Improving oligo-conjugated antibody signal in multimodal single-cell analysis

Terkild Brink Buus[1,2*], Alberto Herrera[1], Ellie Ivanova[1], Eleni Mimitou[3], Anthony Cheng[4,5], Ramin Sedaghat Herati[6], Thales Papagiannakopoulos[1], Peter Smibert[3], Niels Ødum[2], & Sergei B. Koralov[1*]

[1]Department of Pathology, New York University School of Medicine, New York, NY, USA.
[2]LEO Foundation Skin Immunology Research Center, Department of Immunology & Microbiology, University of Copenhagen, Copenhagen, Denmark.
[3]Technology Innovation Lab, New York Genome Center, New York, NY, USA.
[4]Department of Genetic and Genome Sciences, University of Connecticut School of Medicine, Farmington, CT, USA.
[5]Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA, USA.
[6]NYU Langone Vaccine Center, Department of Medicine, New York University School of Medicine, New York, NY

* Correspondence to: Terkild Brink Buus (Terkild.Buus@sund.ku.dk) or Sergei B. Koralov (Sergei.Koralov@nyulangone.org)

Abstract

Simultaneous measurement of surface proteins and gene expression within single cells using oligo-conjugated antibodies offers high resolution snapshots of complex cell populations. Signal from oligo-conjugated antibodies is quantified by high-throughput sequencing and is highly scalable and sensitive. In this study, we investigated the response of oligo-conjugated antibodies towards four variables: Concentration, staining volume, cell number at staining, and tissue. We find that staining with recommended antibody concentrations cause unnecessarily high background and that concentrations can be drastically reduced without loss of biological information. Reducing volume only affects antibodies targeting abundant epitopes used at low concentrations and is counteracted by reducing cell numbers. Adjusting concentrations increases signal, lowers background and reduces costs. Background signal can account for a major fraction of the total sequencing and is primarily derived from antibodies used at high concentrations. Together, this study provides new insight into the titration response and background of oligo-conjugated antibodies and offers concrete guidelines on how such panels can be improved.

Impact statement

Oligo-conjugated antibodies are a powerful tool but require thorough optimization to reduce background signal, increase sensitivity, and reduce sequencing costs.

Introduction

Analysis of surface proteins in multimodal single-cell genomics such as cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) is a powerful addition to conventional single-cell RNA sequencing (scRNA-seq) [1-3]. Unlike flow- and mass cytometry, CITE-seq is not limited by spectral overlap nor availability of distinguishable isotopes [4, 5]. This is due to the practically unlimited number of distinct oligo barcodes and discrete sequence counting, allowing high numbers of antibodies to be included in individual experiments.

While signal acquisition in CITE-seq is different, the reagents and staining procedure is highly analogous to staining for flow cytometry. Traditional titration for flow or mass cytometry aims to identify the conjugated antibody concentration allowing the best discrimination between the signal from positive and negative cells [4, 6]. Multiple factors may affect antibody binding and subsequent signal including antibody concentration, total amount of antibody, as well as the level of target expression (epitope amount). Epitope amount is governed by the number of cells and the per-cell expression of the target epitope. These factors are in turn influenced by the cellular composition of the sample as well as their activation and differentiation state. Nonspecific binding is expected to increase as the total amount of antibody molecules greatly exceed the epitopes present in a sample. As such, nonspecific binding is dependent on the total number of antibody molecules rather than the antibody concentration [5]. This makes staining volume, cell composition and cell number important parameters for optimal staining [6]. Consequently, flow and mass cytometric optimization aims to use antibody concentrations that reaches the highest signal to noise ratio (often reached at the “saturation plateau”) in a minimal volume (and thus minimal number of antibody molecules)[4, 7].

Oligo-conjugated antibody signal has been shown to be highly analogous to fluorochrome-conjugated antibodies of the same clone in flow cytometry in regards to the concentration needed to reach the “saturation plateau” [8]. However, unlike flow cytometry, where antibody (fluorescence) signal intensity has no influence on analysis cost, oligo-conjugated antibody signal is analyzed by counting sequencing reads, making costs strictly dependent on signal intensity (by requiring increased sequencing depth). This is particularly important for methods sequencing vast numbers of cells stained with a high number of antibodies such as single cell combinatorial indexed cytometry by sequencing (SCITO-seq), where shallow sequencing is paramount for the economic feasibility of such methods [9]. Thus, while an optimal antibody concentration in flow cytometry aims to get the highest signal-to-noise ratio, oligo-conjugated antibody staining conditions should be titrated to get sufficient signal-to-noise at the lowest possible signal intensity. In practice, this means that concentrations of most antibodies in

an optimized CITE-seq panel are not intended to reach their "saturation plateau" but should be within their linear concentration range (where doubling the antibody concentration leads to twice the signal). Such concentrations are much more sensitive to the number of available epitopes (i.e. cell number and cell composition) than an optimized flow cytometry panel. Unlike flow and mass cytometry where the major source of background is autofluorescence, spillover between neighboring channels and nonspecific binding of the antibodies [5, 10], a major source of background signal for oligo-conjugated antibodies appears to be free-floating antibodies in the cell suspension [11]. In droplet-based single-cell sequencing methods, these free-floating antibodies will be distributed between cell-containing and empty droplets. As signal from empty droplets can only be distinguished from signal from cell-containing droplets after sequencing and due to the much higher number of empty than cell-containing droplets, background signal can make up a considerable fraction of the sequenced reads, and thus sequencing costs.

In this study, we present a limited but practically applicable titration of four variables in a 5'-CITE-seq panel of 52 antibodies: 1. Antibody concentration (four-fold dilution response), 2. Staining volume (50 µl vs 25 µl), 3. Cell count ($1x10^6$ vs. $0.2x10^6$) and 4. Tissue of origin: peripheral blood mononuclear cells (PBMCs) from healthy donor versus immune cell compartment from a lung tumor sample. We find that oligo-conjugated antibodies show high background and little response to titration when used above 2.5 µg/mL and that most antibodies appear to reach their saturation plateau at concentrations between 0.62 and 2.5 µg/mL. Many antibodies can be further diluted despite being at their linear concentration range without affecting the identification of epitope-positive cells. Reducing staining volume has a minor effect on signal and only impacts signal from antibodies used at low concentrations targeting highly expressed epitopes; this effect is counteracted by reducing the number of cells present during staining. We compare samples stained with pre-titration and adjusted concentrations of the same antibody panel and find that adjusting concentrations increase signal, lowers background, and reduces both sequencing and antibody costs. Finally, we find that background signal in empty droplets can constitute a major fraction of the total sequencing reads and is skewed towards antibodies used at high concentrations targeting epitopes present in low amounts.

Results

Four-fold antibody dilution in PBMC and lung tumor immune cells

A panel of 52 oligo-conjugated antibodies was allocated into several groups of starting concentrations based on previous experience with each antibody or targeted epitope abundance by CITE-seq, flow cytometry and vendor recommendations (ranging from 0.05 to 10 µg/mL; Supplementary Table S1). We stained two samples of either 10^6 PBMCs or $5x10^5$ lung tumor leukocytes in 50µl of antibody mixture with various starting concentrations, hereafter referred to as "Dilution Factor (DF) 1". To determine how the signal from each antibody changed by dilution across the two tissues, we stained the same number of cells in the same volume with a four-times diluted antibody mixture (DF4).

