PBMCs labeled with the AbBs in PBS with 1% BSA and 3) PBMCs first blocked with dextran sulfate before labeling with AbBs. For the third condition, 0.2 mg/ml of dextran sulfate (DS) in PBS with 1% BSA was added to cells for 10 min on ice to reduce non-specific binding of the negatively charged DNA barcodes to the cellular membrane (**Supplementary Fig. 5**). The AbB mix (n = 45) consisted of 1 μl per antibody (total of 45 μL) and the final reaction volume was 200 μL (0.1 mg/ml DS in PBS with 1% BSA).

Cells were labeled with the AbB mix for 15 min on ice. After staining, 12 ml of the wash buffer (PBS with 1% BSA or 0.1 mg/ml DS in PBS) was added to the cells. The samples were spun down at 300*g* for 4 min at 4 °C, and the supernatant was removed with a new 10-ml serological pipette for each sample to prevent cross-contamination. The supernatant was carefully removed, leaving volume of ~50 μl remaining to avoid disturbing the cell pellet. The washing step was repeated three times, and a final wash was conducted in PBS with 1% BSA for all samples.

After the last wash, PBMCs blocked with dextran sulfate before labeling with AbBs were split into four conditions: 1) no magnetic enrichment, 2) CD3+ magnetic enrichment (Miltenyi, 130-050-101), 3) CD11b+ magnetic enrichment (Miltenyi, 130-049-601), and 4) CD19+ magnetic enrichment (Miltenyi, 130-050-301). Miltenyi microbeads were used to positively select for CD3+ T cells, CD11b+ myeloid cells, or CD19+ B cells using MS columns (Miltenyi, 130-042-201) following manufacturer's instructions.

Flow cytometry PBMC immunophenotyping. Frozen PBMCs (AllCells, PB003F, LotA5051 (same donor as for REAP-seq)) were quick-thawed in a 37 °C water bath, washed with 30 mL pre-warmed RPMI Complete Medium (RPMI 1640 + 10% FBS + Pen/Strep + ʟ-glutamine), and washed again with 25 mL Wash Buffer (1× DPBS + 2% FBS + 1 mM EDTA). Viability was assessed using the Vi-Cell Cell Counter (Beckman Coulter) using default settings. Cells were washed a second time in 25 mL Wash Buffer, and the cell pellet was resuspended in a viability stain solution (5.4 mL 1× DPBS + 13.5 μL Fixable Viability Dye eFluor506 (eBioscience)). After 30 min incubation on ice, the cells were washed in 25 mL wash buffer. One million viable cells were aliquoted into microfuge tubes for cell surface staining. Fc-receptors were blocked in a solution containing 45 μL wash buffer, 5 μL Human TruStain FcX (BioLegend), and 2% normal mouse serum (Jackson ImmunoResearch). A cocktail of cell-surface-staining antibodies (**Supplementary Table 5**) was added directly to cells in blocking buffer. Cells were stained for 30 min on ice and then washed with wash buffer. All samples were then fixed with 100 μL freshly diluted 1.6% formaldehyde (Electron Microscopy Sciences) in DPBS per tube for 15 min at room temperature. The cells were washed with 1 mL wash buffer, resuspended in 500 μL wash buffer, and acquired on a LSR II flow cytometer (BD Biosciences). Cytobank (v5.6.1) was used for data analysis[21]. All flow cytometry data shown were initially gated on live cells (**Supplementary Fig. 24a**).

