

Quantification of extracellular proteins, protein complexes and mRNAs in single cells by proximity sequencing

Received: 18 January 2022

Accepted: 13 October 2022

Published online: 01 December 2022

 Check for updates

Luke Vistain^{1,2,6}, Hoang Van Phan^{1,2,6}, Bijentimala Keisham^{1,2},
Christian Jordi³, Mengjie Chen^{4,5}, Sai T. Reddy³ & Savaş Tay^{1,2} ✉

We present proximity sequencing (Prox-seq) for simultaneous measurement of proteins, protein complexes and mRNAs in thousands of single cells. Prox-seq combines proximity ligation assay with single-cell sequencing to measure proteins and their complexes from all pairwise combinations of targeted proteins, providing quadratically scaled multiplexing. We validate Prox-seq and analyze a mixture of T cells and B cells to show that it accurately identifies these cell types and detects well-known protein complexes. Next, by studying human peripheral blood mononuclear cells, we discover that naïve CD8⁺ T cells display the protein complex CD8–CD9. Finally, we study protein interactions during Toll-like receptor (TLR) signaling in human macrophages. We observe the formation of signal-specific protein complexes, find CD36 co-receptor activity and additive signal integration under lipopolysaccharide (TLR4) and Pam2CSK4 (TLR2) stimulation, and show that quantification of protein complexes identifies signaling inputs received by macrophages. Prox-seq provides access to an untapped measurement modality for single-cell phenotyping and can discover uncharacterized protein interactions in different cell types.

Single-cell measurements have expanded our understanding of many aspects of cellular function, such as enabling identification of rare cell subsets, tracking transient cellular states, and incorporating noise and variability into our understanding of cellular phenotypes^{1–3}. These phenotypes are emergent properties of both biomolecules and their interactions. Many biological functions such as signaling, differentiation, development and cellular decision-making are driven by changes in the arrangement and interaction of protein molecules. Particularly, signaling is primarily mediated by the formation and dissociation of protein complexes, and thus cannot be studied from mRNA expression or protein expression alone. Therefore, the ability to measure individual

proteins and their complexes at the single-cell level is among the most informative approaches for understanding cellular function. Despite the apparent value, there are major hurdles in performing highly multiplexed measurements of single-cell proteins and their complexes, because the number of pairwise complexes a protein can form scales quadratically with the number of proteins that is being measured. This demands a method that can encode a large number of outputs, as each measurement must enable identification of both the protein of interest and other proteins in its proximity.

Motivated by this unmet need in multi-omic analysis of single cells, we developed a single-cell assay called Prox-seq and demonstrated an

¹Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL, USA. ²Institute for Genomics and Systems Biology, The University of Chicago, Chicago, IL, USA. ³Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland. ⁴Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL, USA. ⁵Department Human Genetics, The University of Chicago, Chicago, IL, USA.

⁶These authors contributed equally: Luke Vistain, Hoang Van Phan. ✉e-mail: tays@uchicago.edu

end-to-end experimental and computational pipeline for proteomic analysis (Fig. 1a). Prox-seq simultaneously measures extracellular proteins, their protein complexes and mRNAs by combining single-cell RNA sequencing (scRNA-seq) with a proximity ligation assay (PLA)⁴. Prox-seq uses pairs of DNA-conjugated antibodies, called Prox-seq probes, that are designed such that, upon being in proximity, the DNA oligonucleotides (oligomers) on the antibodies are ligated⁴. This yields a ligated PLA product that is read out with next-generation sequencing. From the count of PLA products, we can infer the protein abundance, which is similar to assays such as CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) and REAP-seq (RNA expression and protein sequencing assay)^{5,6}, and protein complex information (Fig. 1b). Prox-seq has potential for highly multiplexed proteomic analysis, because the number of possible protein complexes scales quadratically with the number of targeted proteins (Fig. 1b). In addition, Prox-seq can readily quantify gene expression, enabling multimodal analysis of single cells.

For each protein target, we generated a pair of Prox-seq probes (probe A and B), each of which is an antibody conjugated to a single-stranded DNA oligomer (Fig. 1a). The ratio of oligomer-to-antibody was selected to ensure that probes retain their ability to bind their targets (Supplementary Fig. 1). The oligomers were designed such that each member of probe A can ligate with any member of probe B through a universal connector region (Fig. 1c). The complete PLA product spans 119 bases, with 20 of those bases hybridized to a connector. Based on estimates of ssDNA and dsDNA length, we expect that Prox-seq has a range of 53.8–73.7 nm⁷. At this length scale, PLA products can span the entire length of typical protein complexes⁸. After probe binding and ligation, the cells were processed through scRNA-seq methods that utilize poly-A capture, including droplet-based sequencing (Drop-seq), Smart-seq2 and 10x Genomics Chromium^{9,10}, to retrieve both PLA products and mRNAs. Because ligation requires both a probe A and a probe B, only ligated products can be measured by Prox-seq. Unligated Prox-seq probes are automatically discarded during the library preparation step. The oligomers used to form PLA products include several key features (Fig. 1d). The complete PLA product includes a unique molecular identifier (UMI) region for PCR bias correction, two barcode regions to identify the protein targets of the A and B antibodies, a 3' poly-A tail for capture, and a primer binding site for PCR. The universal connector regions enable proximity ligation, and only ligated products can be PCR amplified.

PLA-type assays have previously been used for sensitive detection of proteins^{11–13}. While these methods have been separately applied to make single-cell measurements^{11,12,14,15} and to use a sequencing readout to measure proteins¹³, these two properties have not been combined into a single assay. Furthermore, while PLA has been used to detect protein complexes in situ, it has not been paired with a sequencing output to measure a high number of protein complexes at the same time⁴. Currently, single-cell protein complex measurements are limited to less than ten complexes per cell¹⁶. Prox-seq expands this number to hundreds of protein complexes. Further, Prox-seq can measure both proteins and whole-transcriptome mRNA thus recapitulating the functionality of REAP-seq and CITE-seq^{5,6}.

Results

Prox-seq measures proteins, protein complexes and mRNA simultaneously

We first sought to show that PLA products can be measured using scRNA-seq, and that the PLA data display cell-type-specific differences. Eleven protein targets were selected corresponding to T cell and B cell markers (Supplementary Table 1). Prox-seq probes were made for these targets along with two isotype controls. This panel was applied to a mixture of T cells (Jurkat) and B cells (Raji), which was then analyzed using the Drop-seq pipeline¹⁰ (Supplementary Table 2).

Prox-seq measurements showed that cells could be accurately clustered using mRNAs, proteins or total PLA products (Fig. 1e–h). The protein abundance is estimated by taking the total number of times the protein target's DNA barcode is detected, from either Prox-seq probe A or B (Supplementary Methods). We found that clustering of cells by mRNA or protein identified the same cell types (Fig. 1e,f,h). Similarly, cells could be clustered using all 169 PLA products, which includes protein proximity information in addition to protein abundance (Fig. 1g). Regardless of the data type used, Prox-seq displayed good concordance between gene expression and the protein abundance once the cells were clustered (Fig. 1i). However, we found that the correlation between mRNA and protein for individual cells varied greatly between genes, and was typically modest, similarly to other studies (Fig. 1i and Supplementary Fig. 2)^{5,6}. We also found that PD1–CD3 and CD3–CD3 complex PLA products were two of the most significantly enriched PLA products in the Jurkat cluster (Wilcoxon rank-sum test, Benjamini–Hochberg-adjusted P value = 2.2×10^{-55} and 5.1×10^{-53} , respectively; Fig. 1j). Flow cytometry confirmed CD3 and PD1 as Jurkat-specific proteins (Extended Data Fig. 1). For the Raji cluster, ICAM1–HLA-DR and HLA-DR–HLA-DR were two of the most significantly enriched PLA products (Wilcoxon rank-sum test, Benjamini–Hochberg-adjusted P value = 2.7×10^{-54} and 2.4×10^{-46} , respectively; Fig. 1j). Flow cytometry confirmed that both ICAM1 and HLA-DR were indeed uniquely expressed on Raji cells (Extended Data Fig. 2).

We next sought to show that Prox-seq quantifies protein expression in single cells. We treated Jurkat and Raji cells with a panel of 13 Prox-seq probes and analyzed the PLA products using a plate-based sequencing method (Supplementary Methods). The plate-based method was chosen because such methods typically yield more UMIs per cell¹⁷. This panel allowed us to measure up to 91 potential pairwise protein complexes (Fig. 1b). We observed minimal nonspecific antibody binding (Supplementary Fig. 3). Comparing flow cytometry to Prox-seq showed high correlation (Spearman's correlation coefficient, 0.88) between mean fluorescence intensity and UMIs (Extended Data Fig. 3). Prox-seq probes that fail to find a partner do not contribute to quantification. To ensure that this property does not interfere with protein quantification, we performed a modified Prox-seq protocol that enables measurement of both ligated and unligated Prox-seq probes (Extended Data Fig. 4). We found that more than 90% of Prox-seq probes were ligated with other probes, which offers a straightforward explanation of why Prox-seq quantification agrees with flow cytometry (Extended Data Figs. 3 and 4b). These results demonstrated that Prox-seq accurately characterizes protein species in single cells, and

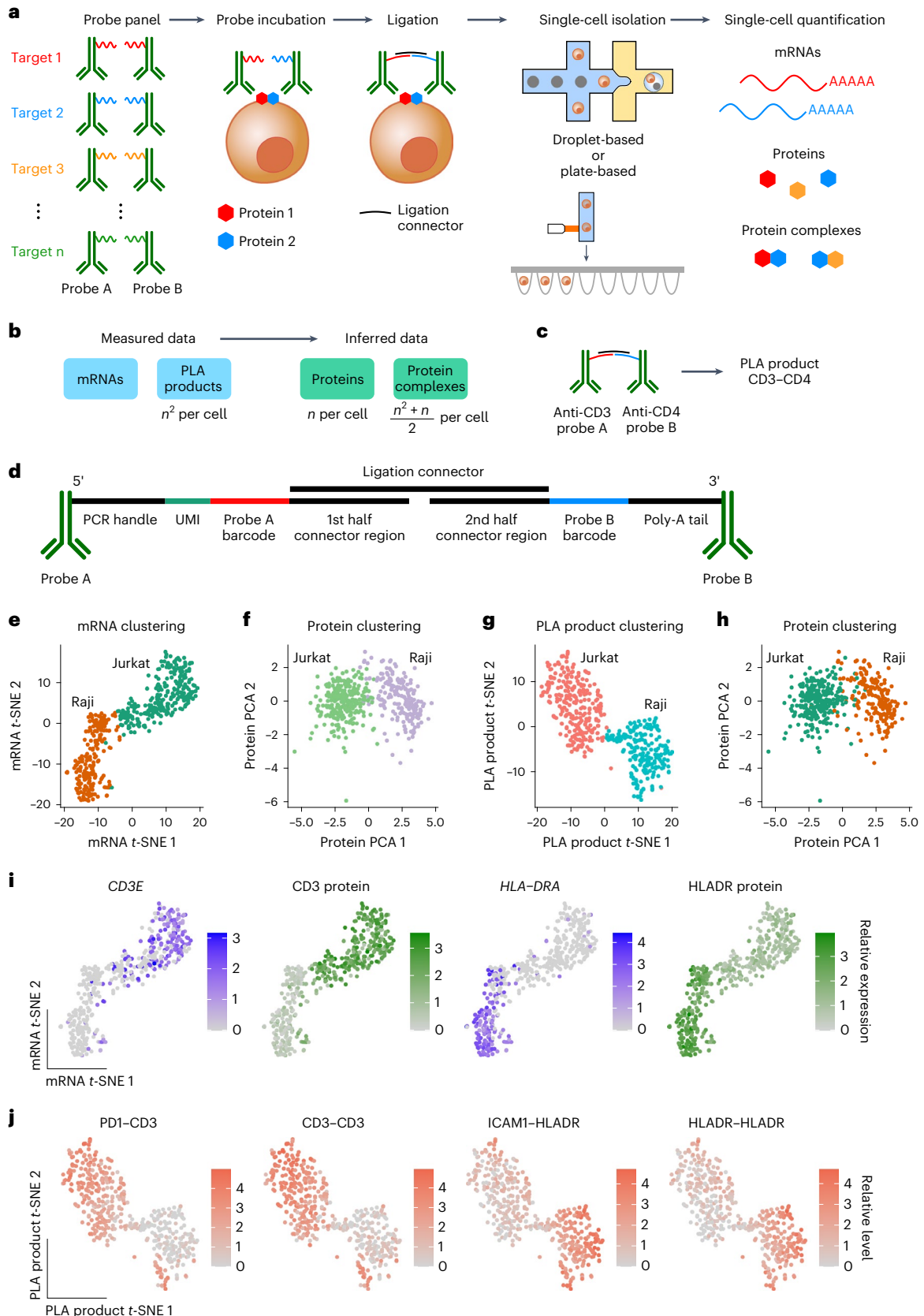
Fig. 1 | Overview of Prox-seq for joint proteomic and transcriptomic analysis of single cells. **a**, Prox-seq workflow: cells are stained with a panel of Prox-seq probe pairs (Prox-seq probe A and B), ligated and processed using a droplet-based or plate-based scRNA-seq protocol. **b**, The measurement output of Prox-seq is the transcript count for each single cell, and the count of n^2 PLA products for each single cell, where n is the number of targeted proteins. **c**, Naming convention of PLA products. If its probe A targets protein CD3, and its probe B targets protein CD4, then the PLA product is called CD3–CD4. **d**, The design of the DNA barcode oligomer. **e**, t -distributed stochastic neighbor embedding (t -SNE) plot showing single cells (T cells/Jurkat and B cells/Raji) clustered with mRNA data. **f**, Principal-component analysis (PCA) plot showing single cells

clustered with protein abundance data (which is calculated from PLA product data). **g**, t -SNE plot showing single cells clustered with PLA product data. **h**, Concordance between cell-type clusters is displayed using the same PCA plot as in **f**, but with cluster labels obtained from mRNA data as in **e**. **i**, Cluster-level concordance between protein and mRNA levels is shown for the *CD3E* gene, CD3 protein, *HLA-DRA* gene and HLA-DR protein. The single cells are colored by the relative expression of mRNAs and proteins. **j**, Plots showing the expression levels of two of the most significant PLA product markers for each cell type (P values $< 10^{-40}$, two-sided Wilcoxon rank-sum test with Benjamini–Hochberg correction). The single cells are colored by the relative level of PLA products.

recapitulates the protein quantification feature of other assays such as REAP-seq and CITE-seq^{5,6}.