Single-cell gene expression was assessed by shallow sequencing (~4,000 reads per cell) to assign cells into major cell lineages (Figure 1A) and cell types (Figure 1B) based on their transcriptional profile (see Supplementary Figure S1 for gene detection and UMI distributions and details on cell type annotation). Leukocytes from lung tumor samples exhibited distinct transcriptional profiles within each cell type but showed overall good co-clustering with similar cell types (Figure 1C). To allow direct comparison of unique molecular identifier (UMI) counts from the different conditions, we reduced the number of cells included in analysis from each condition to contain the same number of cells from each cell type. By only using the gene expression modality for cell type assignment, we can directly compare antibody-derived tag (ADT) UMI counts at different staining conditions within transcriptional sub-clusters without risk of having differences in ADT signal interfere with cell type assignment.

Comparing the total ADT UMI counts from each condition, we saw fewer UMIs from samples stained with DF4 as compared with DF1, both at 77 % sequencing saturation (Figure 1D). However, the reduction in UMI counts from DF1 to DF4 by 38% (761,350 to 474,404) and 51% (1,121,940 to 548,393) in PBMC and Lung, respectively, was markedly less than the four-fold difference (75% reduction) in antibody concentrations used in staining. It is worth noting, that 4/52 antibodies used at the highest concentration (10µg/mL) accounted for more than 20% of the total UMI counts irrespective of tissues and dilution factors and without showing any clearly positive populations (Figure 1D, E; gating thresholds shown in Supplementary Figure S2). Indeed, we found that the majority of antibodies used in concentrations at or above 2.5 µg/mL showed minimal response to four-fold titration, both in terms of total UMI counts (Figure 1F) as well as UMI counts at the 90th quantile of the cell type with the highest overall expression level (Figure 1G; expressing cell types identified in Figure 1E), reflecting the response within the positive population where such could be identified. In contrast, antibodies used in concentrations at or below 0.62 µg/mL all showed close to linear response to four-

fold dilution (shown as a reduction around two “logs” on a log2 scale; Figure 1F, G). This indicates that the signal for many antibodies reach their saturation plateau in the range between 0.62 and 2.5 µg/mL, and that higher concentrations are likely to only increase the background signal.

In the present panel, the response to four-fold dilution can be divided into five categories (Figure 2 and Supplementary Figure S3), that warrant different considerations in the choice of whether to reduce concentration or not. For category A (Fig 2A) reducing concentration is always the right choice. For the other categories (Figure 2B-E), the choice of whether to reduce concentration or not, depends on the balance between the need for signal and the economic cost of signal (see Table 1).

Reducing staining volume primarily affects highly expressed markers

To investigate the effect on ADT signal caused by further reducing the staining volume, we included PBMC samples stained with the same concentration of antibodies in 50 µl or 25 µl (effectively using half the amount of antibodies at twice the cell density). In both samples, we used the DF4 panel on 10^6 cells to assess the worst-case scenario of the reduction, as the amount of epitopes in this setting are likely to be competing for antibodies that are not in excess. Despite having many antibodies responding linearly to concentration reduction (Figure 1), we found much less response to reduced staining volume, both in regard to total number of UMIs (9% reduced; 469,541 to 428,680) and on a marker by marker basis (Figure 3A-C). As expected, antibodies used in low concentrations (0.0125 to 0.025 µg/mL) targeting highly abundant epitopes were most severely affected by the reduced staining volume (such as CD31, CD44 and CD45; Figure 3D, E and Supplementary Figure S4), whereas antibodies targeting less abundant epitopes were largely unaffected (such as CD8 and CD19; Fig 3F).

Reducing cell number during staining increases signal for antibodies at low concentration

To determine if the limited effect of reduced staining volume on ADT signal could be counteracted by simultaneously reducing the number of cells at the time of staining (effectively reducing the total amount of epitopes), we analyzed two PBMC samples with either $1x10^6$ or $0.2x10^6$ cells stained with the same concentration of antibodies (DF4) in 25µl. Similar to reducing staining volume, the majority of the included antibodies were largely unchanged by lowering the cell density at staining, as reflected by only a minor 8% increase in detected UMIs (from 428,680 to 462,916), and also reflected by the analogous distribution of individual markers (Figure 4A-C). Encouragingly, reducing the cell number

at staining increased the signal from the antibodies used at low concentration and targeting highly expressed epitopes (Figure 3D, E and Supplementary Figure S4), thus largely mitigating the loss of signal observed when the staining volume was reduced from 50 to 25 µl (Figure 3B-D and Supplementary Figure S5). Interestingly, despite reducing the cell density at staining 5-fold (from 40 to $8x10^6$ cells/mL) the resulting signal did largely not supersede that of the sample stained in 50 µl with an intermediate cell density of $20x10^6$ cells/mL (Supplementary Figure S6).

Adjusting antibody concentration improves signal, lowers background and reduces cost and sequencing requirements

To evaluate the benefits of adjusting antibody concentrations, we stained $200x10^3$ PBMCs in a staining volume of 25µl using the same antibody panel, with individual antibody concentrations adjusted based on their assigned categories (individual concentrations can be found in Supplementary Table S1 and how each category was adjusted is described in Table 1). On average the adjusted panel used 1.9 fold less antibody than the DF1 staining and 8.4 fold less than the vendor recommended starting concentration (Supplementary Table S2). Together with the reduced staining volume this decreased antibody costs per sample to 50 USD which is a 3.9 and 33.6 fold reduction from DF1 (195 USD) and vendor recommendations (1690 USD), respectively (based on list price of 325 USD per 10µg; Supplementary Table S2).

To allow direct comparison with the DF1 sample, we integrated and down-sampled the DF1 and adjusted samples to include similar numbers cells within each cell type (Figure 5A). We then down-sampled the sequenced ADT reads to yield similar UMI totals of 522,469 and 521,331 across the comparable cell populations for the DF1 and adjusted sample, respectively (Figure 5B). As expected, antibodies used at reduced concentrations yielded relatively fewer UMIs (Category A, B and some from E) whereas increased concentrations yielded more (Category E and some from C). Importantly, we found that antibodies with unchanged concentration yielded more UMIs at similar sequencing depth (Figure 5B, C). This is primarily due to a reduction of Category A antibodies which accounted for 25% of the sequenced UMI sequences in the DF1 sample and only 10% in the adjusted sample.

Due to the cost-of-signal in these sequencing-based approaches, an optimal panel would, ideally use similar number of UMIs per positive cell for each antibody (Figure 5D) and exhibit approximately the same positive signal (UMIs above background; Figure 5E). While some markers should be further reduced (such as CD4, CD45RA, CD62L, CD107a and TCRab) and some adjustments were too

extreme (such as CD14, CD19 and HLA-ABC), the adjusted sample exhibited close to a two-fold increase in the median UMIs per positive cell and a 57 % increase in the median positive signal (from 7 to 11 UMIs; Figure 5E). Importantly, all the markers with the lowest positive signal as well as number of UMIs per positive cell were all increased, reflecting a more balanced sequencing library.