aCD27 *ex vivo* CD8+ T-cell activation assay. Human buffy coats from three healthy volunteers were obtained from the Stanford Blood Center. All blood samples were collected from voluntary donors after obtaining informed consent, in accordance with institutional review board protocols and in compliance with state and federal regulations. The method used for purification of naive CD8+ T cells (which make up typically ~0.4–2.6% leukocytes) involved two negative-selection enrichment steps, where the first enriched for CD8+ cells and the second enriched for naive CD8+ T cells, using a cocktail of antibodies targeting lineage-associated (Lin) antigens. In the first enrichment step, human CD8+ T cells were obtained from buffy coats using RosetteSep Human CD8+ T Cell Enrichment Cocktail (StemCell Technologies) via negative selection according to manufacturer's instructions. Briefly, cells were stained with the RosetteSep cocktail of antibodies and centrifuged over a buoyant density medium Ficoll-Paque Plus (GE, Healthcare). The enriched CD8+ T cells were removed from the plasma and density medium interface and lysed with ACK lysis buffer (Thermo Fisher). In the second enrichment step, naive CD8+ T cells were then negatively selected for using a Human Naive CD8+ T Cell Enrichment Set (BD Biosciences) containing biotinylated monoclonal antibodies (CD4, CD11b, CD19, CD16, CD41a, CD45RO, gδ-TCR, Glycophorin A) to remove CD4+ T cells, monocytes, B cells, NK cells, granulocytes, platelets, memory T cells, γ/δ T cells, and erythroid cells. Cells were frozen in liquid nitrogen for future use. It has been shown that rare cell types such as hematopoietic stem cells (HSCs) which do not express Lin antigens can be enriched by 50- to 200-fold from peripheral blood using cocktails of antibodies to remove Lin+ cells[22,23]. Since our purification method used a double negative selection, it is plausible that a rare population of cells such as common myeloid progenitors could be enriched (>200-fold) and detected at ~2–5% of total cells.

Humanized anti-human CD27 antibody (MSD) was bound to Dynabeads (M-450 Epoxy, Life Technologies) by mixing overnight with prewashed Dynabeads at room temperature. BSA was added to block remaining free binding sites on the beads. Beads were washed and resuspended to ~1.5 × 10^5 beads per μL in 0.1 M sodium phosphate buffer.

For the activation assay to determine an aCD27-specific gene and protein signature, CD8+ naive T cells were thawed and resuspended to 7.5 × 10^5 cells/ml in DMEM-F12 (Gibco), 5% heat-inactivated human serum (Sigma), 50 μM 2-mercaptoethanol (Sigma), 100 U/ml penicillin, and 100 μg/ml streptomycin (Lonza). 1.5 × 10^5 cells were cultured in the presence of a sub-optimal dose of aCD3 (0.025–0.15 mg/ml clone OKT3, Biolegend) and aCD28 (1 mg/ml clone 15E8, Millipore) with or without 1 μL of bead-bound anti-CD27 in a flat-bottom 96-well plate for 3 d in a 37 °C, 5% CO2 incubator. Following stimulation, an aliquot of cells was washed and frozen in 90% FBS (Hyclone) and 10% DMSO (Sigma) for downstream REAP-seq processing. Another aliquot of cells was stained with a fixable viability dye (eFluor506, eBioscience) for 30 min at 4 °C. Excess dye was removed by washing and the cells were blocked with TruStain FcX (Biolegend). The cells were then incubated and stained with a cocktail of antibodies (**Supplementary Table 5**) and analyzed by flow cytometry (BD Biosciences LSRFortessa) for comparison to REAP-seq proteomic measurements. All flow cytometry data shown were initially gated on live cells (**Supplementary Fig. 24b**).

aCD27-treated and untreated cells from donors 1, 2, and 3 were thawed at 37 °C for 3 min and resuspended by adding 10 ml of PBS with 1% BSA slowly

in a dropwise manner. The cells were blocked with 0.2 mg/ml of dextran sulfate in PBS with 1% BSA for 10 min on ice to reduce non-specific binding of the negatively charged oligonucleotides to the cellular membrane. The AbB mix (**Supplementary Table 1**) consisted of 1 µl of each antibody (80 antibodies, total of 80 µl) and the labeling was performed as previously described above for the REAP-seq PBMC experiment.

A Sony SH800 sorter (Sony Biotechnology) was used to enrich for live cells stained with the LIVE/DEAD Viability/Cytotoxicity kit (Thermo Fisher). Cells were resuspended in 300 µl of calcein AM (0.2 µM) and ethidium homodimer-1 (0.2 µM) in PBS with 1% BSA to discriminate live from dead cells where viability ranged from ~50–75%. REAP-seq was conducted on cells treated with aCD27 (donor 1, 4,246; donor 2, 4,044; donor 3, 3,550 cells) and without aCD27 (donor 1, 950; donor 2, 622; donor 3, 406). Fewer aCD27 untreated cells were analyzed due to lower (~2–3×) initial starting cell numbers and also loss of a sample due to an emulsion break (donor 2 and 3, ~2× decrease).