A unique feature of Prox-seq, and a major advantage over existing single-cell proteomic techniques, is that it reveals pairwise protein interactions for each of the targeted proteins (Fig. 1). Interactions between

proteins may be due to the formation of a stable complex or due to random (transient) proximity of proteins. PLA product counts alone do not distinguish these possibilities (Supplementary Methods)¹⁸. Therefore, we sought to identify the PLA products that represent protein complexes. In the absence of complexes, the probability of a PLA



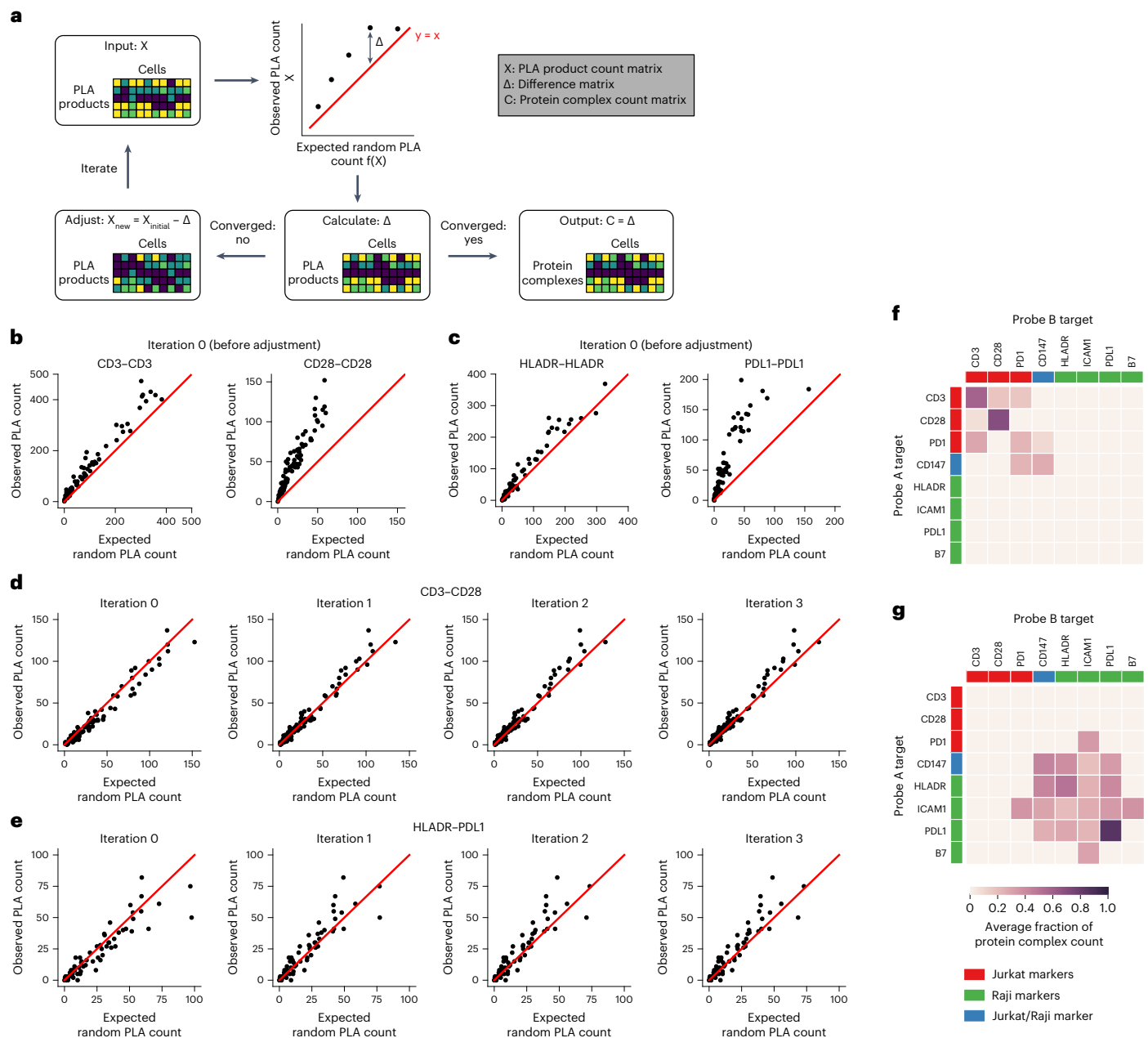


Fig. 2 | Quantification of protein complexes and the proximity ligation background. **a**, Prox-seq measures raw read counts for each PLA product, and the maximum extent of background signal (expected random PLA count) is calculated for each complex from this data (Supplementary Methods). The estimation of random counts is further improved by using an iterative algorithm. **b**, Scatterplots showing PLA counts for two complexes from single Jurkat cells before application of the algorithm. In the scatterplots, each dot represents a single cell, the x axis indicates the expected amount of PLA products from random ligation, and the y axis indicates measured PLA product counts. **c**, Scatterplots showing PLA counts from single Raji cells before application of

the algorithm. **d**, Complex detection algorithm reveals additional complexes with lower abundance. Scatterplots showing the changes in the expected random count of CD3–CD28 in Jurkat cells, after the first three iterations of the algorithm. The algorithm converges between iteration two and three, and the values remain unchanged. **e**, Scatterplots showing the changes in the expected random count of HLA-DR–PDL1 in Raji cells, after the first three iterations of the algorithm. **f, g**, Heat maps showing the final protein complex abundance as a fraction of observed PLA count (fraction of X–Y complex = X–Y complex UMI count/X–Y PLA product UMI count), averaged across all Jurkat (**f**) and Raji (**g**) cells. Complex abundances were calculated after algorithm convergence on each cell.

product forming by random proximity is determined by the concentration of its corresponding probe A and B on the surface of the cell. Using this assumption, we calculated an expected random count for each PLA product based on the Prox-seq probe abundance. This expected random value reflects the maximum amount of a PLA product from random ligation, that is, if none of the targeted proteins were in a complex with one another (Supplementary Methods). When we compared these values to our experimental data, we found several PLA products that were present

at a higher abundance than the expected random value, indicating the presence of stable protein complexes (Fig. 2 and Supplementary Figs. 4 and 5). For example, CD28–CD28 and CD3–CD3 homodimers were high abundance complexes in Jurkat cells (Fig. 2b), whereas the PDL1–PDL1 homodimer was present at very high abundances on Raji cells (Fig. 2c), as expected. The difference (Δ) between the measured and expected random counts indicates the PLA product counts attributed to the stable protein complexes on each cell (Fig. 2).

To further improve the estimation of random proximity background, we developed a computational approach (Fig. 2a–e and Supplementary Methods). Raw Prox-seq data provides matrices for measured PLA product counts, from which we calculated the maximum extent of background for each protein complex. Then, we executed an iterative algorithm to further refine this background estimation. First, the algorithm calculates the expected random count of each PLA product as a first guess of the background. The algorithm then solves a system of quadratic equations describing all possible protein complexes, and produces a new estimate. To account for single-cell variation, we performed a one-sided *t*-test with Benjamini–Hochberg correction (once per iteration for all complexes). If a protein complex estimate is not statistically significant, then the algorithm predicts that the PLA product does not correspond to a stable complex, and the protein complex estimate from the previous iteration is left unchanged (Supplementary Methods). If a complex estimate is statistically significant (adjusted *P* value < 0.05), then the algorithm predicts that the PLA product corresponds to a stable protein complex, and the complex count is updated with the current iteration's estimate (Fig. 2a). Next, the updated protein complex count is used to adjust the PLA product counts, and the algorithm starts the next iteration. The algorithm converges when the absolute change in protein complex count between two successive iterations is below the convergence threshold (Supplementary Methods). As we iterate, we updated our estimate of the background component, hence the expected random counts change with each iteration (Fig. 2d,e).

The difference between the measured counts and the final refined background reveals several other complexes that are at low abundance, yet still significantly above the random ligation background (Fig. 2f,g). We applied our algorithm to Jurkat cells and Raji cells and found that four proteins were calculated to have more than 50% of their PLA product counts attributed to protein complexes: CD3 and CD28 homodimers in Jurkat cells, and PDL1 and HLA-DR homodimers in Raji cells (Fig. 2f,g). Similar results were obtained with a Fisher's Exact Test, also identifying the main protein complexes we found (Supplementary Fig. 6).

Identification of the CD3 and CD28 homodimers in the T cells is noteworthy because they serve as positive controls in our panel. The CD3 Prox-seq probes target the CD3 ϵ protein, two of which are part of the TCR complex¹⁹. CD28 is known to form a stable homodimer on the cell surface through a disulfide bridge²⁰. While it is unclear from previous studies if PDL1 forms a homodimer on the cell surface, all crystal structures of PDL1 feature a homodimer²¹. HLA-DR is thought to exist in an equilibrium between monomers and homodimers on the B cell surface^{22,23}. Therefore, our Protein Complex Estimation Algorithm correctly identified the presence of four known protein complexes. However, B7 and ICAM1 are both thought to undergo some degree of homodimerization^{24,25}. ICAM1 does indeed have the highest number of PLA products attributed to homodimers but, due to its very high expression level, the homodimer represents a small percentage of ICAM1 UMIs (approximately 27%; Fig. 2g). The absence of B7 homodimers raises the possibility that the monoclonal antibody in this panel is unable to bind to the dimerized form. In summary, the proposed algorithm allowed us to determine additional, low-abundance PLA products that correspond to protein complexes and provided a statistical framework to identify and quantify these complexes in our data.

Highly multiplexed quantification of protein complexes in peripheral blood mononuclear cells

We next explored the potential of Prox-seq to measure a large number of protein complexes, and tested its scalability. We determined the effect of panel size on nonspecific antibody binding by comparing Jurkat and Raji cell probes with overlapping Prox-seq panels of different sizes (Extended Data Fig. 5). There was negligible increase in nonspecific binding with increasing panel size (Extended Data Fig. 5d).

This is consistent with the low nonspecific binding levels previously reported in REAP-seq and CITE-seq^{5,6}, which also use barcoded antibody probes for protein detection. We then generated a panel of Prox-seq probe pairs targeting 38 immune cell markers, with a primary focus on T cell markers (Supplementary Table 3). This panel measures up to 741 unique protein complexes. We applied this panel to single human peripheral blood mononuclear cells (PBMCs) and analyzed the sample using two different methods: the plate-based method to maximize our ability to measure potentially rare protein complexes, and the droplet-based 10x method to simultaneously measure mRNA and PLA products in a high-throughput manner.

The plate-based data showed protein measurements that clearly identified the expected cell types: CD8⁺ T cells, CD4⁺ T cells and non-T cells (which do not express CD3; Fig. 3a). Our complex detection algorithm identified 20 protein complexes present in these cells at different levels (Fig. 3b). As before, we identified several known homodimers including the CD3 homodimer, the CD28 homodimer and the CD9 homodimer^{19,20,26} (Fig. 3b). In addition, we identified the existence of both the CD3–CD8 and CD3–CD4 protein complexes (Fig. 3b). Formation of both complexes is consistent with stimulation of T cells with the anti-CD3 antibody in our cocktail^{27,28}. Single-cell heat maps of an example CD4⁺ T cell (Fig. 3c) and CD8⁺ T cell (Fig. 3d) showed clear differences, both in terms of detected PLA products and detected protein complexes.

CD8 associates with CD9 on naïve CD8⁺ T cells

Beyond these known protein complexes, we also identified a potentially new interaction between CD9 and CD8. For CD8⁺ T cells, we observed that cells could be split into two clear subpopulations. In one subpopulation, CD9 PLA products were primarily identified as paired with themselves (CD9–CD9). The other subpopulation displayed CD9 PLA products primarily paired with proteins other than themselves (Fig. 3e). We then sought to identify which protein was interacting with CD9 when the CD9–CD9 PLA product was disfavored. Interestingly, analysis of the CD9–CD9 PLA product-low subpopulation identified the existence of the CD9–CD8 protein complex (Fig. 3f). This is not a previously known complex. However, CD9 is known to participate in immune synapse formation, colocalize with CD3 and coprecipitate with CD3 protein^{29,30}. The appearance of this protein complex is not clearly attributable to changes in protein expression levels, as CD3, CD8 and CD9 were all similarly expressed in both cell populations (Fig. 3g). While CD4⁺ T cells also displayed these two subpopulations to a lesser degree, no CD4–CD9 protein complexes were identified in these cells (Extended Data Fig. 6).

To explore the interplay between protein complexes and mRNA, and to identify the two CD8⁺ T cell subpopulations, we performed a matching experiment using a 10x workflow. This experiment yielded simultaneous measurements of mRNA, protein complexes and protein levels for over 8,700 single cells. We were able to cluster cell types based on their mRNA levels. PLA product information correlated well with the cell types identified by the mRNA information (Fig. 4a,b).

Next, we investigated the correlation between mRNA and protein levels for each of our targets. We once again found that mRNA and protein are correlated on the level of clusters, but only modestly correlated on the single-cell level (Fig. 4c and Supplementary Fig. 7). PLA products reflected levels of protein and complexes for various clusters (Fig. 4d). This cocktail enables measurement of up to 741 protein complexes. Of those 741 potential complexes, we identified 37 as being present, which largely overlaps with the 20 complexes identified by plate-based methods (Extended Data Fig. 7, Supplementary Table 4 and Supplementary Data 1). Of those 37 protein complexes, 21 of them are supported in the literature or the IntAct protein complex database³¹ (Supplementary Table 4). Prox-seq failed to identify 8 protein complexes found in the IntAct database (Supplementary Table 4). Each of these complexes included a protein with a median expression of fewer than five UMIs per cell.

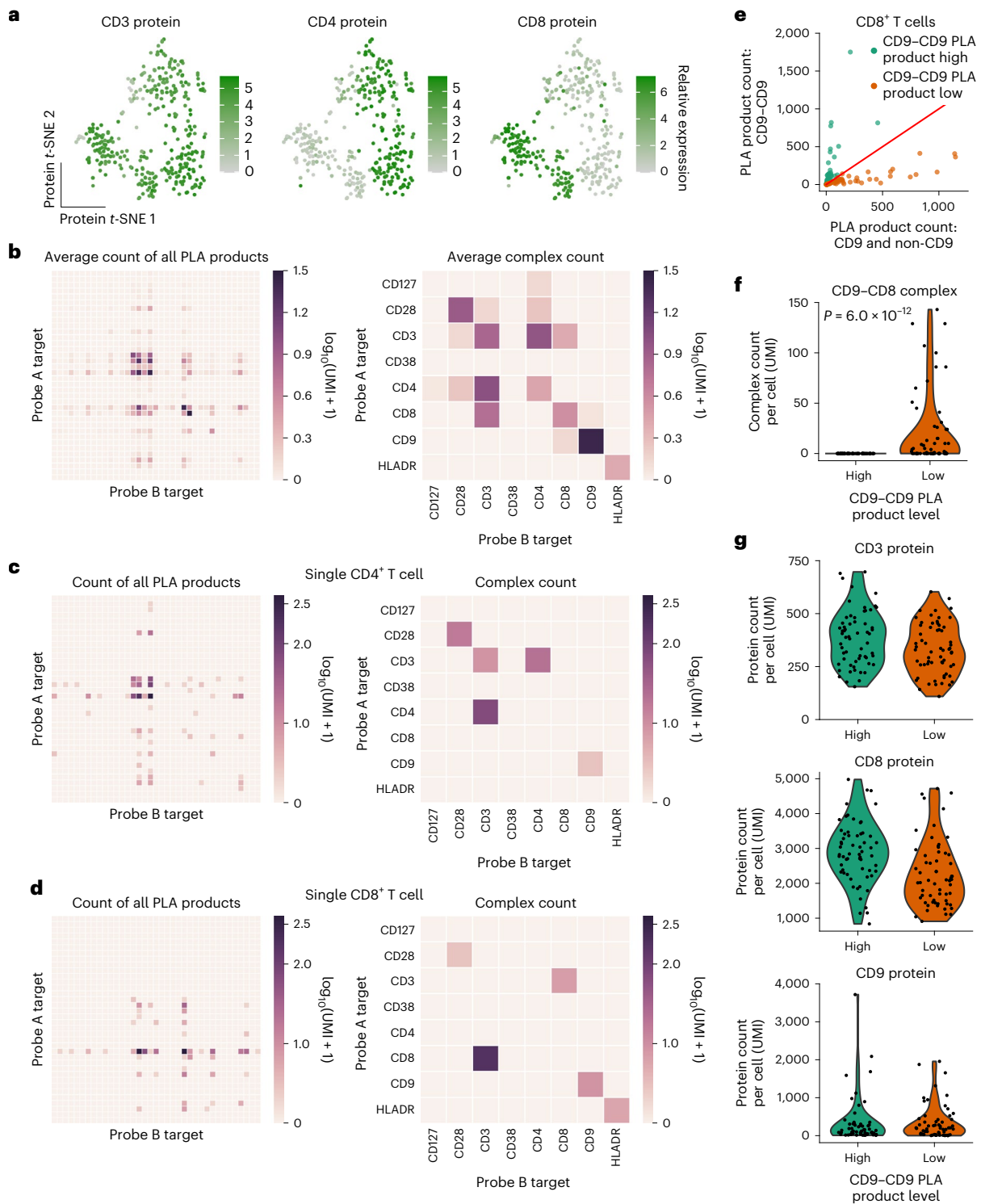


Fig. 3 | Prox-seq reveals a new CD9-CD8 interaction in peripheral blood mononuclear cells. **a**, *t*-SNE plots, using protein data, showing that Prox-seq can identify CD8⁺ and CD4⁺ T cells from CD3, CD4 and CD8 protein expression. The single cells are colored by relative expression of CD3, CD4 and CD8 proteins. **b**, Heat maps showing the average count of PLA products and complexes predicted from the complex detection algorithm across all single cells. The counts were log-transformed before the average is calculated. **c,d**, Heat maps showing the count of all PLA products and detected complexes of an example single CD4⁺ T cell (**c**) and CD8⁺ T cell (**d**). **e**, The presence of two

CD9 dimerization states can be seen from scatterplot of CD9 homodimer counts compared to CD9 heterodimer counts in CD8⁺ T cells. Cells were divided into two groups based on the red line, which indicates $y = x$. **f**, Violin plot showing the distribution of the protein complex CD9-CD8 in the two subpopulations of CD8⁺ T cells (one-sided Wilcoxon rank-sum test). If the algorithm does not detect a protein complex in a cell, a value of 0 is assigned to that read count. **g**, Violin plots showing the distribution of proteins CD3, CD8 and CD9 in the two subpopulations of CD8⁺ T cells.