Importantly, while exhibiting approximately the same relative background signal as assayed by proportion of reads within empty droplets (35-45%; data not shown), the adjusted sample generally showed much lower percentage of UMIs being assigned to background (Figure 5F). This was particularly remarkable for CD86 which went from 76.5 % to 12.6 % and thus yielded comparable positive signal while using 4.8 fold fewer UMIs (from 23,971 to 4,998; Figure 5G). In fact, the exception to this was primarily found within Category E antibodies for which concentrations were increased due to having very low UMI counts in the DF1 sample (CD56, CD127 and CD196; see Supplementary Figure S8 for data on all markers). In these cases, the increased concentration yielded better definition of expected positive populations (Figure 5H). To balance the sequencing requirements of the panel, we reduced concentrations of most category B antibodies. Except CD19 (Figure 5I), all reduced category B antibodies showed no change in resolution of positive versus negative populations despite a marked reduction in their UMI usage (Figure 5C, D) and concomitant reduction their positive signal. For instance, when reducing anti-CD5 from 0.62 to 0.16 µg/mL, it showed largely identical distribution despite using 65% less UMIs (from 22,068 to 7,740; Figure 5J). Category C and E antibodies showed consistently increased positive signal (Figure 5D) and consequently allowed better identification of populations known to express these markers, such as naïve T cells and monocytes for CD45RA and CD45RO, respectively (Figure 5K, L). Individual titration plots for each marker can be found in Supplementary Figure S7.

Background signal from oligo-conjugated antibodies is dependent on antibody concentration and abundances of epitopes

Free-floating antibodies in the solution has been shown to be one of the major contributors to background signal for ADT [11]. Similar to cell-free RNA, background ADT signal can be assayed from empty droplets. To determine the background signal of the different antibodies in our panel, we split the captured barcodes into cell-containing and empty droplets based on the inflection point of the barcode-rank plot for the gene expression UMI counts (Supplementary Figure S8). Despite being a “super-loaded” 10X Chromium run targeting 20,000 cells, the number of empty droplets vastly

outnumber the cell-containing droplets. Consequently, several antibodies exhibited more cumulated UMIs within empty droplets than within cell-containing droplets (Figure 6A). This was particularly prevalent within antibodies used at concentration at or above 2.5 µg/mL, thus drastically skewing the frequency of these antibodies within the empty droplets as compared with cell-containing droplets (Figure 6A, B). Conversely, antibodies targeting highly abundant epitopes were enriched within cell-containing droplets, irrespective of their staining concentration (such as CD44 and CD107a, HLA-ABC, HLA-DR; Figure 6C). Enrichment of antibodies targeting abundant epitopes (such as CD3, CD4, CD8 and CD45RA) within the cell-containing droplets despite high numbers of UMIs within empty droplets was also observed within three publicly available datasets using an identical 17 antibody panel on 1,000 or 10,000 cells using two different capture approaches (3'- and 5' capture; Supplementary Figure S9). We found that ADT signal in empty droplets (i.e. background) was highly correlated with the UMI cutoff for detection (Figure 6D; Supplementary Figure S10). Markers with low background generally showed low UMI cutoff and exhibited high dynamic range allowing identification of multiple levels of expression (as seen for CD4 and CD19; Figure 6D, E). In contrast, markers with high background showed high UMI cutoff regardless of whether they exhibited cell type-specific signal (such as CD86 and CD279; Figure 6F) or whether their positive signal was absent or obscured by the high background (such as TCRγδ; Figure 6G).

Discussion

In this study, we show that titration of oligo-conjugated antibodies for multimodal single-cell analysis can improve the sensitivity, lower background signal and reduce costs and sequencing requirements, and that such optimizations go beyond (and even against) the need to reach the "saturation plateau". We show that for a representative panel of 52 antibodies, most antibodies used in concentrations at or above 2.5 µg/mL show high background signal and minimal loss in sensitivity upon a four-fold reduction in concentration. Antibodies used at concentrations between 0.625 and 2.5 µg/mL show limited (non-linear) response whereas most antibodies used at concentrations below 0.625 µg/mL show linear or close to linear response. It should be noted, that these estimates may be inherently biased given that the starting concentrations were based on our prior experience with the individual antibody clones and our assumptions regarding abundance of targeted epitopes. This has favored using higher concentrations for antibodies known to have low performance and for antibodies with unknown performance. Nonetheless, for antibodies with unknown performance, our results highlight the benefits of conducting titration experiments or initially using the antibodies at concentrations in the 0.625 to 2.5

µg/mL range rather than the 5 to 10 µg/mL range recommended by published antibody staining protocols and by commercial vendors. This is particularly important when adding new antibodies to existing panels, where antibodies added in a high concentration may account for a drastically disproportionate usage of the total sequencing reads without providing any biological information (as seen for CD86, CD152, CD183, CD197 and TCRγδ in the DF1 panel). Our results also show that concentrations of antibodies targeting highly expressed epitopes can be further reduced without affecting resolution of positive and negative cells, even when these antibodies are already used within their linear concentration range (such as CD5, CD8 and CD19). By reducing the concentration of these antibodies, the allocation of reads to each antibody becomes more balanced between more and less abundant epitopes allowing the overall sequencing depth to be reduced and maximizing the yield of a sequencing run.

By using varying starting concentrations based on prior experience and by titrating the full panel together, our study does not necessarily identify the antibody concentration providing the optimal signal-to-noise ratio. This could have been achieved by using saturating starting concentrations and additional serial dilutions as has been previously done for a few markers [8]. However, due to the cost of signal in these cytometry-by-sequencing methods, using all antibodies at their highest signal-to-noise ratios would require much deeper sequencing, as highly expressed markers would use the vast majority of the total sequencing reads. Instead, we aimed to get sufficient signal-to-noise while keeping the sequencing allocated to each marker balanced. A further complication for titration experiments that start with saturating amounts of antibody, is the observation that background signal can be largely attributed to free-floating antibodies in the solution. Thus, using high concentrations for all markers in one or more condition would increase the background in all conditions if these were multiplexed into the same droplet segregation. This would likely obscure the positive signals and possibly titration response at lower concentrations (similar to what we see for category A antibodies). To avoid this, each condition would have to be run in its own droplet segregation, making traditional titration experiments prohibitively costly.

In this study, we used commercially available antibody clones that have been extensively used for other applications such as flow and mass cytometry and we do see high concordance between ADT signals and the expected antigens within each cell type. Our approach did not allow us to formally test whether each antibody is specific to its intended antigen, as we inferred based on our understanding included cell types and looked for concordance with gene expression signature of the cells. However, it should

be noted that, when using antibody clones that are unfamiliar or have not undergone extensive testing, it is important to assure their specificity.

Reducing staining volume for 10^6 PBMCs from 50 µl to 25 µl only showed minor effect on signal and this minimal impact was primarily observed for antibodies used at very low concentrations (0.0125 to 0.025 µg/mL) targeting highly expressed epitopes (such as CD31, CD44 and CD45). This effect was readily counteracted by concomitantly reducing the number of cells at staining to $0.2x10^6$ PBMCs in 25 µl. In flow cytometry, while the binding of antibody is strictly dependent on its concentration, background signal is dependent on the ratio between the total amounts of antibody and epitopes [6]. Consequently, background can be reduced by increasing the number of cells (increasing the amount of epitope) or decreasing staining volume (effectively reducing the amount of antibody without changing its concentration). For antibodies optimized to reach their "saturation plateau" (common in flow cytometry), both of these approaches can be applied without changing the true signal. In contrast, for oligo-conjugated antibodies used in sequencing based single-cell approaches, operating in the linear range, signal from highly abundant epitopes stained with low concentration of antibody will be affected. In such cases, the cells can be stained in multiple steps adjusting the staining volume while keeping the concentration the same – i.e. staining in a smaller volume for antibodies with high background and subsequently staining antibodies at low concentration in a higher volume. In this regard, when multiplexing samples, pre-staining each sample with hashtags and pooling prior to staining with additional CITE-seq antibodies may provide multiple advantages: 1. All samples are stained at the same time with the exact same antibody mixture – making cross-sample comparison more accurate. 2. By having more cells in a smaller total volume, less total antibody is used in the presence of more epitopes conceivably reducing the background signal. 3. Samples where cell number at staining is a limiting factor, such as small tissue biopsies, will be exposed to the same local concentrations of antibody as more abundant samples (such as PBMCs) removing potential differences between samples by antibodies being "sponged" by differences in overall epitope abundance. However, this approach is only available when all samples are similarly affected by the staining procedure and can tolerate the additional washes needed (after both hashing and CITE-seq staining).