REAP-seq assay. Cell number and viability were measured using a Countess II Automated Cell Counter (ThermoFisher). Cells were loaded onto the Chromium Controller (10x Genomics) targeting 10,000 cells per lane; for samples with <10,000 cells, the entire sample was loaded. The 10x Genomics v1 single cell 3′ RNA-seq reagent kit protocol (10x Genomics) was used to process samples according to manufacturer's protocol. Following the reverse transcription and the silane bead cleanup step there is a 0.6× SPRI size-selection step (Beckman Coulter, B23318) where ~80 µL of supernatant (50 µl sample + 30 µl SPRI) is discarded. This supernatant contains the shorter (~155 bp) cell barcoded AbBs (CB-AbBs) and is retained for the proteomic part of this assay (**Supplementary Fig. 3**).

SPRI beads (45 µl) were added to the supernatant containing the CB-AbBs (~80 µl) and incubated at room temperature for 5 min. The SPRI beads were washed twice with 80% EtOH and CB-AbBs were eluted from the beads into 34 µl of nuclease-free water (Ambion). An exonuclease treatment was then conducted to remove any excess unbound single-stranded cell barcodes (~114 bp) or AbB (~65-66 bp). To 34 µl of eluted cDNA, 4 µl of 10× Exonuclease 1 Reaction buffer and 2 µl of Exonuclease I (New England BioLabs) was added for a final reaction concentration of 1 U/µl Exonuclease I. The sample was incubated at 37 °C for 60 min followed by denaturation at 80 °C for 20 min. After the denaturation step, 60 µL of SPRI beads (1.5× v:v) was added and incubated at room temperature for 5 min. The SPRI beads were washed twice with 200 µl of 80% EtOH, and the CB-AbBs were eluted from the beads into 33 µl of buffer EB (Qiagen).

The CB-AbBs were amplified with KAPA HiFi HotStart Polymerase (Kapa Biosystems) in a 50 µl reaction volume following manufacturer's protocol. During amplification, the sample index and P5 adaptor were added to the CB-AbB using three primers; P7, P5, and P5-sample index-Part of Nextera Read 1 (see **Supplementary Table 6** for sequence). Each primer was used at the final reaction concentration of 0.3 µM. The samples were cycled as follows: 95 °C 3 min, 18 cycles of: 98 °C 20 s, 62 °C 30 s, and 72 °C 30 s; then a final extension step of 72 °C for 4 min. The final library was cleaned up by adding 1.0× volume of SPRI beads (50 µL). The SPRI beads were washed twice with 200 µl of 80% EtOH and CB-AbBs were eluted from the beads into 33 µl of buffer EB (Qiagen).

Samples were run on the Agilent Bioanalyzer High Sensitivity DNA chip (Agilent Technologies) to confirm the desired protein library product of ~185 bp. If PCR impurities were present, the desired fragment was isolated using a BluePippin (Sage Science) in a broad range selection mode from 160 to 215 bp using a 3% agarose gel cassette and the Q2 marker as an internal standard (Sage Science). A Qubit dsDNA HS assay (ThermoFisher Scientific) was used to determine DNA concentration for sequencing.

Libraries from the proteomic assay were sequenced on an Illumina HiSeq 2500 using a HiSeq Rapid SBS V2 50 cycle kit (Illumina) and a HiSeq Rapid PE Cluster kit (Illumina). Libraries were diluted to a final concentration of 10 pM where the library consisted of 60% protein library, 30% PhiX v3 (Illumina), and 10% custom oligonucleotide (P5-[N_8]-Nextera Read 1- [N_{26}]- TruSeq Read 2-[N_{26}]- P7, IDT) to increase library diversity. Custom sequencing primers were used for read 1 and the i7 index read (**Supplementary Table 7**) following Illumina guidelines for custom primers. Hybridization buffer, HT1 (Illumina), was used to dilute each custom primer to a final concentration of 0.5 µM and a final volume of 5 ml. Sequencing reads were performed in the following order: 1) read 1 was 26 cycles where cycles 1–14 sequenced the cell barcode and cycles 15–26 were

included because of the 26-cycle minimum read 1 length; i7 was 8 cycles for the antibody barcode, i5 was 8 cycles for the sample index, and read 2 was 10 cycles for the unique molecular index (UMI).