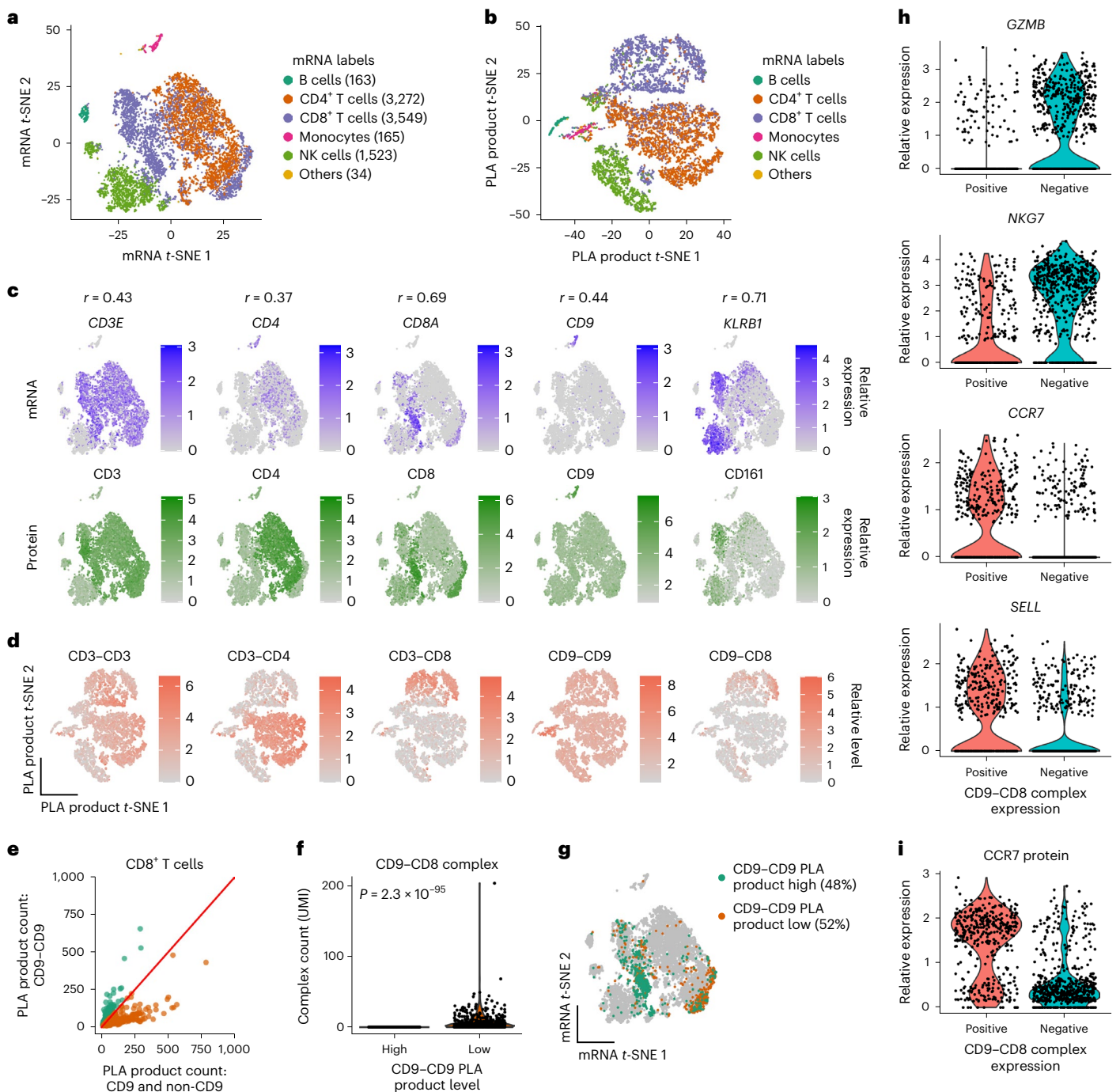


Fig. 4 | Simultaneous protein and mRNA measurements by Prox-seq on 8,700 single peripheral blood mononuclear cells. a, b, *t*-SNE plots of single cells clustered on mRNA (**a**) or PLA (**b**) products (proteins and complexes). In **a** and **b**, the cells are labeled using mRNA data. NK, natural killer. **c**, *t*-SNE plots showing the correlation between mRNA and protein levels in single cells. The plots also show the Pearson's correlation coefficient (r) for each mRNA–protein pair. **d**, *t*-SNE plots showing the level of select PLA products. **e**, Scatterplot showing two subpopulations of CD8⁺ T cells, according to CD9–CD9 PLA product level. Cells were divided into two groups based on the red line, which indicates $y = x$. For better visualization, three outlier single cells with a CD9 and non-CD9 PLA

product count higher than 1,000 were not plotted. **f**, Violin plot showing that the CD9–CD8 protein complex is present in the CD9 homodimer-low subpopulation of CD8⁺ T cells (one-sided Mann–Whitney U test). If the algorithm does not detect a protein complex in a cell, a value of 0 was assigned to that read count. **g**, *t*-SNE plot showing the location of the two subpopulations of CD8⁺ T cells. **h**, Violin plots showing that the subpopulation of cells expressing CD9–CD8 protein complex are downregulated in activation markers (*GZMB* and *NKG7* mRNAs) and upregulated in naive T cell markers (*CCR7* and *SELL* mRNAs). **i**, Violin plot confirming expression of *CCR7* at the protein level.

Measurements using the 10x genomics pipeline reproduced the findings from the plate-based method that CD8⁺ T cells are separated into two subpopulations based on CD9–CD9 PLA product levels (Fig. 4e). In the subpopulation with low CD9–CD9 PLA product, CD9 was found to be in a protein complex with CD8 (Fig. 4f). With

the benefit of mRNA information, we found that these two cell types displayed very different transcriptional profiles (Fig. 4g). Cells without the CD9–CD8 protein complex showed upregulation of *GZMB* and *NKG7* genes (Fig. 4h). Each of these genes is a marker of activated lymphocytes³² (Supplementary Data 2). Conversely,

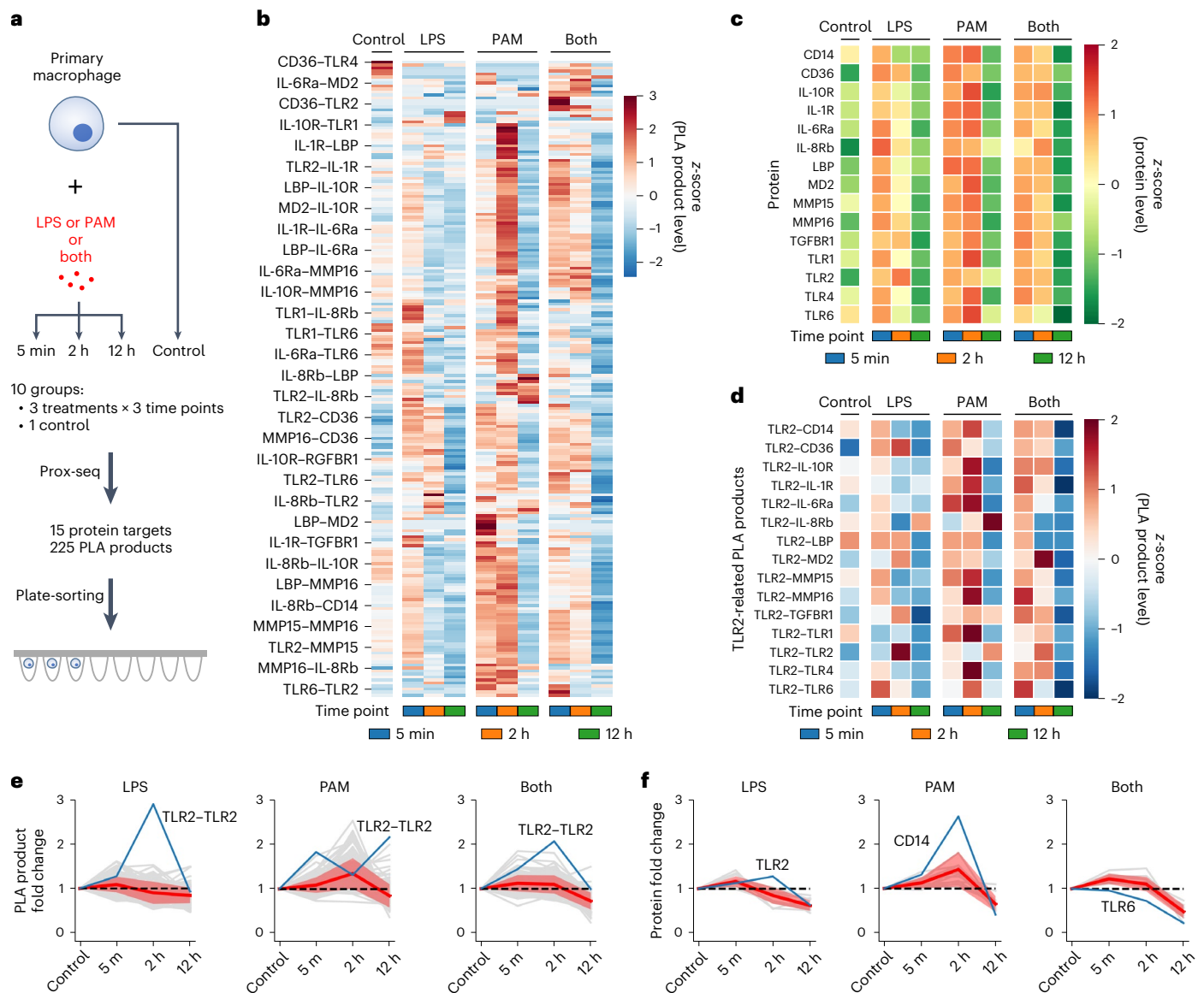


Fig. 5 | Prox-seq enables the study of receptor interactions under combined Toll-like receptor stimulation in macrophages. **a**, Primary human macrophages were treated with LPS, PAM or both LPS and PAM for 5 min, 2 h or 12 h, fixed, and then processed with Prox-seq. Cells were then processed with plate-based Prox-seq. **b**, The general time course of stimulation response can be seen from a heatmap showing the average expression of all protein and protein complex products across the ten conditions. For visualization purposes, the rows are clustered with hierarchical clustering (Euclidean distance metric and complete linkage), and the dendrogram is hidden. **c**, Heat map showing the average expression of all proteins across the ten conditions. **d**, Some binding partners for TLR2 differed from the average protein expression values, as displayed in a heatmap showing the average of all binding partners with TLR2.

In **b–d**, the UMI counts are log-transformed, then averaged by condition and standardized to calculate the row-wise z-score. **e, f**, The average fold change of all PLA products (**e**), and all proteins for each type of ligand (**f**). A pseudocount of one UMI was added to the numerator and denominator for fold-change calculations. The gray lines indicate the fold change of individual PLA products or proteins, the red lines indicate the average of all PLA products or proteins, and the red bands indicate the standard deviation. In **e**, $n = 155, 166$ and 159 PLA products are shown for the LPS, PAM and both treatment groups, respectively. The blue lines indicate the fold change of TLR2–TLR2 for each stimulation condition. In **f**, $n = 15$ individual proteins are shown for all three treatment groups. A select protein is shown with a blue line.

cells with the CD9–CD8 protein complex displayed upregulation of *SELL* and *CCR7* genes, both of which are markers for naïve T cells (Fig. 4h)³². Furthermore, we also observed the differential expression of *CCR7* protein (Fig. 4i). Taken together, these data suggest that the presence of the CD9–CD8 protein complex is a marker of naïve CD8⁺ T cells. We note that it is unlikely that the activation status displayed by some cells is a response to our Prox-seq cocktail. While the cocktail does include stimulatory antibodies, the entire time course of antibody exposure is 30 min, far less than is typically required to activate T cells³³.

Prox-seq shows macrophage signaling additivity via receptor dynamics

We developed a panel of Prox-seq probe pairs targeting 15 surface proteins known to be involved in the nuclear factor kappa B (NF- κ B) signaling pathway, a central mediator of innate immunity³ (Supplementary Table 5). This panel measures up to 225 protein dimers on each cell. First, primary human macrophages were exposed to ligands that activate NF- κ B in the form of lipopolysaccharide (LPS), Pam2CSK4 (PAM) or both. LPS activates TLR4, PAM activates TLR2, and both receptors signal to the NF- κ B pathway. Untreated cells are included as the control.

For each ligand, cells were stimulated for 5 min, 2 h or 12 h (Fig. 5a). Then, the cells were harvested, fixed and processed with plate-based Prox-seq. Fixation was used to preserve the receptor interaction for the 5-min stimulation group and to prevent the antibodies from inducing stimulation and introducing artifacts.

Overall, there was a clear trend of increasing PLA products following stimulation through 2 h, and a sharp decline at 12 h (Fig. 5b). However, this trend was not universal, with some PLA products rising through the entire time course or appearing only at 12 h (Fig. 5b). In contrast, total protein levels were consistently lower at 12 h (Fig. 5c). We found that the tendency for a protein to produce a pair is not strictly a result of protein expression levels. For example, TLR2 displayed major changes in its preferred PLA product partners, depending on both time and stimulant, that do not always track with the protein levels for these partners (Fig. 5d). Consistent with previous single live-cell imaging studies of NF- κ B dynamics, LPS stimulation displayed a faster response with most PLA products peaking at 5 min, whereas PAM displayed a slower response that peaked at 2 h³⁴ (Fig. 5e).

Prox-seq is well suited for studying how signals are integrated when cells encounter two different signals. When LPS and PAM were used simultaneously for combinatorial stimulation of macrophages, the change in PLA products on average showed traits of both stimuli in an additive manner across stimulation durations, with a broad peak that was sustained until finally dropping at 12 h (Fig. 5e). Proteins showed a similar trend as PLA products (Fig. 5f). This simple additivity suggests that for the proteins that we measured, LPS and PAM are operating independently, without synergy. This result is consistent with previous studies that used live-cell microscopy to identify non-integrative signaling between LPS and PAM³⁴.