We compared ADT signal from PBMCs stained with the same antibody panel at the starting concentration with a sample stained at concentrations adjusted following the titration experiment. While some markers could benefit from further adjustments, the sample stained with the adjusted panel was more balanced in its distribution of sequencing reads among markers, having twice the median UMIs per positive cell. Despite intentionally reducing signal in category B antibodies, we found an

overall 57 % increase in the median positive signal. Concomitantly, the adjusted panel exhibited 43% lower background signal (median of 26.3 % to 14.9 % UMIs assigned to background) despite increasing the concentrations of many category C and E antibodies. Consequently, the adjusted concentrations greatly improved the overall performance of the panel. We took precautions to make the samples as comparable as possible by down-sampling the sequencing depth to the same level and comparing similar numbers of similar cells (at the mRNA level). Nonetheless, as these samples were from different preparations and different donors, we cannot exclude that some of the observed differences can be attributed to these factors. For instance, we found that the monocytes in the adjusted sample exhibited higher unspecific binding (as seen from the isotype controls) than in the DF1 sample, despite being treated similarly with Fc-blocking reagents (which should minimize such biding [12]).

Due to the 10-1000 fold higher numbers of proteins as compared to coding mRNA [13], ADT libraries have high library complexity (unique UMI content) and are rarely sequenced near saturation. Thus, either sequencing deeper or squandering less reads on a few antibodies increase signal from all (other) included antibodies. We found that by simply reducing the concentration of the five antibodies used at 10 µg/mL, we gained 17 % more reads for the remaining antibodies. Consequently, assuming we are satisfied with the magnitude of signal we got from all other antibodies using the starting concentration, this directly translates to a 17 % reduction in sequencing costs. Due to different antibodies being adjusted in different directions for different reasons (according to their assigned categories) it is difficult to convert the overall improved utilization of sequencing reads into exact savings calculation. However, assuming signal is improved or unchanged, the savings on sequencing for each marker can be estimated by how many UMIs are needed to acquire a given signal. In the case of CD86, we found that the signal was dramatically improved by reducing concentration from 10 to 0.667 µg/mL while also using 79 % fewer UMIs and consequently a much lower number of sequencing reads.

Empty droplets have been shown to be useful for determining the background signal of CITE-seq [11]. This suggests that the major source of background signal for ADT libraries can be attributed to free-floating antibodies (or oligos) in the solution, rather than unspecific antibody binding to cell surfaces. In the present study, the samples were multiplexed by hashing antibodies and pooled after oligo-conjugated antibody staining and then run in the same 10X Chromium lane. This effectively obscures the contribution of each sample to the total amount of free-floating antibodies in the final cell suspension which is conceivably skewed towards the samples stained in high volume with the highest concentration of antibodies – as these samples contain the highest total amount of antibody. Consequently, as free-floating antibodies are the major source of background, this would explain why

we do not observe reduced background in the cell stained at the lowest concentrations (i.e. dilution factor 4). As such, for markers with no specific signal due to high background (such as CD183, CD197 and TCRgd), the titration responses may be underestimated due to specific signal being lost within the high background. This also means that for markers with high background signal, our proposed reductions in concentrations are conservative, as we would expect to see decreased background in samples stained with reduced amount of antibodies (as seen in the comparison with the adjusted concentrations). In droplet-based single-cell analyses, background signal is not only diminishing the sensitivity and resolution of true signals but is also a major contributor to sequencing cost of ADT libraries. Due to empty droplets vastly outnumbering cell-containing droplets, we found that ADT signal from empty droplets can easily account for 20-50% of the total sequencing reads and consequently 20-50% of the sequencing cost. The number of antibodies used in CITE-seq-related platforms is only expected to expand. Additionally, the numbers cells included in each experiment is continuously being increased (as seen for methods such as SCITO-seq [9]). As such, reducing background signal from oligo-conjugated antibodies should be a priority. The source of the free-floating antibodies is not completely understood. Observations from this study suggest that antibodies used at high concentration targeting absent or sparse epitopes are highly enriched within the empty droplets as compared with the cell-containing droplets. This indicates that residual unbound antibody from the staining step is a major contributor, despite several washing steps. Practically, this suggests that additional washing after cell staining would be beneficial when the number and type of cells in the samples allow it. Optimal washing is achieved by repeated washing steps while assuring that maximal residual supernatant is removed after each centrifugation and followed by gentle but complete resuspension in a large buffer volume.

More and more advanced CITE-seq-related cytometry-by-sequencing platforms are rapidly being developed. However, while these platforms utilize different methods to assure single-cell resolution, and use different approaches to label the cells, they all use high-throughput sequencing to count signal from a variety of oligo-conjugated probes (such as antibodies with both surface and intra-cellular targets, MHC-peptide multimers, B-cell receptor antigens etc.) [1, 2, 9, 14-18]. Most of the observations, results and conclusions from this study will thus also be applicable to a variety of platforms where improving oligo-conjugated probe signal is critical for ensuring their broad utility and economic feasibility.

Materials and methods

Clinical samples

Lung adenocarcinoma patient sample (female, 57 year old, former smoker: 15 pack years, treated with chemotherapy) was collected at New York University Langone Health Medical Center in accordance with protocols approved by the New York University School of Medicine Institutional Review Board and Bellevue Facility Research Review Committee (IRB#: i15-01162 and S16-00122).

Cell isolation, cryopreservation and thawing

Peripheral blood mononuclear cells (PBMCs) were isolated from a leukopak and whole blood from healthy donors (New York Blood Center) for the pre-titration and adjusted samples, respectively. PBMCs were purified by diluting in PBS and subsequent gradient centrifugation using Ficoll-Paque PLUS (GE Healthcare) and 50 mL conical tubes (Falcon). PBMCs in the interphase were collected and washed twice with PBS 2% FBS. Lung tumor sample were cut into small pieces with a razor blade and enzymatically digested (100 U/mL Collagenase IV, Sigma-Aldrich, C5138-1G; 50 µg/mL DNase 1, Worthington, LS002138) for 35 minutes being rotated at 37°C in HEPES buffered RPMI 1640 containing 0.5% FBS. After digestion, the sample was forced through a 100 µm cell strainer to make a single-cell suspension. Single-cell suspensions from both PBMCs and lung tumor were cryopreserved in freezing medium (40% RPMI 1640, 50% FBS and 10% DMSO) and stored in liquid nitrogen. On the day of the experiments, cryopreserved samples were thawed for 1-2 minutes in a 37°C water bath, washed twice in warm PBS containing 2% FBS and re-suspended in complete media (RPMI 1640 supplemented with 10% FBS and 2mM L-Glut).