Single-cell RNA sequencing computational analyses. Raw sequencing data were processed using the 10x Genomics Cell Ranger pipeline (version 1.3). First, cellranger mkfastq demultiplexed libraries based on sample indices and converted the barcode and read data to FASTQ files. Second, cellranger count took FASTQ files and performed alignment (UCSC hg19, STAR)[24], filtering, and unique molecular identifier (UMI) counting. Lastly, cellranger aggr took outputs from multiple runs of cellranger count, and normalized these runs to the same sequencing depth and recomputed the gene-barcode matrices and analysis on the combined data (**Supplementary Tables 2** and **4**). Then we selected cells that had a mitochondrial read rate < 20% and expressed > 250 genes (PBMCs) or >500 genes (aCD27 assay) to filter out low-quality cells from our data set. Count data of each cell was first scaled to a total of 1×10^4 molecules and then log transformed using the R package Seurat (v1.4)[25]. Normalized and scaled data were used for all downstream analyses. Specifically, we performed principal component analysis (PCA) using a normalized matrix of genes with high dispersion. We defined highly variable genes as the mean normalized counts > 0.1 and the s.d. of normalized counts > 2. In order to cluster cells into groups, we used a graph-based algorithm to detect clusters based on the similarities of top principal components between cells in Seurat. The graph-based cluster algorithm consists of two steps. The first step is to calculate the connectivity across cells by computing the pair-wise eigenvalue-weighted Euclidean distance with significant PC scores of any pair of cells. The second step is to apply a "community" detection algorithm[26] to partition the graph derived from the first step into graph "modules" based on connectivity. This algorithm is implemented in the Seurat package [v1.4]. We selected the top significant principal components and then employed the t-distributed stochastic neighbor embedding (t-SNE) algorithm to visualize resulting clusters. For the differential expression analysis, we used the FindAllMarkers function in Seurat to identify genes differentially expressed between conditions (aCD27 treated vs. untreated, fold change >1.3) or clusters (outlier cluster vs. rest of cells, fold change >1.5). The *P*-values were corrected for multiple testing using the Bonferroni correction and the threshold used for statistical significance was an adjusted *P*-value < 0.01.

Single-cell protein sequencing computational analyses. BCL files were converted to FASTQ files using Illumina bcl2fastq (version 2.17.14). In the proteomic assay read 1 and read 2 are switched relative to 10x Genomics standard single-cell RNA-seq pipeline (v1 chemistry), so the "R1" and "R2" names in FASTQ files were switched to "R2" and "R1", respectively, in order to be compatible with the cellranger count pipeline (version 1.3). A post-sorted genome BAM file containing the cell barcode (14 bp), UMI (10 bp), and antibody barcode (8 bp) was generated using the cellranger count pipeline. The BAM file was converted to a SAM file using SAMtools (version 1.2)[27].

We developed an aligner called REAP-seq (RS) Aligner to process the SAM files containing the antibody barcode (8 bp) sequences. First, an AbB dictionary file was created that associates antibody names to the unique 8-bp DNA sequences that were conjugated to each antibody. Second, a hash table of cell barcodes was generated to parse the cell barcode information. Reads were grouped by their cell barcodes and UMIs were used to avoid double counting sequence reads that arose from the same AbB molecule. Third, the reads were aligned to the antibody dictionary using a Hamming distance <1. Finally, a protein digital expression matrix was generated that contained UMI-corrected read counts for each antibody and cell barcode. In order to filter out the cells representing background noise, we used the inflection point of detected cell numbers as a function of cumulative read counts as a cutoff (**Supplementary Tables 2** and **4**). For each sample, we generated a protein expression matrix for cells also identified in the single-cell RNA-seq data. The protein expression matrices were log normalized, first scaling each cell to a total of 1×10^4 molecules using Seurat and then PCA was performed on the normalized data. The t-SNE plots were generated to visualize the clusters (**Fig. 2a** and **Supplementary Fig. 10b**). For the differential expression analysis, we used the FindAllMarkers function in Seurat to identify proteins differentially expressed between conditions (aCD27-treated vs. untreated, fold change >1.3) or clusters (outlier cluster vs. rest of cells, fold change >1.5). The *P*-values were corrected for multiple testing using the