Prox-seq identifies signaling inputs received by macrophages

Live-cell microscopy measurements of NF- κ B transcription factors can predict if a cell was stimulated with LPS or PAM³⁴. We reasoned that the changes in receptor organization could also identify the stimulating ligand in a mixed stimulation scenario. We trained a logistic regression classifier using PLA count data at each time point after only LPS or PAM stimulation. For the 2-h time point, our classifier was able to identify PAM-like or LPS-like macrophage responses (Fig. 6a). The single largest coefficient for this classification was the presence of the TLR2–TLR2 PLA product, which was highly elevated in the LPS-treated cells (Fig. 6b and Extended Data Fig. 8d). Fivefold cross-validation verified that the 2-h time point was the best option to build the classifier (Extended Data Fig. 8a–d). This classifier was then applied to single cells co-stimulated with both LPS and PAM to classify them into LPS-like, PAM-like or mixed response cells (Extended Data Fig. 8e–g). Most single cells were classified as either LPS or PAM like, but some exhibited characteristics of both signal types (mixed response cells). Technical artifacts were not able to explain the existence of mixed response cells (Supplementary Fig. 8). Remarkably, similar mixed response cells were also observed in live-cell microscopy studies³⁴. A classifier of similar predictive power could be produced from protein data; however, the total PLA products provided more subtle information than the individual proteins alone (Extended Data Fig. 8h–j). For example, all proteins were found to have lower expression in LPS-treated cells compared to PAM-treated ones, while PLA products IL-8Rb–MD2 and IL-1R–TGFBR1 were higher in the former (Extended Data Fig. 8h). Consistent with the logistic regression classifier's results, the TLR2–TLR2 protein complex appeared 2 h after LPS treatment in macrophages, then disappeared at 12 h (Fig. 6c). In contrast, under PAM stimulation this protein complex was absent at early time points (2 h) and appeared only 12 h after PAM treatment (Fig. 6c). While the TLR2 homodimer is known to exist, it is not previously believed to participate in either LPS or PAM signaling^{35,36}.

Prox-seq reveals signaling variability under Toll-like receptor stimulation

Finally, we explored the variability in single-cell signaling responses displayed in our PLA product data. When we compared the mean and variance among all PLA products, we observed a sharp decrease in variance after all stimulation conditions compared to the control group (Fig. 6d and Supplementary Fig. 9). Such a reduced single-cell variability after ligand stimulation was previously seen in single live-cell microscopy studies of NF- κ B signaling^{3,34}. We observed that the low-variance PLA products all contained a CD36 Prox-seq barcode (Fig. 6e and Supplementary Fig. 10). Histograms of all PLA products containing CD36 show two modes in control cells, separated by the number of UMIs (Fig. 6f and Supplementary Fig. 10). LPS treatment caused cells to shift to the higher UMI mode (Fig. 6f). Because this change is already occurring on the 5-min timescale, the increase in PLA products is unlikely to be a result of increased protein expression. Rather, CD36 is involved in interactions and rearrangement of the other proteins targeted by our probe panel. This is further supported by the appearance of new CD36 protein complexes at 5 min (Fig. 6c). CD36 is a scavenger receptor that recognizes a variety of bacterial lipid and lipoprotein molecules³⁷. It has also been shown to act as a co-receptor for TLR2 and TLR4, both of which are stimulated in response to the ligands in our study^{38,39}. Furthermore, stimulation with oxidized low-density lipoprotein can induce CD36 to form a protein complex with TLR4 and TLR6 (ref. 40). Overall, these results show that Prox-seq can identify rearrangement and cell-to-cell variability of receptor components during signaling.

Read depth requirements for Prox-seq

We next explored if Prox-seq has different read depth requirements compared to other single-cell sequencing modalities. As is typical for single-cell sequencing, we found that plate-based methods offer the highest library complexity (Supplementary Table 2)^{41–43}. The 10x Genomics pipeline offers the best tradeoff between library complexity and cost when mRNA is also recovered. To further explore the relationship between read depth and Prox-seq performance, we performed a downsampling analysis, whereby reads were randomly removed from cells to simulate lower read depth (Extended Data Fig. 9). We found that for the protein and PLA product modalities, the number of UMIs per cell and features per cell (where features can be proteins, PLA products or protein complexes) increased as the mean number of reads per cell approached 10,000 (Extended Data Fig. 9a–d). Our data showed diminishing returns for read counts above 10,000 reads per cell (Extended Data Fig. 9f–g). Therefore, we advise users to apply at least 10,000 reads per cell when sequencing PLA products.

Discussion

In summary, we present a practical and broadly applicable technology for simultaneous measurements of extracellular proteins, protein complexes and mRNAs in single cells, and showed its application in different biological contexts. We expect that Prox-seq will be a valuable tool for understanding signaling, differentiation, development and cellular decision-making, which are largely driven by changes in protein interactions. The compatibility with commonly used single-cell sequencing methods allows its wide adoption by many laboratories. Most importantly, Prox-seq can identify members of pairwise protein complexes, providing a new modularity to single-cell sequencing. In this study, we demonstrated the detection of surface proteins in intact single cells, while in principle Prox-seq can be applied to intracellular proteins as well as cell lysates.

There are some limitations inherent to Prox-seq. Several of these stem from the requirement of antibodies. Monoclonal antibodies were primarily used in this study because they enable confident quantification of homodimers. However, monoclonal antibodies likely suffer from a higher false-negative rate than polyclonal antibodies. Polyclonal antibodies, by virtue of having multiple epitopes, should ameliorate

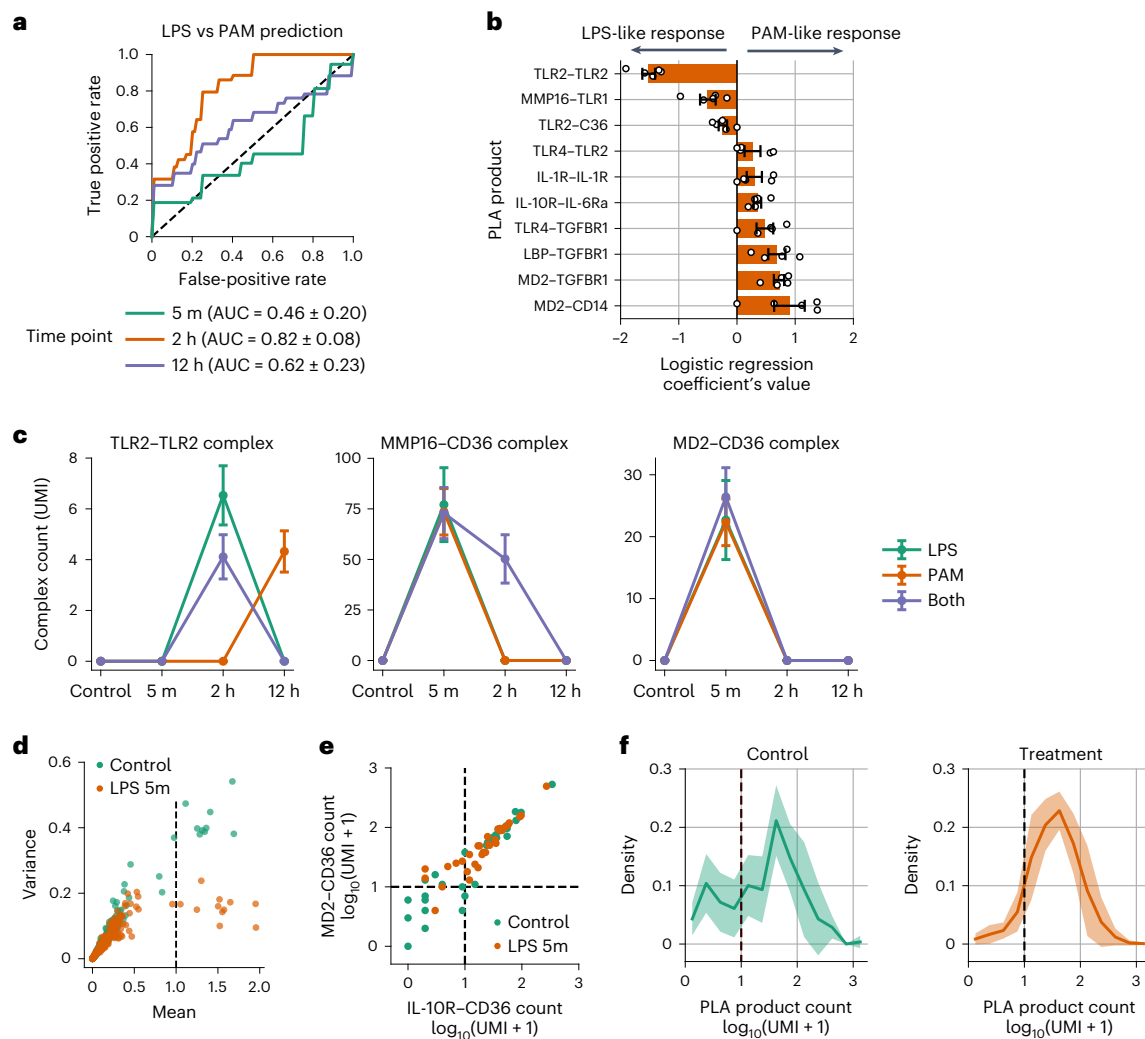


Fig. 6 | Prox-seq reveals single-macrophage variability in TLR signaling and enables identification of immune inputs from protein measurements.

a, Receiver operating characteristic (ROC) curves of a logistic regression classifier trained on PLA product levels from different time points. The classifier is trained to predict whether a single cell was stimulated with LPS or PAM. Each ROC curve represents the mean ROC curve from fivefold cross-validation. The area under the curve (AUC) metric of each time point is presented as the mean \pm s.d. of the AUC metrics from fivefold cross-validation for that particular time point. **b**, Bar plot showing the PLA products that most strongly contributed to the LPS versus PAM prediction at the 2-h time point. Only PLA products with absolute coefficients above 0.2 are shown. A positive value indicates that the PLA product is higher in PAM-treated cells, while a negative value indicates that the PLA product is higher in LPS-treated cells. Each dot represents the value of the logistic regression coefficients from one of the five cross-validation folds. The bars show

the mean \pm s.e.m. of the coefficient's values across $n = 5$ cross-validation folds. **c**, Plots showing the dynamics of three example protein complexes. Data are presented as the mean \pm s.e.m. $n = 31$ single cells for the control group. For the LPS treatment group, $n = 32$ cells for all three time points. For the PAM group, $n = 32$ cells for both 5-min and 2-h time points, and 31 cells for the 12-h time point. For the group treated with both LPS and PAM, $n = 34$ cells for the 5-min time point, and 36 cells for both 2-h and 12-h time points. **d**, Scatterplot showing the relationship between the mean and variance of log-transformed PLA count of control sample, and sample treated with LPS for 5 min. The PLA products with mean values greater than or equal to 1 are CD36 related. **e**, Scatterplot showing relationship between IL-10R-CD36 and MD2-CD36 complexes in control group and 5-min LPS treatment group. **f**, Plots showing the distribution of PLA product counts of the nine CD36-related PLA products in control and treatment samples. The solid lines indicate the mean, and the ribbons indicate the standard deviation of $n = 9$ PLA products.

some false-negative concerns at the cost of losing the ability to reliably quantify homodimers. Similarly to other antibody-based assays, antibodies should be validated for compatibility with Prox-seq. In addition, antibody assays are typically stimulatory when the antibody is directed at a receptor. When this is undesirable, cells should be fixed before Prox-seq analysis. Single-cell sequencing methods have recently been proposed for fixed cells; however, there is usually some loss of data quality^{44,45}.

In this study, we also developed an algorithm for better prediction of the random ligation background between surface proteins, which allows the identification of additional, low-abundance complexes. The data structure of Prox-seq results in coupling between PLA products,

influencing the accurate quantification of protein complex abundance. Our prediction algorithm addresses this challenge, but still has certain limitations. Cells must be clustered before protein complex quantification. Some of the algorithm's parameters are chosen heuristically, which can lead to changes in the predicted protein complexes depending on the parameters. Also, the algorithm currently does not take into account the distribution of PLA product counts. We expect that utilizing the distribution of PLA product counts could further improve statistical power and reduce false-positive rate.

Despite these limitations, Prox-seq accurately identified various cell types, measured expected and uncharacterized protein complexes in human PBMCs, and studied protein rearrangements and complex

formation during TLR signaling in macrophages. We detected known protein complexes such as the CD3 homodimer and the CD28 homodimer in T cells. We also identified a new receptor interaction between CD8 and CD9 on human primary naïve CD8⁺ T cells. Lastly, we observed different temporal changes in receptor arrangements under LPS and PAM stimulation in macrophages and showed additive integration of TLR signals, which are supported by previous live-cell microscopy and modeling studies in single cells³⁴.

Recent advances in single-cell sequencing technology have enabled comprehensive characterization of the transcriptome, genome and epigenome at the single-cell level^{10,46,47}. Several methods have expanded these approaches to incorporate antibody-based protein measurements^{5,6,48}. Furthermore, the field of single-cell mass spectrometry has been undergoing rapid progress^{49,50}. However, the measurement of protein complexes at the single-cell level has a lagging pace compared to other analytes. Prox-seq provides a quadratically scaled multiplexing capability to greatly increase the number of protein complexes that can be measured. Currently, in order to make highly multiplexed measurements of protein complexes, one is limited to bulk samples that cannot be applied to single cells^{51,52}. Methods suited for single cells are limited in their multiplexing capacity, typically measuring fewer than 10 complexes^{4,53}; whereas with Prox-seq, we have demonstrated the ability to survey 741 possible protein complexes in single PBMCs. Furthermore, Prox-seq incorporates scRNA-seq, thereby providing multiple single-cell data types simultaneously, which greatly enhances multi-omic analysis capability in single cells.

Online content

Any methods, additional references, Nature Portfolio 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/s41592-022-01684-z>.