Oligo-conjugated antibody staining

We modified the published protocol for ECCITE-seq [19] to stain cells in round-bottom 96-well plates (as is common practice for flow cytometry staining in many laboratories). This allowed us to reduce staining volumes and centrifugation time analogous to staining for flow cytometry. After thawing, the intended number of cells were resuspended in 12.5 µl or 25 µl of CITE-seq staining buffer (2% BSA, 0.01% Tween in PBS) for samples stained in a total of 25 µl or 50 µl, respectively. To prevent antibody binding to Fc receptors [12], Fc receptor block from two vendors (TruStain FcX, BioLegend and FcR blocking reagent, Miltenyi) was added to the suspension and incubated for 10 minutes on ice. During incubation, the antibody solution of 52 TotalSeqC antibodies (BioLegend; Supplementary Table S1) was washed on a pre-wet Amicon Ultra-0.5 Centrifugal Filter to remove sodium azide. The volume of the resulting antibody pool was adjusted to 2X of final concentrations and 12.5 µl or 25 µl was added to the cells to achieve a total staining volume of 25 µl or 50 µl, respectively. 10µg/mL of a unique

hashing antibody was added to each sample and incubated for 30 min on ice. After staining, cells were washed four times in 1x150 µl and 3x200 µl CITE-seq staining buffer.

Super-loading of 10X chromium

Individually hashed samples were counted using a hemocytometer and pooled in equal ratio at high concentration. Pooled sample was strained through a 70µm cell strainer and counted again using a hemocytometer. To achieve approximately 20,000 cells after doublet removal, cell concentration was adjusted to 1314 cells/µl to achieve the target of 41,645 cells in 31.7µl for super-loading of the 10X Chromium Chip A. Gene expression (using 5' v1 chemistry; 10X Genomics) and Antibody Derived Tag (ADT) and hashtag-oligo (HTO) libraries were constructed using reagents, primers and protocol from the published ECCITE-seq protocol [19]. All libraries from the titration run were sequenced together with other samples on an Illumina NovaSeq6000 S1 flow cell. The post-titration (adjusted) sample (using 5' v1.1 chemistry; 10X Genomics) was multiplexed and sequenced together with other samples not included in this study on Illumina NovaSeq6000 SP and S1 flow cells.

Alignment and counting of single-cell sequencing libraries

The multiplexed gene expression library was aligned using kallisto (v0.46)-bustools (v0.39.0) [8, 9]. Given the polyA selection inherent in the 10X genomics protocol, reads were aligned against a reference transcriptome based on the GTF file included in the Cell Ranger software (refdata-cellranger-GRCh38-3.0.0/genes/genes.gtf; 10X Genomics) that does not include as many non-polyA transcripts as the human transcriptome included by kallisto-bustools by default. From the 77,507,446 reads assigned to the gene expression library 66.9% aligned to the transcriptome. ADT and HTO libraries were counted using the *kallisto indexing and tag extraction* (KITE) workflow (https://github.com/pachterlab/kite) resulting in 82,527,351 and 65,875,774 counted reads, respectively. Number of UMIs and genes detected per cell across cell lineages can be found in Supplementary Figure S1A, B.

Single-cell demultiplexing, preprocessing and down-sampling

To allow detection of UMI counts within non-cell-containing droplets, unfiltered count matrices from each modality was loaded into a 'Seurat' (v3.1.4) object [20]. Samples were demultiplexed by their unique hashtag oligos (HTO) using the Seurat function 'MULTIseqDemux' yielding 19,560 demultiplexed cells. This allowed the removal of 3,724 (19%) cross-sample doublets. Due to the shallow sequencing of the mRNA library (~4000 reads/cell), expression at least 60 genes and a percent mitochondrial reads below 15 % were used to remove barcodes from non-viable cells or debris (2,499

or 15% of cells removed). Intra-sample doublets were removed using the 'scDblFinder' (v1.1.8) R package (392 cells removed). UMI counts from antibody derived tags (ADTs) were normalized using default configuration of the DSB (v0.1.0) R package with ADT signal from HTO-negative droplets used as empty drop matrix and using included isotype controls [11]. Gene expression was preprocessed using the default Seurat v3 pipeline and fine-grained clusters were identified using the 'FindClusters' function with a resolution of 1.2. Clusters were annotated by lineages and cell types using their distinct expression of markers within the mRNA or (ADT) modality and aided by cell-by-cell annotation from the SingleR R package (v1.4.0) using the "Monaco reference" from the celldex R package (v1.0.0) made from bulk RNA-seq samples of sorted immune cell populations from GSE107011 [21]. Top 5 differentially expressed marker genes for each cluster can be found in Supplementary Figure S1D. To allow direct comparison of UMI counts across conditions, each condition was down-sampled by tissue of origin to include the same number of cells within each fine-grained cell type cluster (resulting in 1,777 cells from each PBMC sample and 1,681 cells from each Lung tumor sample).

Integration and sub-sampling for pre- and post-titration comparison

The post-titration (adjusted) sample was pre-processed as described above. Together with the DF1 sample, the adjusted sample was normalized and integrated based on their mRNA expression using the SCTransform and IntegrateData functions from the Seurat package as described in the Seurat integration vignette [20, 22]. After mRNA-based clustering using FindClusters at resolution 1.2, similar number of comparable cells was selected by taking the nearest neighbors in PCA-space for each cell in the sample with the fewest cells within the given cluster. This sampling assured that similar number of comparable cells (at the mRNA level) were selected for comparison, thus minimizing the effect of the sample differences. To allow direct comparison of UMI counts and eliminate differences in sequencing depth as a factor, we down-sampled the FASTQ files from the ADT modality of the adjusted sample to achieve similar totals of UMIs within the DF1 (522,469) and Adjusted (521,331) samples.

Comparing ADT signal from cell-containing and empty droplets

For comparison of UMI counts within cell-containing and non-cell-containing (empty) droplets for the present dataset and the 10X Genomics datasets, we divided the unfiltered count matrices by the inflection point in their ranked per cell UMI sum from the mRNA library. Barcodes above the inflection point were then used to extract UMI counts within cell-containing droplets from each antibody oligo modality. All UMIs that were not included in cell-containing droplets were considered from empty droplets.

Data and code availability

All code and commands used to process the data and to generate all plots and figures are available at GitHub: https://github.com/Terkild/CITE-seq_optimization

UMI count matrices from the optimization experiment have been deposited at FigShare with DOI: https://doi.org/10.6084/m9.figshare.c.5018987. The feature barcode 3' and 5' VDJ 10X datasets are available from the 10X Genomics website.

Conflict of interest

PS is co-inventor of a patent related to the single cell technology utilized in this study (US provisional patent application 62/515–180).

Acknowledgements

Work in Dr. Koralov's laboratory was supported by the NIH R01 grant (HL-125816), LEO Foundation Grant (LF-OC-20-000351), NYU Cancer Center Pilot Grant (P30CA016087), the Judith and Stewart Colton Center for Autoimmunity Pilot grant, and a grant from the Drs. Martin and Dorothy Spatz Foundation. TBB and NØ are supported by the Danish Cancer Society (Kræftens Bekæmpelse), the Danish Council for Independent Research (Danmarks Frie Forskningsfond) and the LEO Foundation. We thank NYU Genome Technology Center for technical assistance and support and acknowledge NYU Center for Biospecimen Research & Development and NYU Perlmutter Cancer Center for their support in acquiring patient biospecimens.