Bonferroni correction and the threshold used for statistical significance was an adjusted P-value < 0.01.

The normalized protein expression matrix was also converted to an FCS file using the CsvToFcs module in GenePattern (Broad Institute) and uploaded into Cytobank for analysis[21]. Cytobank software (v5.6.1) was used to generate histograms and bivariate dot plots with the proteomic data from REAP-seq as is performed with flow cytometry data.

Comparison of single-cell RNA-seq data to protein data. To compare single-cell proteomic and transcriptomic expression, we clustered mRNA or protein data and then projected the mRNA and protein single-cell expression data onto the preserved clustering structure in the t-SNE plot (**Figs. 1b** and **2g**, and **Supplementary Figs. 7c** and **18a**). For each protein marker in **Supplementary Fig. 7c**, we calculated the Pearson correlation coefficient (PCC) between the mRNA and protein data across all cells. and generated a table containing information on raw UMI counts data for genes and protein markers (**Supplementary Table 3**).

Gene set enrichment analysis. Gene set enrichment analysis was performed on differentially expressed genes or proteins through the use of Ingenuity Pathways Analysis (Ingenuity Systems, version 3355992), GO enrichment analysis[28,29], and MetaCore Pathway analysis (Clarivate Analytics, version 6.31, **Supplementary Discussion**). Enriched canonical pathways were defined as significant if adjusted P-values were < 0.05. Comparative analysis of differentially expressed proteins and genes by Gene Set Enrichment Analysis (GSEA)[30,31] was conducted. Differential gene expression between aCD27-treated and untreated cells was used to generate a rank order gene list, which was compared to a rank order protein list (abs(log fold change) > 0.2 and adjusted P-value < 0.05). Normalized enrichment scores (NES) and false-discovery rate (FDR) q-values were calculated for each donor (**Supplementary Fig. 17**).

Bulk RNA-seq analysis. In addition to the Cell Ranger pipeline, RNA-seq FASTQ files generated from the PBMC assay were processed using the customized bulk RNA-seq data analysis pipeline in OmicSoft ArraySuite, version 9. Specifically, raw reads were first filtered based on quality control (QC) and then aligned to the reference genome (human genome Ensembl GRCh38) using OSA[32]. After alignment, gene expression levels (raw read counts and FPKMs) were quantified by the RSEM algorithm[33] with the human gene model Ensembl.R82. Gene counts were normalized by total counts and scaled by a factor of 1×10^6. The top 5,000 highest expressed genes were correlated and the coefficient of variation (R^2) was calculated between different treatment conditions 1) AbB versus No AbB and 2) AbB with blocking buffer versus AbB without blocking buffer (**Supplementary Fig. 9**).

Biacore characterization of CD69 AbB. A Series S CM5 Chip (GE Healthcare) was immobilized via an amine coupling kit (GE Healthcare) to >8,000 RU with a rabbit anti-mouse Fc capture polyclonal antibody (GE Healthcare, BR100838). Antibodies anti-CD69 (FN50), anti-CD69 + DNA barcode (FNF50), and Mouse IgG1 isotype control (MOPC-21) were captured to ~130–160 RU on the active flowcells (Fc-2,3,4) and the reference flowcell (Fc-1) was left blank (**Supplementary Fig. 15c–e**). Binding stability to either human CD69-his recombinant protein (R&D Systems) or recombinant human CD47-his (R&D Systems),

diluted to 588 nM and 1,000 nM respectively, was measured and compared. All of the reagents were prepared in 1× HBS-EP+ (GE Healthcare) running buffer and the binding measurements were performed on a Biacore T200 at 25 °C.