References

1. Grün, D. et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255 (2015).
2. Rizvi, A. H. et al. Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nat. Biotechnol.* **35**, 551–560 (2017).
3. Tay, S. et al. Single-cell NF- κ B dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267–271 (2010).
4. Söderberg, O. et al. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* **3**, 995–1000 (2006).
5. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
6. Peterson, V. M. et al. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936–939 (2017).
7. Chi, Q., Wang, G. & Jiang, J. The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measurements using mean field theory. *Phys. A: Stat. Mech. Appl.* **392**, 1072–1079 (2013).
8. Birnbaum, M. E. et al. Molecular architecture of the $\alpha\beta$ T cell receptor–CD3 complex. *Proc. Natl Acad. Sci. USA* **111**, 17576–17581 (2014).
9. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
10. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
11. Albayrak, C. et al. Digital quantification of proteins and mRNA in single mammalian cells. *Mol. Cell* **61**, 914–924 (2016).
12. Lin, J. et al. Ultra-sensitive digital quantification of proteins and mRNA in single cells. *Nat. Commun.* **10**, 3544 (2019).
13. Darmanis, S. et al. ProteinSeq: high-performance proteomic analyses by proximity ligation and next generation sequencing. *PLoS ONE* **6**, e25583 (2011).
14. Genshaft, A. S. et al. Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol.* **17**, 188 (2016).
15. Darmanis, S. et al. Simultaneous multiplexed measurement of RNA and proteins in single cells. *Cell Rep.* **14**, 380–389 (2016).
16. Leuchowius, K.-J. et al. Parallel visualization of multiple protein complexes in individual cells in tumor tissue. *Mol. Cell. Proteomics* **12**, 1563–1571 (2013).
17. Ziegenhain, C. et al. Comparative analysis of single-cell RNA-sequencing methods. *Mol. Cell* **65**, 631–643 (2017).
18. Alsemarz, A., Lasko, P. & Fagotto, F. Limited significance of the in situ proximity ligation assay. Preprint at *bioRxiv* <https://doi.org/10.1101/411355> (2018).
19. Van Der Merwe, P. A. & Dushek, O. Mechanisms for T cell receptor triggering. *Nat. Rev. Immunol.* **11**, 47–55 (2011).
20. Esensten, J. H., Helou, Y. A., Chopra, G., Weiss, A. & Bluestone, J. A. CD28 costimulation: from mechanism to therapy. *Immunity* **44**, 973–988 (2016).
21. Zak, K. M. et al. Structural biology of the immune checkpoint receptor PD-1 and its ligands PD-L1/PD-L2. *Structure* **25**, 1163–1174 (2017).
22. Cherry, R. J. et al. Detection of dimers of dimers of human leukocyte antigen (HLA)-DR on the surface of living cells by single-particle fluorescence imaging. *J. Cell Biol.* **140**, 71–79 (1998).
23. Cochran, J. R., Cameron, T. O. & Stern, L. J. The relationship of MHC-peptide binding and T cell activation probed using chemically defined MHC class II oligomers. *Immunity* **12**, 241–250 (2000).
24. Bhatia, S., Edidin, M., Almo, S. C. & Nathenson, S. G. Different cell surface oligomeric states of B7-1 and B7-2: Implications for signaling. *Proc. Natl Acad. Sci. USA* **102**, 15569–15574 (2005).
25. Reilly, P. L. et al. The native structure of intercellular adhesion molecule-1 (ICAM-1) is a dimer. Correlation with binding to LFA-1. *J. Immunol.* **155**, 529–532 (1995).
26. Kovalenko, O. V., Yang, X., Kolesnikova, T. V. & Hemler, M. E. Evidence for specific tetraspanin homodimers: inhibition of palmitoylation makes cysteine residues available for cross-linking. *Biochem. J.* **377**, 407–417 (2004).
27. Thome, M., Germain, V., Disanto, J. P. & Acuto, O. The p56lck SH2 domain mediates recruitment of CD8/p56lck to the activated T cell receptor/CD3/ ζ complex. *Eur. J. Immunol.* **26**, 2093–2100 (1996).
28. Collins, T. L. et al. p56lck association with CD4 is required for the interaction between CD4 and the TCR/CD3 complex and for optimal antigen stimulation. *J. Immunol.* **148**, 2159–2162 (1992).
29. Rocha-Perugini, V. et al. Tetraspanins CD9 and CD151 at the immune synapse support T cell integrin signaling. *Eur. J. Immunol.* **44**, 1967–1975 (2014).
30. Toyo-Oka, K. et al. Association of a tetraspanin CD9 with CD5 on the T cell surface: role of particular transmembrane domains in the association. *Int. Immunol.* **11**, 2043–2052 (1999).
31. Kerrien, S. et al. IntAct—open source resource for molecular interaction data. *Nucleic Acids Res.* **35**, D561–D565 (2007).
32. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
33. Trickett, A. & Kwan, Y. L. T cell stimulation and expansion using anti-CD3/CD28 beads. *J. Immunol. Methods* **275**, 251–255 (2003).

34. Kellogg, R. A., Tian, C., Etzrodt, M. & Tay, S. Cellular decision-making by non-integrative processing of TLR inputs. *Cell Rep.* **19**, 125–135 (2017).
 35. Ozinsky, A. et al. The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between Toll-like receptors. *Proc. Natl Acad. Sci. USA* **97**, 13766–13771 (2000).
 36. Qiu, Y. et al. Divergent roles of amino acid residues inside and outside the BB loop affect human Toll-like receptor (TLR)2/2, TLR2/1 and TLR2/6 responsiveness. *PLoS ONE* **8**, e61508 (2013).
 37. Silverstein, R. L. & Febbraio, M. CD36, a scavenger receptor involved in immunity, metabolism, angiogenesis and behavior. *Sci. Signal.* **2**, re3 (2009).
 38. Biedroń, R., Peruń, A. & Józefowski, S. CD36 differently regulates macrophage responses to smooth and rough lipopolysaccharide. *PLoS ONE* **11**, e0153558 (2016).
 39. Seimon, T. A. et al. Atherogenic lipids and lipoproteins trigger CD36–TLR2-dependent apoptosis in macrophages undergoing endoplasmic reticulum stress. *Cell Metab.* **12**, 467–482 (2010).
 40. Stewart, C. R. et al. CD36 ligands promote sterile inflammation through assembly of a Toll-like receptor 4 and 6 heterodimer. *Nat. Immunol.* **11**, 155–161 (2009).
 41. Mereu, E. et al. Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nat. Biotechnol.* **38**, 747–755 (2020).
 42. Ding, J. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **38**, 737–746 (2020).
 43. Zhang, X. et al. Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems. *Mol. Cell* **73**, 130–142 (2019).
 44. Phan, H. V. et al. High-throughput RNA sequencing of paraformaldehyde-fixed single cells. *Nat. Commun.* **12**, 5636 (2021).
 45. Denisenko, E. et al. Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.* **21**, 130 (2020).
 46. van Dijk, D. et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729 (2018).
 47. Stevens, T. J. et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
 48. Shahi, P., Kim, S. C., Haliburton, J. R., Gartner, Z. J. & Abate, A. R. Abseq: ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci. Rep.* **7**, 44447 (2017).
 49. Specht, H. et al. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol.* **22**, 50 (2021).
 50. Zhu, Y. et al. Proteomic analysis of single mammalian cells enabled by microfluidic nanodroplet sample preparation and ultrasensitive NanoLC-MS. *Angew. Chem. Int. Ed. Engl.* **57**, 12370–12374 (2018).
 51. Huttlin, E. L. et al. The BioPlex Network: a systematic exploration of the human interactome. *Cell* **162**, 425–440 (2015).
 52. Rual, J.-F. et al. Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**, 1173–1178 (2005).
 53. Grant, D. M. et al. Multiplexed FRET to image multiple signaling events in live cells. *Biophys. J.* **95**, L69–L71 (2008).
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
- Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.
- © The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

Methods

Prox-seq probe preparation

Antibodies were DNA conjugated using previously published methods⁵⁴. Briefly, antibodies were concentrated and buffer exchanged into PBS before conjugation using a concentrator with a 50,000 molecular-weight cutoff (EMD Millipore). The antibodies were then reacted with dibenzocyclooctyne-PEG4-*N*-hydroxysuccinimidyl ester (DBCO; Sigma, 764019) in dimethylsulfoxide (DMSO; Sigma Aldrich). This was done by combining the antibody solution with a 2-mM DBCO solution at a 10:1 volume-by-volume ratio. This reaction was incubated on ice for 1–2 h. After incubation, the DBCO-conjugated antibodies were purified using a concentrator with a 50,000 molecular-weight cutoff and the antibody-to-DBCO ratio was measured via UV-Vis (Nanodrop)⁵⁴. Around 1–2 μg DBCO-conjugated antibodies (at 3–13 μM) were combined with an equal volume of 80 μM azide-functionalized PLA oligomer (IDT) dissolved in PBS (Life Technologies) and allowed to react overnight at 4 °C. For probes that were stored long term, the reaction mixture was then brought to 50% glycerol/PBS (Sigma Aldrich).

Cell culture

Jurkat and Raji cell lines were a generous gift from J. Huang⁵⁵. Both were maintained at 37 °C with 5% CO₂ in RPMI (Gibco, Thermo Scientific) supplemented with 10% fetal bovine serum (FBS, Hyclone, Fisher Scientific).

Frozen PBMCs and macrophages were purchased from STEMCELL Technologies. They were quickly thawed at 37 °C and washed three times by suspension in 10 ml RPMI + 10% FBS and centrifugation at 300g for 3 min. Cells were then allowed to rest overnight at 37 °C with 5% CO₂ in RPMI + 10% FBS.

Flow cytometry

Jurkat and Raji cells were plated in a 96-well plate (Corning) at 100,000 cells per well. Cells were centrifuged at 500g for 5 min, media was removed and replaced with 30 μl 5 nM Prox-seq probes (2.5 nM probe A and 2.5 nM probe B) in probe binding buffer (PBS, 0.1% BSA (Thermo Scientific), 0.1 mg ml⁻¹ sonicated salmon sperm DNA (Invitrogen), 6.7 nM of each isotype). Cells were incubated with probes for 30 min at 37 °C. Cells were then washed three times by centrifuging at 500g for 5 min and resuspending in 100 μl 1% BSA/PBS. Cells were then resuspended in a 1:100 dilution of secondary antibody (Supplementary Table 6) in 1% BSA/PBS and incubated for 20 min at room temperature. Cells were centrifuged and washed two times as before. Finally, cells were analyzed using a Fortessa 4–15 (BD Biosciences) with a high-throughput screening module.

Jurkat/Raji sample preparation (Drop-seq-based Prox-seq)

In total, 150,000 Jurkat and 150,000 Raji cells were counted and spun at 500g for 3 min, washed once with 1% BSA/PBS and spun again. Cells were then combined and plated in 96-well U-bottom plates. Cells were then centrifuged at 300g for 3 min and fixed with 4 mM 3,3'-dithiobis(sulfosuccinimidyl propionate, Life Technologies) in PBS at 37 °C for 30 min. Cells were washed once with 1 ml 1% BSA/PBS and 90 μl probes were added at 5 nM each pair (2.5 nM probe A + 2.5 nM probe B) in probe binding buffer. Cells were then incubated at 37 °C for 60 min, washed twice with 1 ml 1% BSA/PBS, and ligated with 300 μl ligation solution. Finally, cells were unfixated with 30 mM dithiothreitol at 37 °C for 30 min, centrifuged at 300 g for 3 min, and resuspended in 0.1% BSA/PBS.

Jurkat/Raji sample preparation (plate-based Prox-seq)

Cells underwent different processing procedures depending on whether they were analyzed by a droplet-based or plate-based protocol. For the plate-based analysis, 50,000 Jurkat cells and 50,000 Raji cells were collected, centrifuged at 500g for 3 min, and washed once in 1% BSA/PBS. Jurkat cells were resuspended in 5 μM carboxyfluorescein diacetate

succinimidyl ester (BioLegend) in PBS for 20 min at room temperature to identify Jurkat cells specifically during cell sorting. The cells were then resuspended in 30 μl Prox-seq probes in probe binding buffer. Each probe pair was at 5 nM (2.5 nM probe A + 2.5 nM probe B). Cells were incubated at 37 °C for 60 min. They were then centrifuged and washed three times by centrifuging and resuspending in 1% BSA/PBS as before. Cells were then resuspended in 100 μl ligation solution (50 mM HEPES pH 7.5, 10 mM MgCl₂, 1 mM rATP (New England Biolabs), 9.5 nM connector oligomer (TTTCACGACACGACACGATTAGGTC; IDT), 130 U ml⁻¹ T4 ligase (NEB)) and rotated for 3 h at 37 °C. With 30 min remaining in the incubation, propidium iodide (PI; Invitrogen) was added to the solution to a final concentration of 1 μg ml⁻¹. Cells were then centrifuged at 500g for 3 min, resuspended in 1% BSA/PBS + 1/500 PI. Single PI-negative cells were sorted into each well of two 96-well plates, one for each cell line.

Peripheral blood mononuclear cell sample preparation

One million rested PBMCs were collected and pelleted. All centrifugation steps were performed at 300g for 3 min. The cells were then resuspended in 1 ml 1% BSA/PBS and pelleted. The cells were then resuspended in Fc blocker solution composed of 95 μl 1% BSA/PBS and 5 μl TruStain FcX (BioLegend) and incubated at room temperature for 5 min. Following this step, all buffers were supplemented with 1:1,000 RNase inhibitor (NEB). The cells were then pelleted and resuspended in 300 μl 5 nM Prox-seq probes in probe binding buffer (as above). Due to its size, the antibody cocktail was divided into three parts and administered in series. For each portion, cells were allowed to incubate at 37 °C for 15 min, then pelleted and resuspended in the next portion. Cells were then spun and washed three times with 1 ml 1% BSA/PBS. They were then resuspended in ligation solution, as described above, and incubated at 37 °C for 30 min. For the 10x experiment, cells were transferred to that workflow at this stage (see below). For plate-based experiments, cells were spun and washed in 1% BSA/PBS and resuspended in 500 μl 1% BSA/PBS + 1/500 PI (Invitrogen) and incubated at room temperature for 10 min. Finally, cells were pelleted, resuspended in 500 μl 1% BSA/PBS, and live cells were sorted into plates.

Peripheral blood mononuclear cell sample preparation–10x

After ligation, 1 ml 1% BSA/PBS was added to the cells and they were centrifuged at 300g for 5 min. Cells were resuspended in 1 ml 1% BSA/PBS and centrifuged at 300g for 5 min. Dead cells were removed from this sample using Miltenyi Biotec Dead Cell Removal Kit (130-090-101) following the manufacturer's guidelines. Cells were then counted and diluted to 1,000 cells per μl . This sample was then processed using the recommended 10x protocol and modified library preparation procedure (Supplementary Methods).

Primary macrophage sample preparation

Thawed macrophages were distributed into ten wells of a non-tissue-culture-treated 24-well plate. They were then allowed to rest overnight in RPMI + 10% FBS. While in the plate, cells were centrifuged for 300g for 3 min and stimulated for 12 h, 2 h or 5 min with 100 ng ml⁻¹ LPS, 40 ng ml⁻¹ PAM, or both in RPMI + 10% FBS. All future centrifugation steps occurred at 300g for 3 min. After stimulation, cells were dissociated with TrypLE (Gibco), pelleted and resuspended in 4% paraformaldehyde (PFA, ChemCruz) for 15 min at room temperature. Cells were then spun and washed one time with 1 ml 1% BSA/PBS and resuspended in 95 μl 1% BSA/PBS + 5 μl TruStain FcX. After 5 min at room temperature, cells were pelleted and each sample was resuspended in 54 μl 5 nM Prox-seq probes in probe binding buffer (as above). The cells were allowed to incubate for 30 min at 37 °C. Following probe incubation, the cells were pelleted and washed with 1 ml 1% BSA/PBS three times. The cells were then resuspended in ligation solution as above and allowed to incubate for 30 min at 37 °C. Finally, the cells were pelleted, resuspended in 500 μl 1% BSA/PBS + 1/500 PI (Invitrogen), and live cells were sorted into plates.

Prox-seq

For droplet-based Prox-seq, cells were processed according to the Drop-seq protocol. Briefly, cells were co-encapsulated with bar-coded beads (ChemGenes, Macosko-2011-10(V+)) in droplets using a microfluidic device. Next, the droplets were broken, and the beads were subjected to reverse transcription, exonuclease digestion and whole-transcriptome amplification. The resulting PLA products and cDNAs were processed separately into sequencing libraries.

For plate-based Prox-seq, cells were sorted into 96-well plates containing 4 μ l of Smart-seq2 lysis buffer (0.1% Triton X-100, 1 unit per μ l RNase inhibitor, murine (NEB), 2.5 mM dNTPs, 2.5 μ M SmartSeq2_oligoTVN, 2.5 μ M SmartSeq2_oligoTGT in water). For PFA-fixed primary macrophages, the cells were sorted into 96-well plates containing 6 μ l of modified Smart-seq2 lysis buffer (0.1% Triton X-100, 1,000 units per ml RNase inhibitor, murine (NEB), 20 units per ml proteinase K (NEB), 2.5 mM dNTPs, 2.5 μ M SmartSeq2_oligoTVN, 2.5 μ M SmartSeq2_oligoTGT in TE buffer). For non-PFA-fixed cells, after cell sorting, the plates were frozen at -80°C for storage. When the samples are ready for processing, the plates were thawed on ice and incubated at 72°C for 3 min before library preparation. For PFA-fixed cells, after cell sorting, the plates were incubated at 56°C for 1 h, 95°C for 10 min, 4°C for at least 5 min before storage at -80°C . Afterwards, the plates can be thawed on ice and proceed directly to library preparation. For plate-based Prox-seq library preparation, briefly, 2–4 μ l per well was used for pre-amplification of PLA products, followed by another PCR reaction to attach the single-cell indexes and the sequencing adaptor. More detailed protocols are available in the Supplementary Methods.