Contributions:

TBB, AH, EI and SBK initiated the project and were responsible for experiment design with input from EPM, RH, TP, NØ and PS. TBB and AH performed the experiments. TBB performed the bioinformatic data analysis with input from AC, EPM, PS and SBK. TBB, AH, EI and SBK were responsible for the writing, with all authors providing input on the manuscript.

References

1. Stoeckius, M., et al., *Simultaneous epitope and transcriptome measurement in single cells.* Nat Methods, 2017. **14**(9): p. 865-868.
2. Peterson, V.M., et al., *Multiplexed quantification of proteins and transcripts in single cells.* Nat Biotechnol, 2017. **35**(10): p. 936-939.
3. Mair, F., et al., *A Targeted Multi-omic Analysis Approach Measures Protein Expression and Low-Abundance Transcripts on the Single-Cell Level.* Cell Rep, 2020. **31**(1): p. 107499.
4. Gullaksen, S.E., et al., *Titrating Complex Mass Cytometry Panels.* Cytometry A, 2019. **95**(7): p. 792-796.
5. Hulspas, R., et al., *Considerations for the control of background fluorescence in clinical flow cytometry.* Cytometry B Clin Cytom, 2009. **76**(6): p. 355-64.
6. Hulspas, R., *Titration of fluorochrome-conjugated antibodies for labeling cell surface markers on live cells.* Curr Protoc Cytom, 2010. **Chapter 6**: p. Unit 6 29.
7. van Vreden, C., et al., *Titration of Mass Cytometry Reagents.* Methods Mol Biol, 2019. **1989**: p. 83-92.
8. Stoeckius, M., et al., *Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics.* Genome Biol, 2018. **19**(1): p. 224.
9. Hwang, B., et al., *SCITO-seq: single-cell combinatorial indexed cytometry sequencing.* bioRxiv, 2020: p. 2020.03.27.012633.
10. Au-Yeung, A., et al., *Visualization of Mass Cytometry Signal Background to Enable Optimal Core Panel Customization and Signal Threshold Gating.* Methods Mol Biol, 2019. **1989**: p. 35-45.
11. Mulè, M.P., A.J. Martins, and J.S. Tsang, *Normalizing and denoising protein expression data from droplet-based single cell profiling.* bioRxiv, 2020: p. 2020.02.24.963603.
12. Andersen, M.N., et al., *Elimination of erroneous results in flow cytometry caused by antibody binding to Fc receptors on human monocytes and macrophages.* Cytometry A, 2016. **89**(11): p. 1001-1009.
13. Marguerat, S., et al., *Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells.* Cell, 2012. **151**(3): p. 671-83.
14. Setliff, I., et al., *High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity.* Cell, 2019. **179**(7): p. 1636-1646 e15.
15. O'Huallachain, M., et al., *Ultra-high throughput single-cell analysis of proteins and RNAs by split-pool synthesis.* Commun Biol, 2020. **3**(1): p. 213.
16. Overall, S.A., et al., *High throughput pMHC-I tetramer library production using chaperone-mediated peptide exchange.* Nat Commun, 2020. **11**(1): p. 1909.
17. Gaublomme, J.T., et al., *Nuclei multiplexing with barcoded antibodies for single-nucleus genomics.* Nat Commun, 2019. **10**(1): p. 2907.
18. Katzenelenbogen, Y., et al., *Coupled scRNA-Seq and Intracellular Protein Activity Reveal an Immunosuppressive Role of TREM2 in Cancer.* Cell, 2020. **182**(4): p. 872-885 e19.
19. Mimitou, E.P., et al., *Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells.* Nat Methods, 2019. **16**(5): p. 409-412.
20. Stuart, T., et al., *Comprehensive Integration of Single-Cell Data.* Cell, 2019. **177**(7): p. 1888-1902 e21.
21. Monaco, G., et al., *RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types.* Cell Rep, 2019. **26**(6): p. 1627-1640 e7.
22. Hafemeister, C. and R. Satija, *Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression.* Genome Biol, 2019. **20**(1): p. 296.

Tables

Table 1: Five categories of response to four-fold dilution

Category	Response to 4-fold dilution	Markers	Considerations
Category A (Figure 2A)	Antibodies exhibiting no response to dilution indicating that the positive signal is full saturated, absent, or obscured within high background signal.	*CD1a, CD30, CD86, CD134, CD138, CD152, CD183, CD197, CD279, CD336, IgG1, IgG2A, and TCRgd*	Reducing antibody concentration is always the right choice. These antibodies sequester a large amount of UMIs without any biological information. Reducing concentration may reveal a true positive population obscured by the background signal.
Category B (Figure 2B)	Antibodies that respond by a reduction in signal but without hampering the ability to distinguish positive and negative fractions.	*CD4, CD5, CD8, CD11b, CD19, CD62L, CD69, CD103, CD107a, CD194, CD274, EpCAM, HLA-DR, and TCRab*	Reducing antibody concentrations will be economically beneficial with minimal loss of biological information. For instance, In the Lung at DF1, HLA-DR uses 9% of the total UMI counts within cell-containing droplets and can be reduced at least four-fold without any apparent change in ability to discriminate between positive and negative cells.
Category C (Figure 2C)	Antibodies that respond by a reduction in signal which subsequently changes the ability to distinguish positive from negative cells or bring the cutoff value for positive cells down to only a few UMIs.	*CD1c, CD2, CD3, CD14, CD25, CD26, CD28, CD31, CD39, CD45RA, CD45RO, and CD141*	Reducing antibody concentration will reduce biological information as cells expressing the targeted epitopes may not exhibit sufficient signal. If only cells expressing high levels of the marker need to exhibit signal, these can be slightly reduced.
Category D (Figure 2D)	Antibodies that respond linearly to titration but take up high numbers of UMIs due to targeting (almost) ubiquitously expressed markers.	*CD44, CD45 and HLA-ABC*	These can be reduced if all cells exhibit high UMI counts. Unless these markers have a clear purpose, most experiments will benefit from dropping them from the panel as they tend to sequester a large proportion of total sequencing reads with little biological information.
Category E (Figure 2E)	Antibodies where response is hard to assess due to not showing expected positive population either due to lack of epitopes (no positive cells in either tissue) or lack of antibody binding (non-functional antibody).	*CD24, CD56, CD66b, CD70, CD80, CD117, CD123, CD127, CD196, and CD223*	Should be evaluated individually. Do you have prior information indicating that this marker is expressed by cells in these types of samples? Do any cells in the sample express high levels of the gene encoding the targeted protein? If so, increasing the concentration of the antibody or trying a different clone may yield better signal.