Blocking buffer experiments on bulk PBMCs. PBMCs were blocked with either 1) sheared DNA salmon sperm (1 mg/ml, Thermo Fisher), 2) dextran sulfate (0.2 mg/ml, Sigma Aldrich), or 3) polyanionic competitor (1 μM, SomaLogic) in 1% BSA in PBS for 10 min on ice (**Supplementary Fig. 5a**). A cocktail of 28 AbBs consisting of 1 μL per antibody (total of 28 μL) was added to each sample and the final reaction buffer included either 1) sheared DNA salmon sperm (0.5 mg/ml), 2) dextran sulfate (0.1 mg/ml), or 3) polyanionic competitor (0.5 μM, SomaLogic) in 1% BSA. Cells were labeled with the AbB mix for 15 min on ice. Excess unbound AbBs were washed from cells as previously described. Cells were counted and 1,000 cells were used in the reverse transcriptase step for each condition. SMART-seq v4 reverse transcriptase (Clontech) was used following manufacturer's guidelines but substituting a custom primer (P7-cell barcode-Read-UMI-poly(dT), **Supplementary Table 8**) for the 3′ SMART-Seq CDS Primer II A. After the RT step, all downstream library preparation steps for the REAP-seq proteomic assay were conducted as previously described.

Flow cytometry validation of aCD27 drug blocking anti-CD27 monoclonal antibody on AbB panel. Rhesus whole blood was pre-incubated with unlabeled aCD27 drug (MSD) at varying concentrations; 10 μg/mL, 0.25 μg/mL, or 0 μg/ml for 30 min. Next, a cocktail of cell surface-staining antibodies (**Supplementary Table 5**) including either anti-CD27 monoclonal antibody labeled with APC (clone M-T271, BioLegend) or a mouse IgG1 isotype control labeled with APC were added to the blood and incubated for 30 min on ice. Next, cells were lysed with 2 mL ACK lysis buffer for 5 min at room temperature. Cells were washed with PBS and incubated with a viability stain solution (eFluor506, eBioscience) for 30 min on ice. All samples were then fixed with 1% paraformaldehyde for 15 min at room temperature. The cells were washed and analyzed by flow cytometry (LSR II flow cytometer, BD Biosciences). All flow cytometry data shown were gated on live cells→singlets→PBMCs→CD14⁻/CD20⁻→CD3+ T cells (**Supplementary Fig. 19a**).

Statistical analysis. A **Life Sciences Reporting Summary** is available.

Data availability. All data generated in this project have been deposited to the Gene Expression Omnibus (GEO) with the accession code GSE100501. Custom analysis code is available at https://github.com/kelvinxz/RS_aligner.

21. Kotecha, N., Krutzik, P.O. & Irish, J.M. *Curr. Protoc. Cytom.* **53**, 10.17 (2010).
22. Wognum, A.W., Eaves, A.C. & Thomas, T.E. *Arch. Med. Res.* **34**, 461–475 (2003).
23. Thomas, T.E., Miller, C.L. & Eaves, C.J. *Methods* **17**, 202–218 (1999).
24. Dobin, A. *et al. Bioinformatics* **29**, 15–21 (2013).
25. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F. & Regev, A. *Nat. Biotechnol.* **33**, 495–502 (2015).
26. Waltman, L. & Jan van Eck, N. *Eur. Phys. J. B* **86**, 471 (2013).
27. Li, H. *et al. Bioinformatics* **25**, 2078–2079 (2009).
28. Ashburner, M. *et al.*; The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
29. Gene Ontology Consortium. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).
30. Subramanian, A. *et al. Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
31. Mootha, V.K. *et al. Nat. Genet.* **34**, 267–273 (2003).
32. Hu, J., Ge, H., Newman, M. & Liu, K. *Bioinformatics* **28**, 1933–1934 (2012).
33. Li, B. & Dewey, C.N. *BMC Bioinformatics* **12**, 323 (2011).