All oligonucleotides and primers used for droplet-based and plate-based methods are summarized in Supplementary Tables 7–9.

Next-generation sequencing

For the droplet-based Prox-seq, a NextSeq Mid-output kit v2.5 was used to sequence both mRNA and PLA libraries in the same sequencing run. The cDNA and PLA libraries each received 20% of the total reads. PhiX control was spiked in at 40% concentration according to Illumina's instruction, because of the low diversity of the PLA libraries. Custom read 1 sequencing primer (Read1CustomSeqB), custom read 2 primer (DropPLA_Read2) and custom i7 index read primer (DropPLA_i7Read) were used according to Illumina's instructions. Read distribution was 20 bases for read 1, 85 bases for read 2 and 8 bases for i7 index read. For the plate-based Prox-seq, PLA libraries from four 96-well plates were sequenced with a mid-output NextSeq kit v2.5. PhiX control was spiked in at 40% concentration according to Illumina's instructions. Custom read 1 sequencing primer (SmartPLA_Read1), custom i5 index read primer (SmartPLA_i5Read) and custom i7 index read primer (SmartPLA_i7Read) were used according to Illumina's instructions. Read distribution was 75 bases for read 1, 8 bases for i5 index read and 8 bases for i7 index read.

Sequencing alignment

Drop-seq mRNA-sequencing data were aligned using Drop-seq tools v2.3.0. The 10x mRNA-sequencing data were aligned using Cell Ranger v6.1.1 and the human reference genome GRCh38 version 2020-A from 10x Genomics. The sequencing data of PLA products were aligned using a custom Java program available at <https://github.com/tay-lab/Prox-seq/>. More details are available in the Supplementary Methods.

Reporting summary

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

Data availability

The raw and count data are deposited in NCBI's Gene Expression Omnibus under accession numbers [GSE149574](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149574) and [GSE196130](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE196130). Source data are provided with this paper.

Code availability

The custom program for PLA product alignment and the codes used for alignment and data analysis are available at <https://github.com/tay-lab/Prox-seq/>.

References

- Gong, H. et al. Simple method to prepare oligonucleotide-conjugated antibodies and its application in multiplex protein detection in single cells. *Bioconjug. Chem.* **27**, 217–225 (2016).
- Hui, E. et al. T cell co-stimulatory receptor CD28 is a primary target for PD-1-mediated inhibition. *Science* **355**, 1428–1433 (2017).

Acknowledgements

We thank J. Huang (Pritzker School of Molecular Imaging, University of Chicago) for the generous gift of the Jurkat and Raji cell lines used in this study. We thank A. Basu (Section of Genetic Medicine, Department of Medicine, University of Chicago) and H. Eckart (Section of Genetic Medicine, Department of Medicine, University of Chicago) for their advice on the Drop-seq protocol. We acknowledge both The University of Chicago Genomics Facility, The University of Chicago Cytometry and Antibody Technology facility and The University of Chicago Research Computing Center for their services. We thank U. Landegren (Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University) for advice with PLA. We thank Z. Ren and C. Chen at the CRI Bioinformatics Core for their advice on the alignment program. We thank A. A. Khan (Department of Pathology, University of Chicago) and D. Reiman (Department of Bioengineering, University of Illinois at Chicago) for discussions on data analysis. Automated library preparation was performed by the Cellular Screening Center at the University of Chicago. S.T. was awarded a National Institutes of Health (NIH) R01 grant GM127527 and the Paul G. Allen Distinguished Investigator Award, which supported this work. M.C. was awarded NIH R01 grants GM126553 and HG011883, and an NSF grant 2016307, which supported this work.

Author contributions

L.V., H.V.P., C.J., S.T.R. and S.T. conceived of and designed the project. L.V., H.V.P., B.K. and C.J. performed the experiments, designed the Prox-seq probe components and analyzed sequencing data. H.V.P. and M.C. performed statistical analysis. L.V., H.V.P. and S.T. wrote the manuscript. S.T. supervised the project. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41592-022-01684-z>.

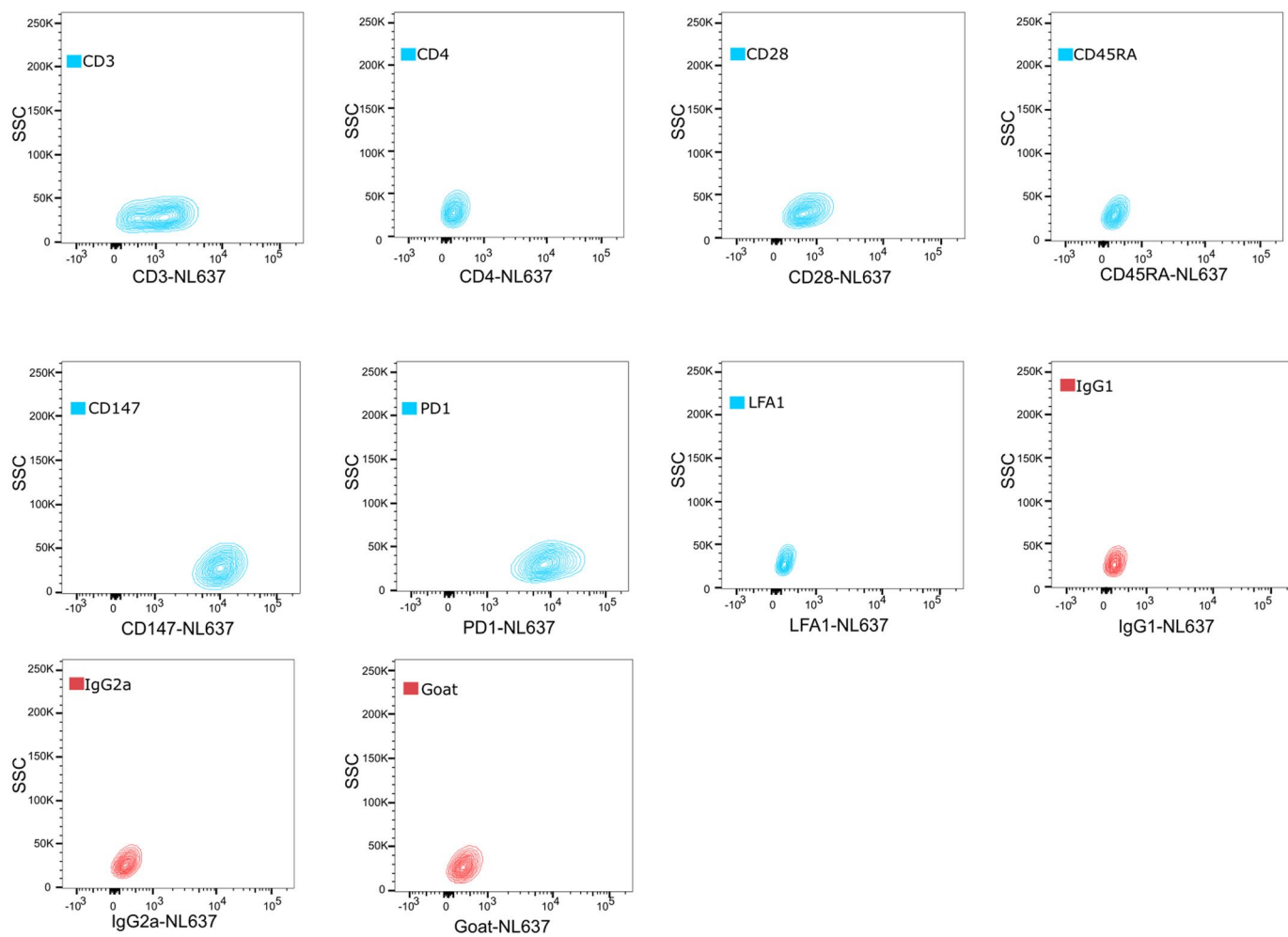
Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41592-022-01684-z>.

Correspondence and requests for materials should be addressed to Savaş Tay.

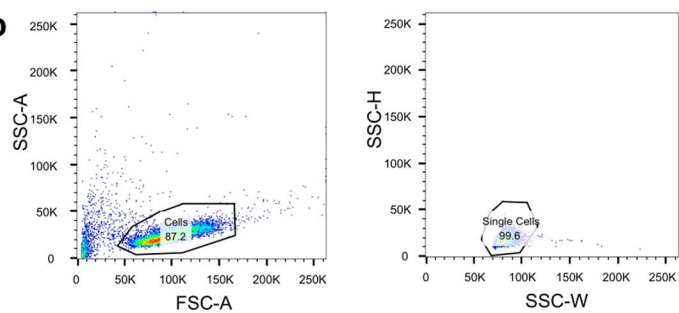
Peer review information *Nature Methods* thanks Nikolai Slavov, Chun Ye, and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editor: Lei Tang, in collaboration with the *Nature Methods* team.