Figures and Figure Legends

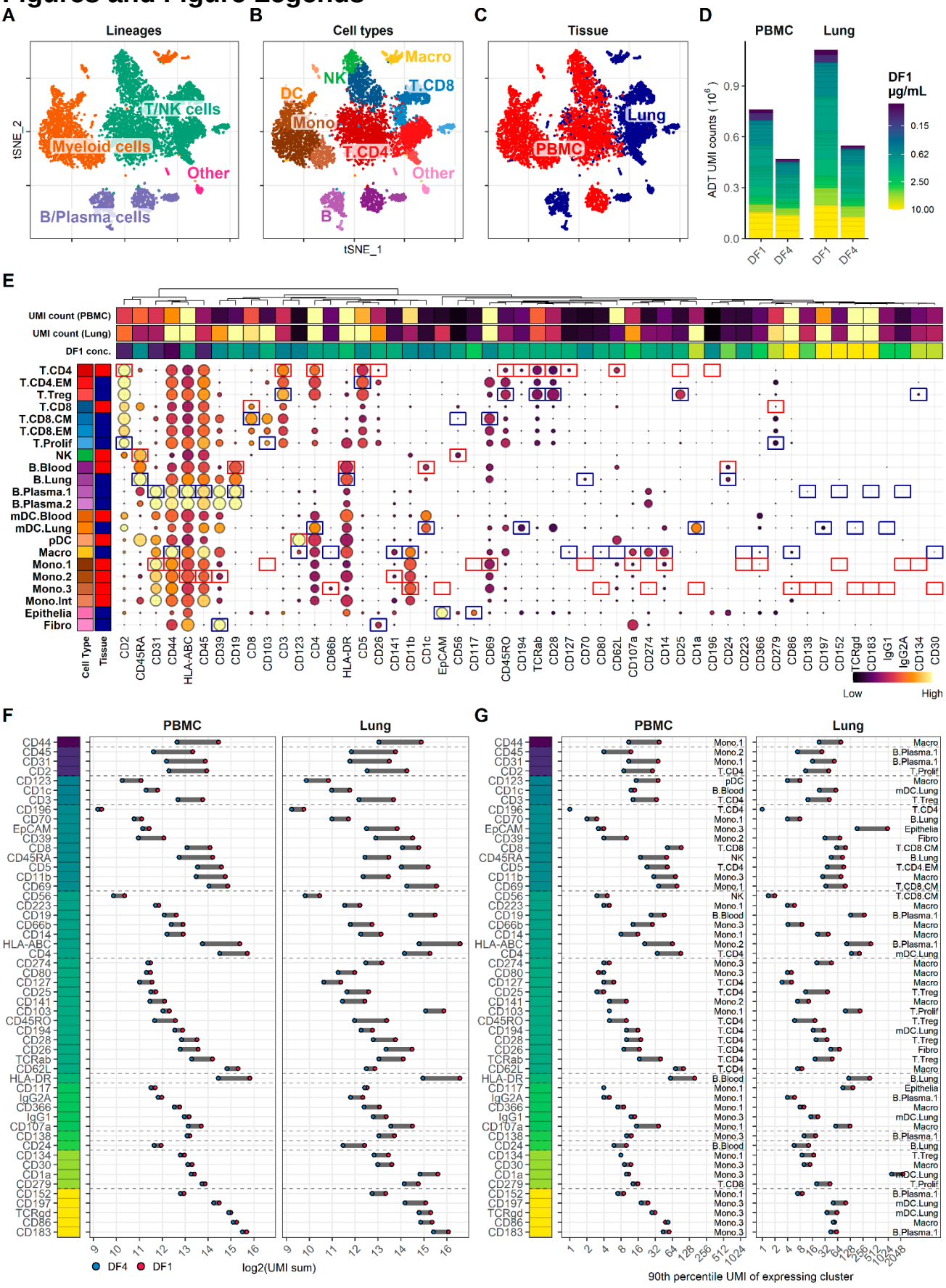


Figure 1: Four-fold antibody dilution response in PBMC and lung tumor immune cells.
A-C. Single cells from all samples and conditions were clustered and visualized according to their gene expression and colored by (A) overall cell lineage, (B) cell type, and (C) tissue of origin. D. Summarized UMI counts within cell-containing droplets segmented by the individual antibodies stained at the starting concentrations (dilution factor 1; DF1) or at a four-fold dilution (DF4) in PBMC and Lung samples (concentrations of each antibody can be found in Supplementary Table S1). Antibody segments are colored by their concentration at DF1. E. Heatmap of normalized antibody-derived tag (ADT) signal within each transcription-based cluster identified in B. Visualized by frequency of positive cells (circle size) and colored by the median ADT signal within the positive fraction (i.e. signal from a marker that is highly expressed by all cells in a cluster will have the biggest circle and be colored yellow). Red and blue colored boxes denote the clusters chosen for evaluating titration response within blood and lung samples, respectively. F, G. Change in ADT signal for each antibody by four-fold dilution. Individual antibodies are colored by their concentration at DF1 and quantified by (F) sum of UMIs within cell-containing droplets assigned to each antibody and (G) 90th percentile UMI count within expressing cell cluster identified in E and annotated by numbers to the right.