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

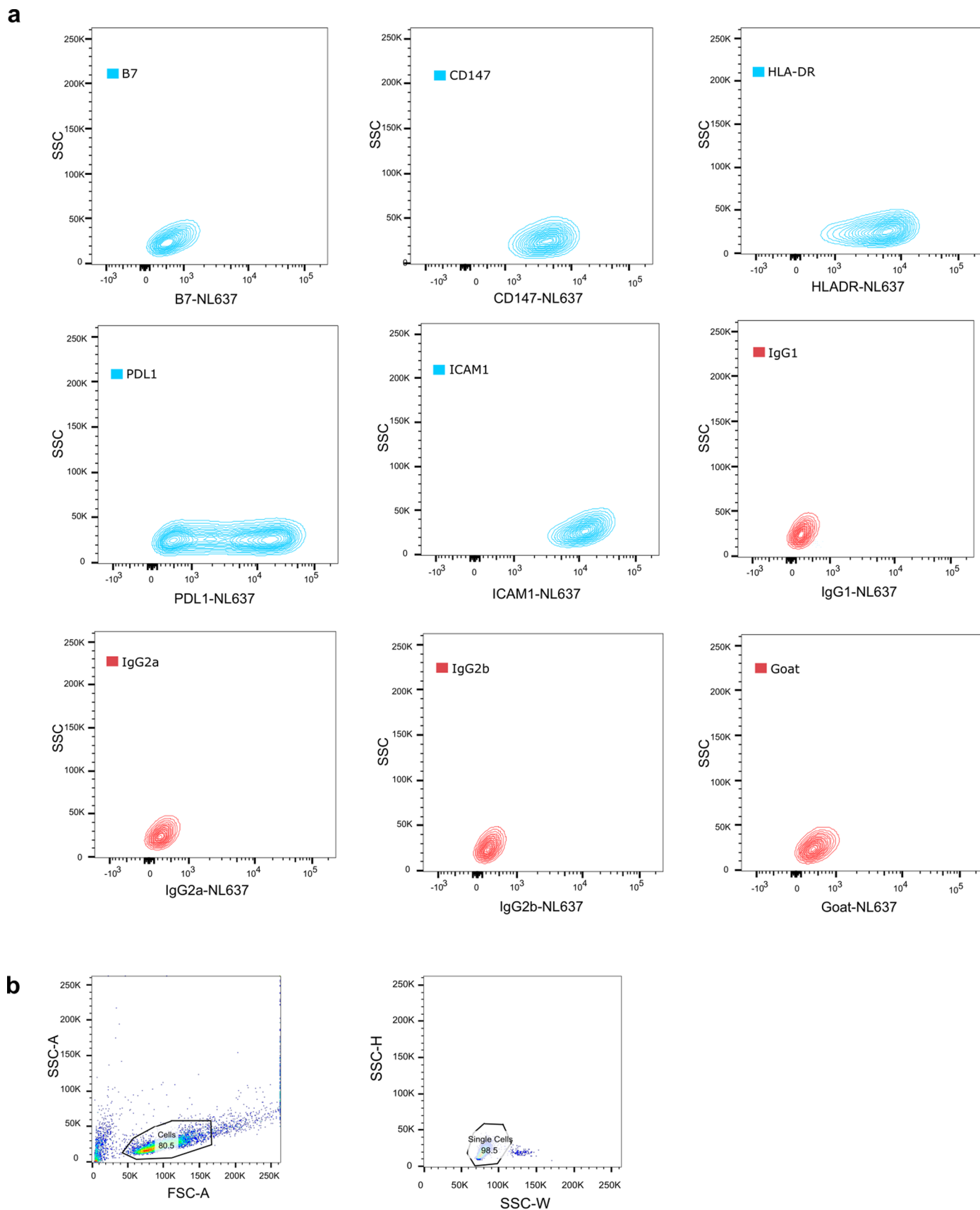
a

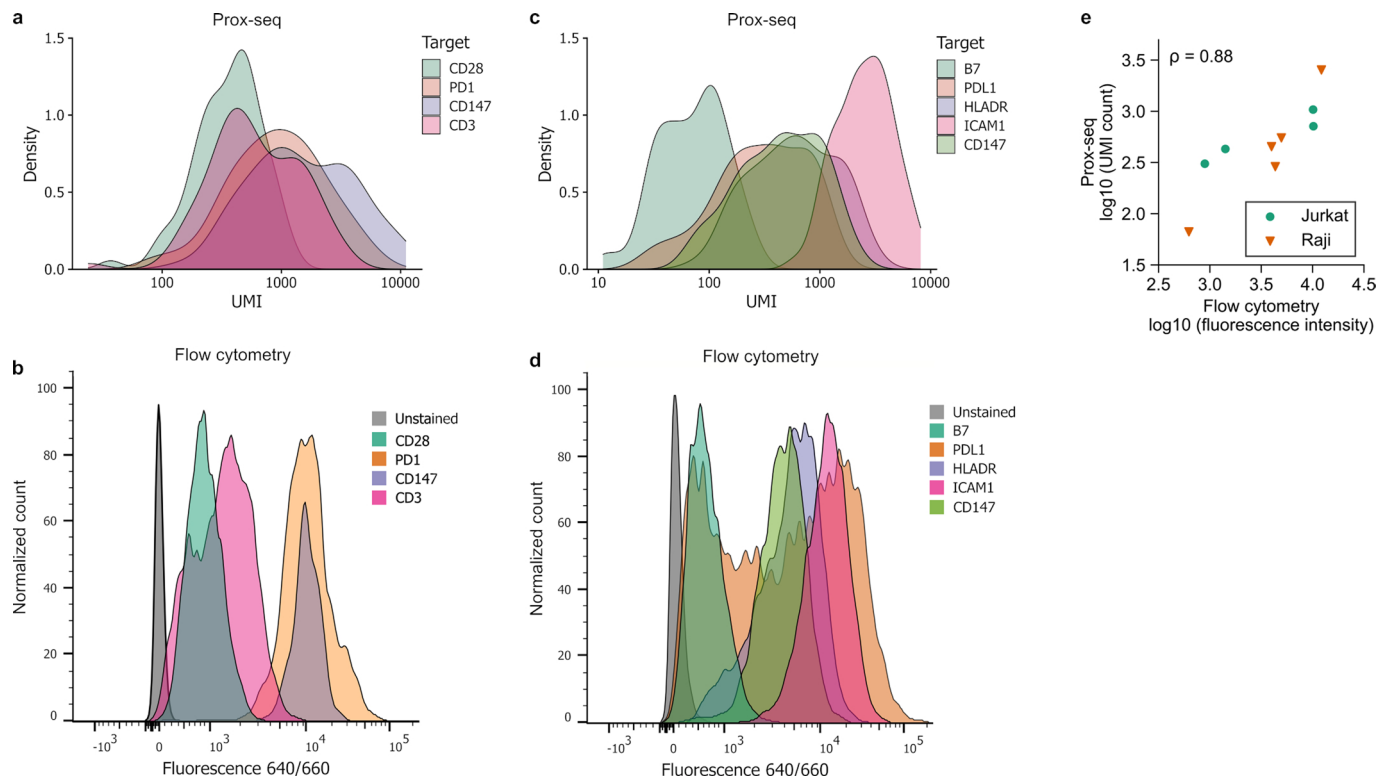


b



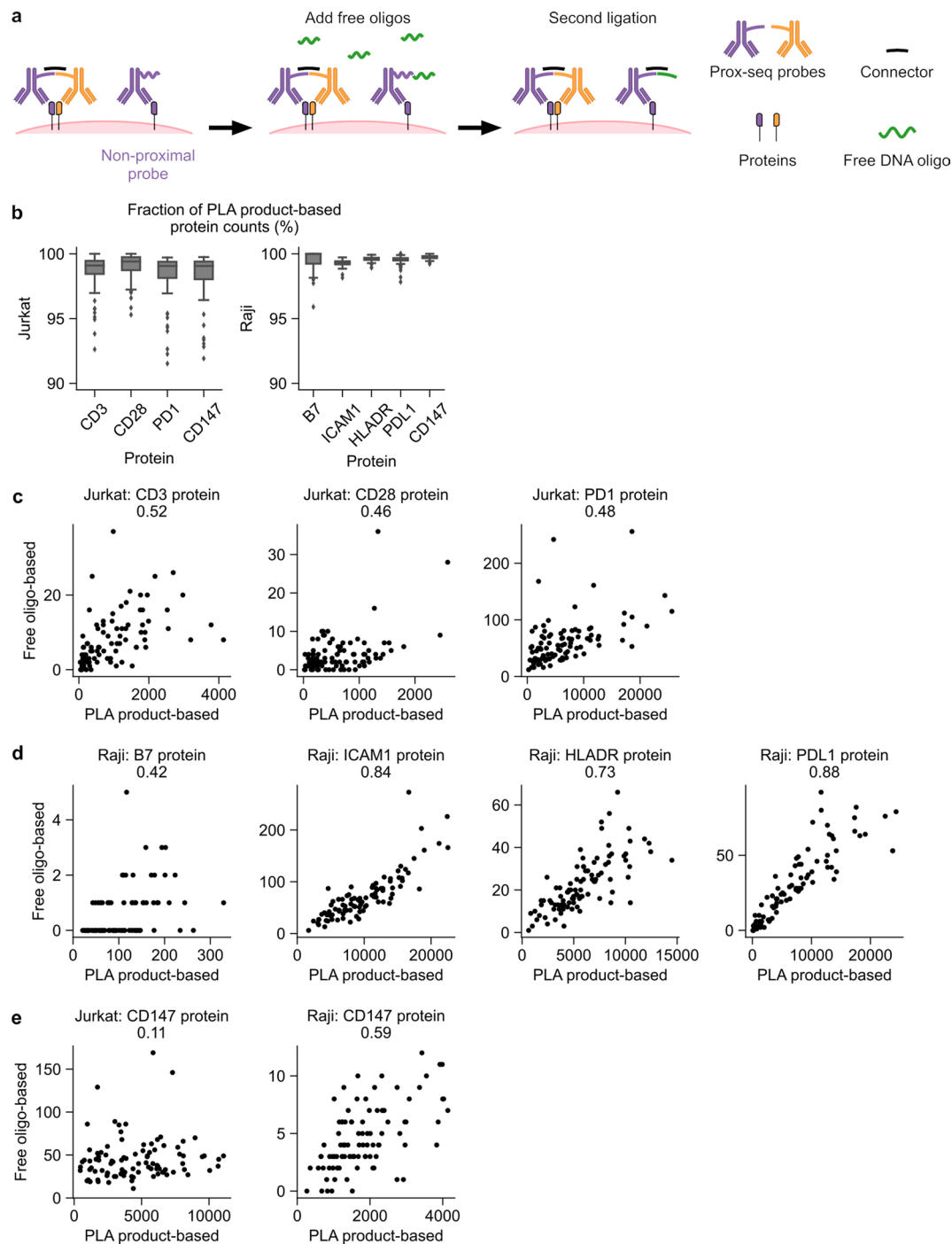
Extended Data Fig. 1 | Jurkat cell protein expression levels. Flow cytometry data showing Prox-seq probe binding on Jurkat cells. (a) Each T cell marker in the panel along with isotype controls. (b) The gating strategy to identify individual cells. CD45RA uses the mouse IgG2a control, CD147 uses the goat control, and the rest use the mouse IgG1 control.





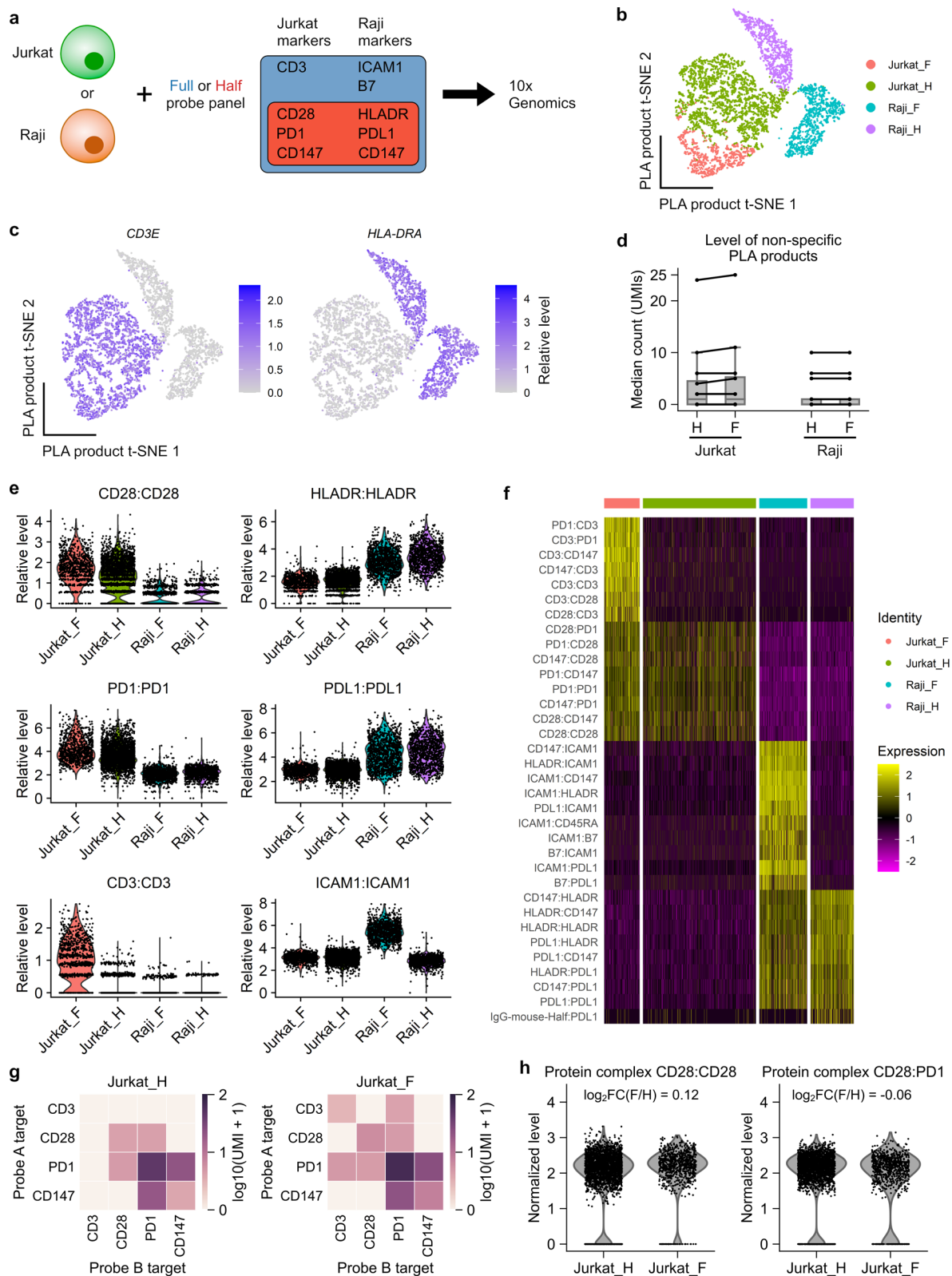
Extended Data Fig. 3 | Comparison of protein quantification between Prox-seq and flow cytometry. (a, b) Distribution of the protein abundance of Jurkat markers as measured by (a) Prox-seq and (b) flow cytometry. (c, d) Distribution of the protein abundance of Raji markers as measured by (c) Prox-seq and (d) flow

cytometry. (e) Scatter plot showing the median protein abundance as measured by flow cytometry or Prox-seq. Each point indicates a protein. The plot also shows the Spearman's correlation coefficient, ρ , between Prox-seq and flow cytometry measurements.



Extended Data Fig. 4 | Benchmarking of protein quantification based on PLA products. (a) Schematic showing how the use of free oligo binding could help measure non-proximal Prox-seq probes. After the ligation step in the standard Prox-seq protocol, free DNA oligos were added so that they can be ligated to probes that are bound to protein but are not proximal to another Prox-seq probe. Antibodies' cartoons were made with BioRender's academic license. (b) Box plots showing the fraction of protein counts calculated from PLA products to the protein counts calculated from PLA products and non-proximal Prox-seq probes ($n = 95$ Jurkat cells and 93 Raji cells). The center line of the box indicates the median, the bottom and top bounds of the box indicate the 25th and 75th

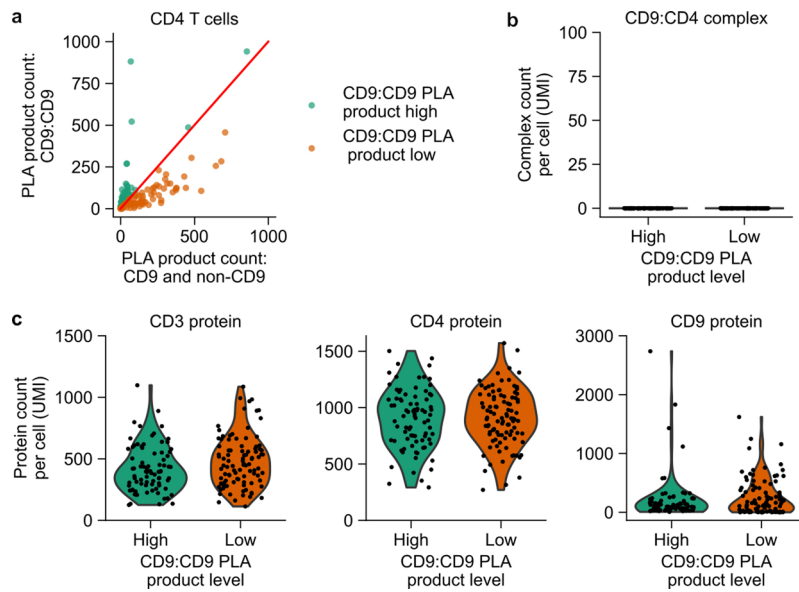
percentiles, and the whiskers extend to $1.5\times$ the interquartile range beyond the box. (c) Scatter plots comparing protein quantification based on PLA products vs. free oligo binding method for Jurkat cell markers. (d) Scatter plots comparing protein quantification based on PLA products vs. free oligo for Raji cell markers. (e) Scatter plots comparing CD147 protein quantification based on PLA products vs. free oligo binding method for Jurkat and Raji cells. In (c–e), the numbers above each panel indicate the Pearson's correlation coefficients. The Free oligo-based estimates are made by taking the number of PLA UMI's that contain one barcode from the indicated protein and one barcode from the free oligo.



Extended Data Fig. 5 | See next page for caption.

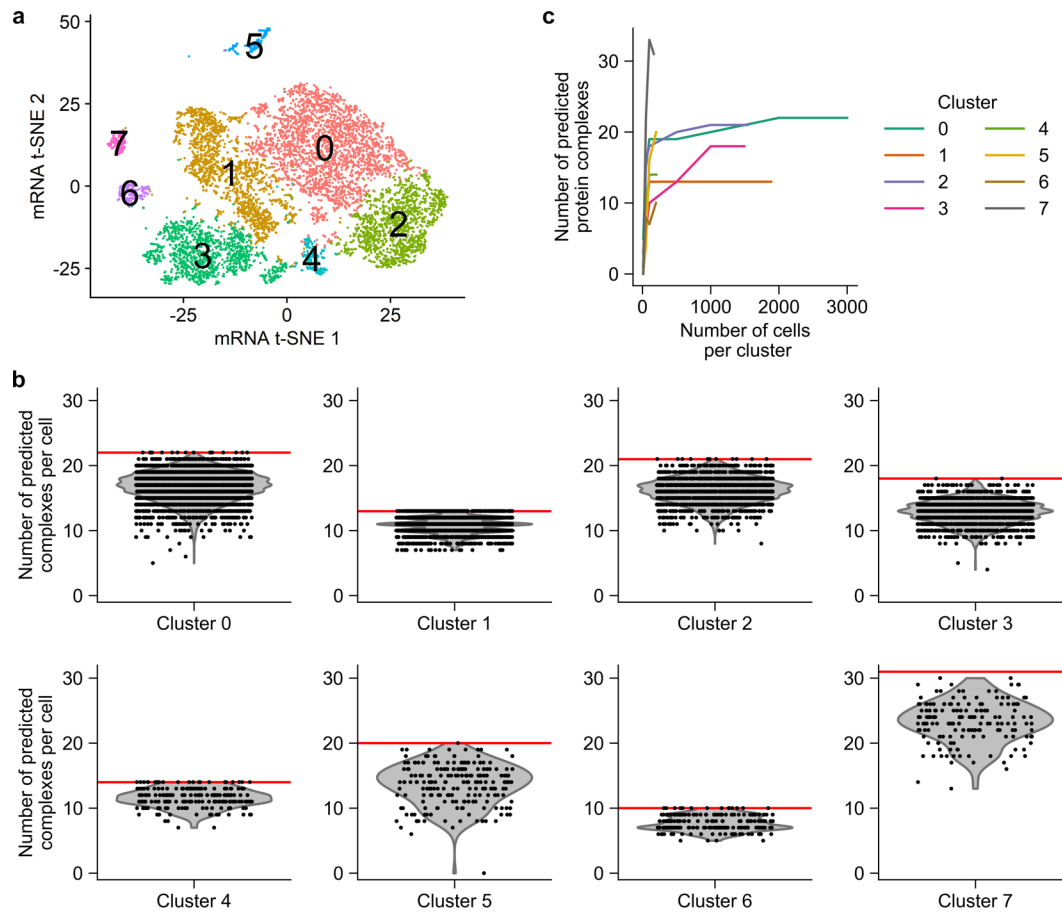
Extended Data Fig. 5 | Analysis of Prox-seq probe non-specific background binding. (a) Schematic of the experiment. Jurkat and Raji cells were separately incubated with the full or half Prox-seq probe panel, then combined and processed with the 10x Prox-seq pipeline. The half probe panel includes Jurkat markers CD28, PDI, and CD147 probes, and Raji markers HLADR and PDL1. The full probe panel contains all the probes in the half panel, plus Jurkat marker CD3 and Raji markers ICAM1 and B7. (b) t-SNE plot based on PLA product data showing the cell type and probe panel identity. F and H stands for full and half panels, respectively (n = 856, 2738, 1159 and 1051 single cells for Jurkat_F, Jurkat_H, Raji_F and Raji_H, respectively). (c) t-SNE plots showing the expression levels of CD3E and HLA-DRA genes. (d) Plots showing the median counts of non-specific PLA products across different cell types and probe panels. The center line of the box indicates the median, the bottom and top bounds of the box indicate the 25th and 75th percentiles, and the whiskers extend to 1.5× the interquartile range

beyond the box. Each black line connects the median counts of a non-specific PLA product in the half and full probe panel. n = 16 non-specific PLA products for both Jurkat and Raji clusters. (e) Violin plots showing the relative levels of Jurkat-specific PLA products CD28:CD28, PDI:PDI and CD3:CD3, and Raji-specific PLA products HLADR:HLADR, PDL1:PDL1 and ICAM1:ICAM1. CD3:CD3 and ICAM1:ICAM1 PLA products were expected to only be detected in full probe panel clusters. (f) Heatmap showing the relative level (row z-score) of top 10 PLA product markers of each of the 4 clusters identified in (c). (g) Heatmaps showing the average log-transformed levels of protein complexes in Jurkat cells. (h) Violin plot showing the normalized levels of protein complexes CD28:CD28 and CD28:PDI in Jurkat cells. The normalized levels were calculated by log-transforming counts per 10,000 UMIs of predicted protein complexes plus a pseudocount of 1. The fold-change was calculated by dividing the average normalized level of the Jurkat full panel cells by that of the Jurkat half panel cells.



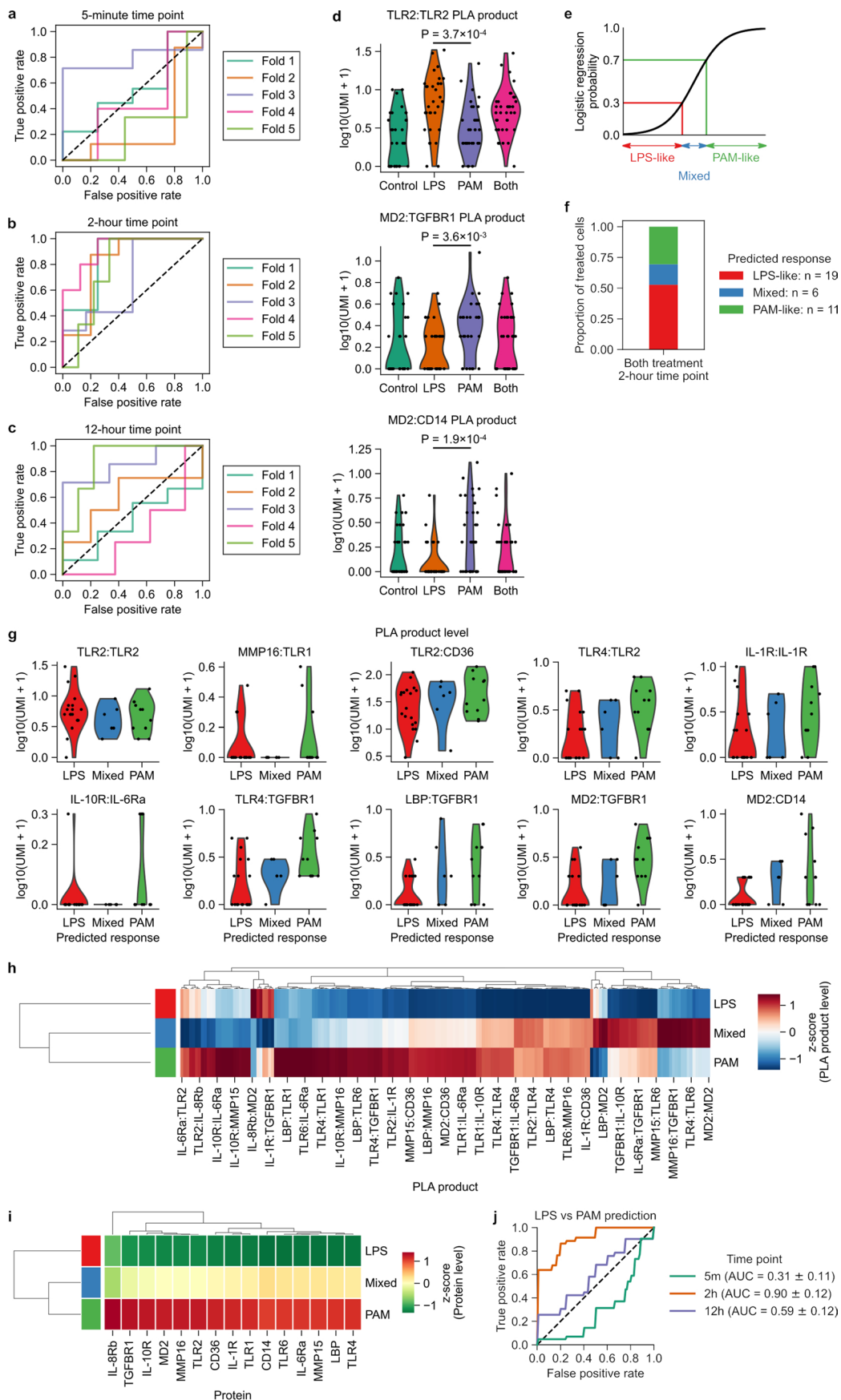
Extended Data Fig. 6 | Characterization of CD4 T cells in PBMCs using plate-based Prox-seq. (a) Scatter plot showing two subpopulations of CD4 T cells, according to CD9-related PLA products level. (b) Violin plot showing that, unlike CD8 T cells, both subpopulations of CD4 T cells do not express the protein

complex CD9:CD8. (c) Violin plots showing the distribution of proteins CD3, CD4 and CD9 in the two subpopulations of CD4 T cells. Note that the complex detection algorithm assigns zero values to low-abundance PLA products that do not pass the statistical test.



Extended Data Fig. 7 | Number of predicted protein complexes across cell types. (a) t-SNE plot of PBMC clusters, obtained using mRNA expression level. (b) Violin plots showing the number of predicted protein complexes per single cell, for each of the 8 clusters identified using mRNA data. The horizontal red lines indicate the total number of predicted protein complexes per cluster.

In total, 61 protein complexes were detected across all 8 clusters, of which 37 complexes are unique. (c) Plot showing the number of protein complexes predicted by the algorithm at different number of cells. Here, various numbers of cells were randomly subsampled from each of the 8 clusters identified in (a), and the complex prediction algorithm was applied on the subsampled cells.

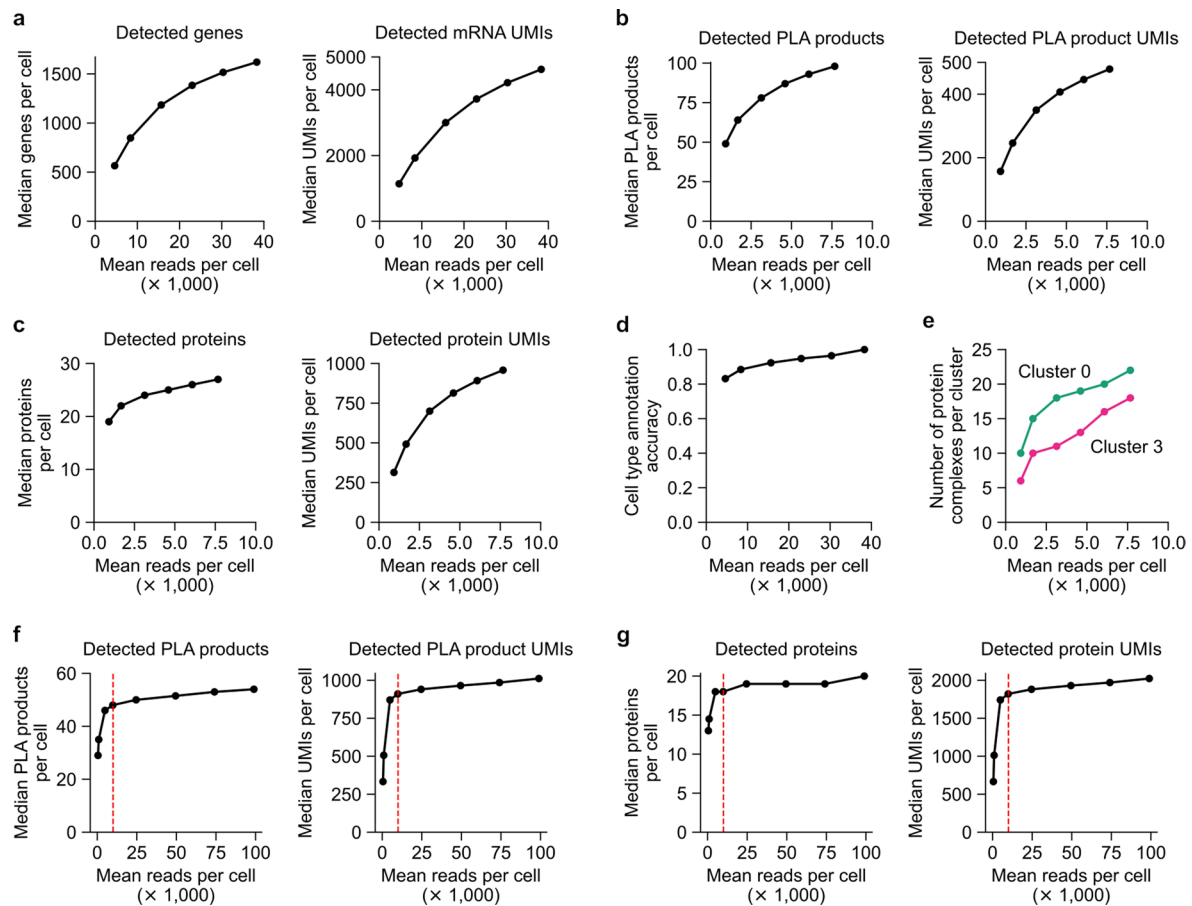


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Analysis of LPS and PAM-treated macrophages.

(a–c) Receiver operating characteristic (ROC) curves of 5-fold cross-validation of a logistic regression classifier that is trained on (a) 5-minute data, (b) 2-hour data, and (c) 12-hour data. The black dashed lines in (a–c) indicate random classification. (d) Violin plots showing the log-transformed count of the top three PLA products of the logistic regression model that is trained on 2-hour data. P-values are calculated using two-sided Welch's t-test ($n = 31, 32, 32,$ and 36 single cells for the control, LPS, PAM, and both treatment groups, respectively). (e) Schematic showing how the logistic regression classifier is used to predict response (LPS-like, PAM-like, and mixed) in cells treated with both LPS and PAM after 2 h. (f) Bar plot showing the proportion of LPS/PAM-treated cells that show LPS-like, PAM-like and mixed response. n indicates the number of cells in each response group. (g) Violin plots showing the log-transformed

count of the top logistic regression coefficients (Fig. 6b) in each predicted response group for cells treated with both ligands. (h, i) Heatmaps showing (h) the relative PLA product levels and (i) the relative protein levels of the LPS-like, PAM-like and mixed response groups. The PLA product (or protein) counts are log-transformed, then averaged by response group, and finally standardized. Hierarchical clustering is performed on the PLA products (or proteins) and response groups using Euclidean distance and complete linkage. (j) ROC curves of a logistic regression classifier trained on protein levels from different time points. The classifier is trained to predict whether a single cell was stimulated with LPS or PAM. Each ROC curve represents the mean ROC curve from 5-fold cross-validation. The area under the curve (AUC) metric of each time point is presented as mean \pm s.d. of the AUC metrics from 5-fold cross-validation for that particular time point.



Extended Data Fig. 9 | Analysis of sequencing depth. (a–c) Effects of sequencing depth on (a) the number of detected genes and transcript counts per single cell, (b) the number of detected PLA products and their UMI counts, and (c) the number of detected proteins and protein UMI counts in 10x-based Prox-seq. (d) Effects of sequencing depth on automated cell type annotation based on mRNA data with singleR package. The cell type annotation at the maximum sequencing depth is used as the ground truth annotation. (e) Effects of sequencing depth on the number of detected protein complexes. Clusters were identified using mRNA data (see Extended Data Fig. 7). Clusters 0 and 3 were chosen as examples because they had the most number of cells per cluster.

In (a–e), the sequencing results of the mRNA and PLA product libraries from the 10x PBMC experiment were downsampled to 10%, 20%, 40%, 60%, and 80% to simulate different sequencing depths. (f, g) Effects of sequencing depth on (f) the number of detected PLA products and their UMI counts, and (g) the number of detected proteins and protein UMI counts in plate-based Prox-seq. In (f, g), the sequencing results of the mRNA and PLA product libraries from the plate-based PBMC experiment were downsampled to 0.5%, 1%, 5%, 10%, 25%, 50%, and 75% to simulate different sequencing depths. The red dashed lines in (e, f) indicate 10,000 mean reads per cell.

Corresponding author(s): Savas Tay

Last updated by author(s): Aug 25, 2022

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 NextSeq 550 system was used to obtain sequencing data.

Data analysis

Drop-seq sequencing data was aligned using Drop-seq tools v2.3.0. 10x sequencing data was aligned using Cell Ranger v6.1.1. Data was analyzed using R v4.2.0, Anaconda/Python v3.8.8, Java v1.8.0_21-b07, Seurat v4.0.4 and v4.1.0, singleR v1.1.0.0, cellDex v1.6.0, numpy v1.20.1, pandas v1.2.4, SciPy v1.6.2, statsmodels v0.12.2, scikit-learn v0.24.1. The custom Java and Python programs are available at <https://github.com/tay-lab/Prox-seq>.

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

The raw sequencing data and processed count data are deposited in NCBI Gene Expression Omnibus (accession numbers GSE149574 and GSE196130).

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](https://www.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 our analyses.	The size of our samples was primarily determined by the number of single cells that could be retrieved. We obtained tens to thousands of cells, which provide a sufficiently large sample for our analyses.
Data exclusions	Cells were filtered based on the number of UMI counts detected, and on the number of genes detected.
Replication	Experiments were done on hundreds to thousands of single cells, providing replicates within each experiment. The protein quantification data was compared with flow cytometry. The exact experimental conditions were tested once. However, various experiments largely overlap. For example, we analyze the same cell line with multiple different single-cell isolation methods (plate-based, Drop-seq, 10x). These near duplicates were all reproducible. When near duplication failed, we found that cell viability was the primary cause.
Randomization	Our single-cell analyses identify cells algorithmically, removing the potential for bias. Furthermore, in most experiments the researchers are unaware of the cell type until this analysis is complete.
Blinding	For drop-seq and 10x data, researchers were blinded to the cell type in question. For plate-based methods researchers knew the identity of the cells in order to validate the method.

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
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	The complete list of antibodies used is displayed in Supplementary Tables 1, 3, 5, 6, 11 and 12. Only secondary antibodies were used as dilutions from the vendor, and this is indicated in the methods section. All antibodies used to make probes were subsequently used at snM. This is indicated in the method section as well.
Validation	Antibodies were validated by the manufacturer for their ability to stain the target protein via flow cytometry. In total, 5 companies supplied antibodies for this study: Biolegend, R&D Systems, EMD Millipore, abcam, and Novus. Validation data can be found on these manufacturers' websites with the antibody catalog number presented in the Supplementary Tables file.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Jurkat and Raji cells were a gift from Prof. Jun Huang. The cells were originally purchased from ATCC.
Authentication	The conclusions we make in the manuscript depend only on the expression level of a few key proteins in our cell lines. Each of these proteins were measured directly via flow cytometry.
Mycoplasma contamination	Not tested
Commonly misidentified lines (See ICLAC register)	None

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 | Samples were taken directly from cell culture without advances pre-processing steps. The Prox-Seq procedure is discussed in detail in the methods section |
| Instrument | Fortessa 4-15 (BD Biosciences) with a High Throughput Screening module |
| Software | Data were collected using FACS Diva and analyzed using FlowJo V_10 |
| Cell population abundance | Cell population abundance was not relevant to this study. |
| Gating strategy | The only gating applied was to isolate single cells. Cells were identified with forward and side scatter, and single cells were identified with height and width if side scatter. |
- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.