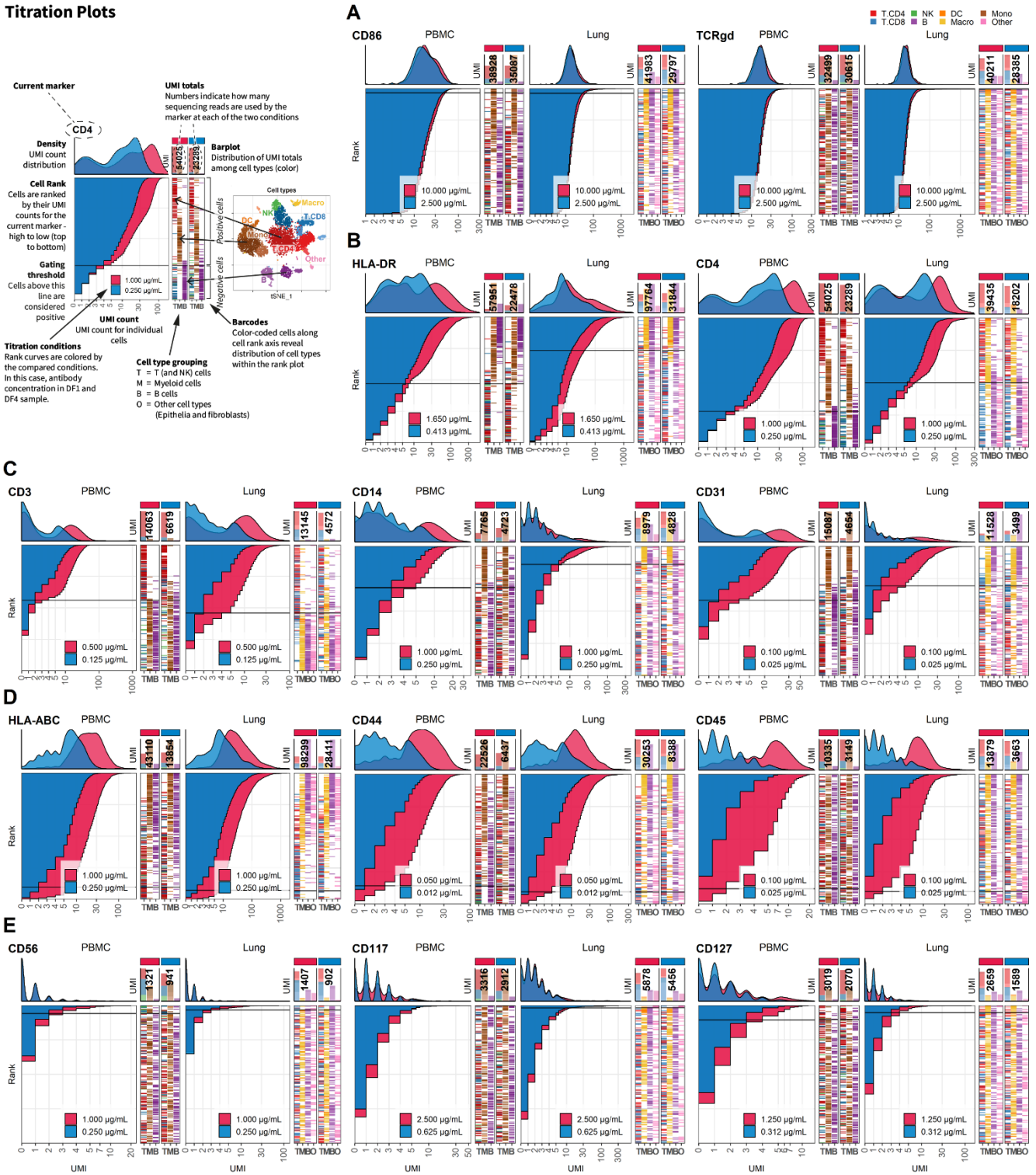


Figure 2: Four-fold antibody dilution response is dependent on epitope abundance.
Titration plots (marker UMI count vs. cell rank) showing response to reduction in antibody concentration from dilution factor 1 (DF1) to DF4 within PBMC (left) and Lung (right). Histogram depicts distribution of UMIs at each condition colored by dilution factor (and annotated with concentration). Numbers within bar plot denote total UMI count within cell-containing droplets at each antibody concentration. Barcodes to the right depicts cell type by color at the corresponding rank to visualize specificity of the antibody. Horizontal line depicts gating threshold for cells considered positive for the marker. Antibody response to four-fold dilution can be divided into five categories exemplified in **A-E**. **A**. Antibodies where the positive signal is obscured within the

background signal (Category A). **B**. Antibodies that respond by a reduction in signal but without hampering the ability to distinguish positive from negative cells (Category B). These antibodies also show strict cell type specificity (i.e. HLA-DR is restricted to non-T cells whereas CD4 is highly expressed in T cells and intermediately expressed in myeloid cells as shown in the barcode plot) **C**. Antibodies that respond by a reduction in both signal and changes the ability to distinguish positive from negative cells (Category C). **D**. Antibodies targeting ubiquitously expressed markers (Category D). **E**. Antibodies that do not show a convincing positive population due to either lack of epitopes (no positive cells in either tissue) or lack of antibody binding (non-functional antibody) (Category E). Titration plots for all markers can be found in Supplementary Figure S3.

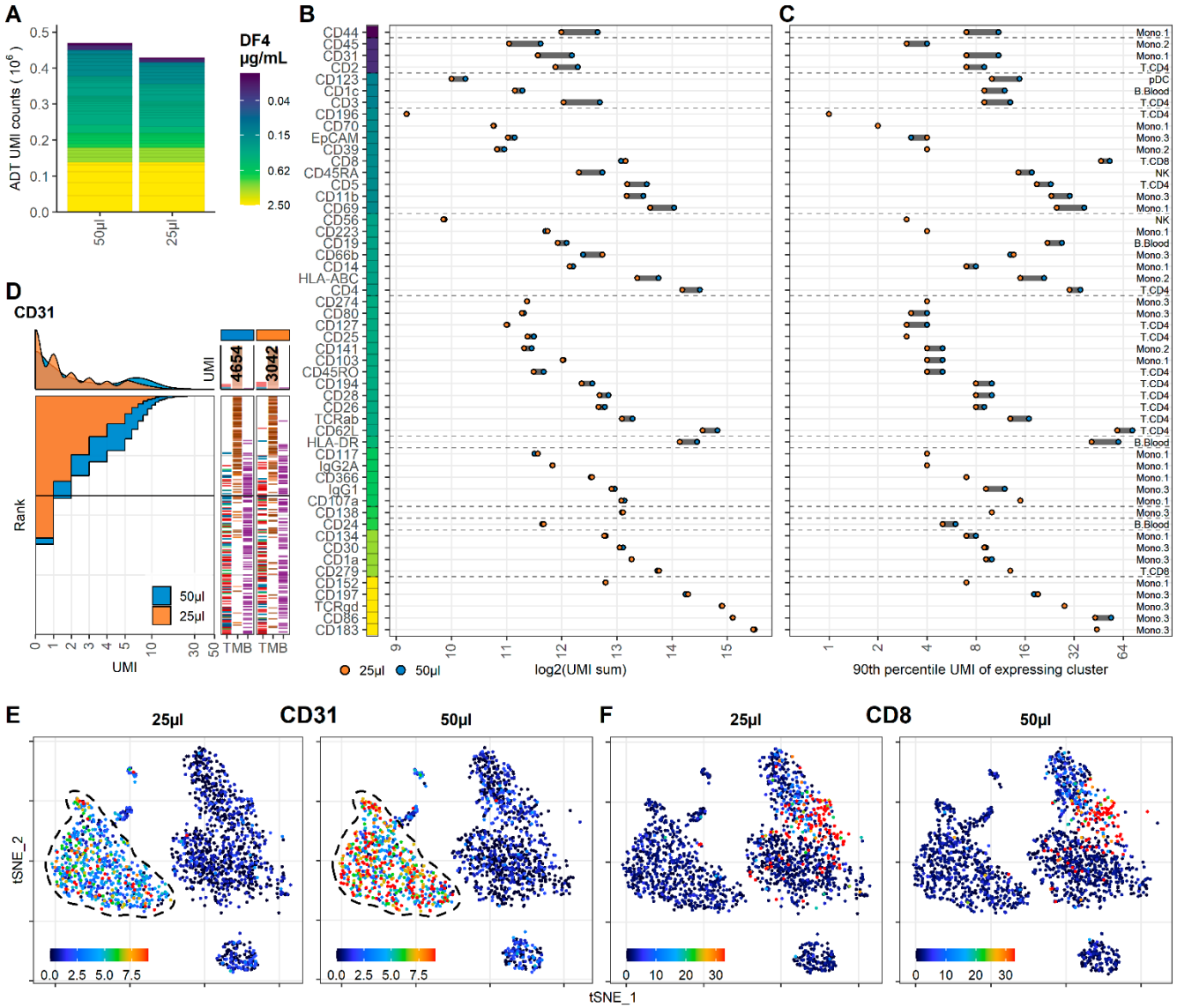


Figure 3: Reducing staining volume primarily affects highly expressed markers.
Comparison of PBMC samples stained in 50µl (same sample as DF4 in Figure 1) or 25µl volume at dilution factor 4. **A**. Summarized UMI counts within cell-containing droplets segmented by the individual antibodies colored by their concentration. **B, C**. Change in ADT signal for each antibody by reducing staining volume from 50 to 25 µl. Individual antibodies are colored by their concentration. Quantified by (**B**) sum of UMIs within cell-containing droplets assigned to each antibody and (**C**) 90th percentile UMI count within the cell type with most abundant expression (the assayed cell type is annotated to the right). **D**. Titration plot (marker UMI count vs. Normalized cell rank) for CD31 signal response when reducing staining volume from 50 µl to 25 µl. Histogram depicts distribution of UMIs at each condition. Barcode to the right depict cell type occurrence at the corresponding rank to visualize cell specificity of the antibody. Numbers on top of the small bar plot denote total UMI count assigned to CD31 within cell-containing droplets from each condition. **E**, **F**. Non-normalized UMI counts visualized on tSNE of an affected (CD31; **E**) or an unaffected (CD8; **F**) marker by the reduction in cell density Dashed line indicate region where expression levels vary between volumes. Titration plots for all markers can be found in Supplementary Figure S4.

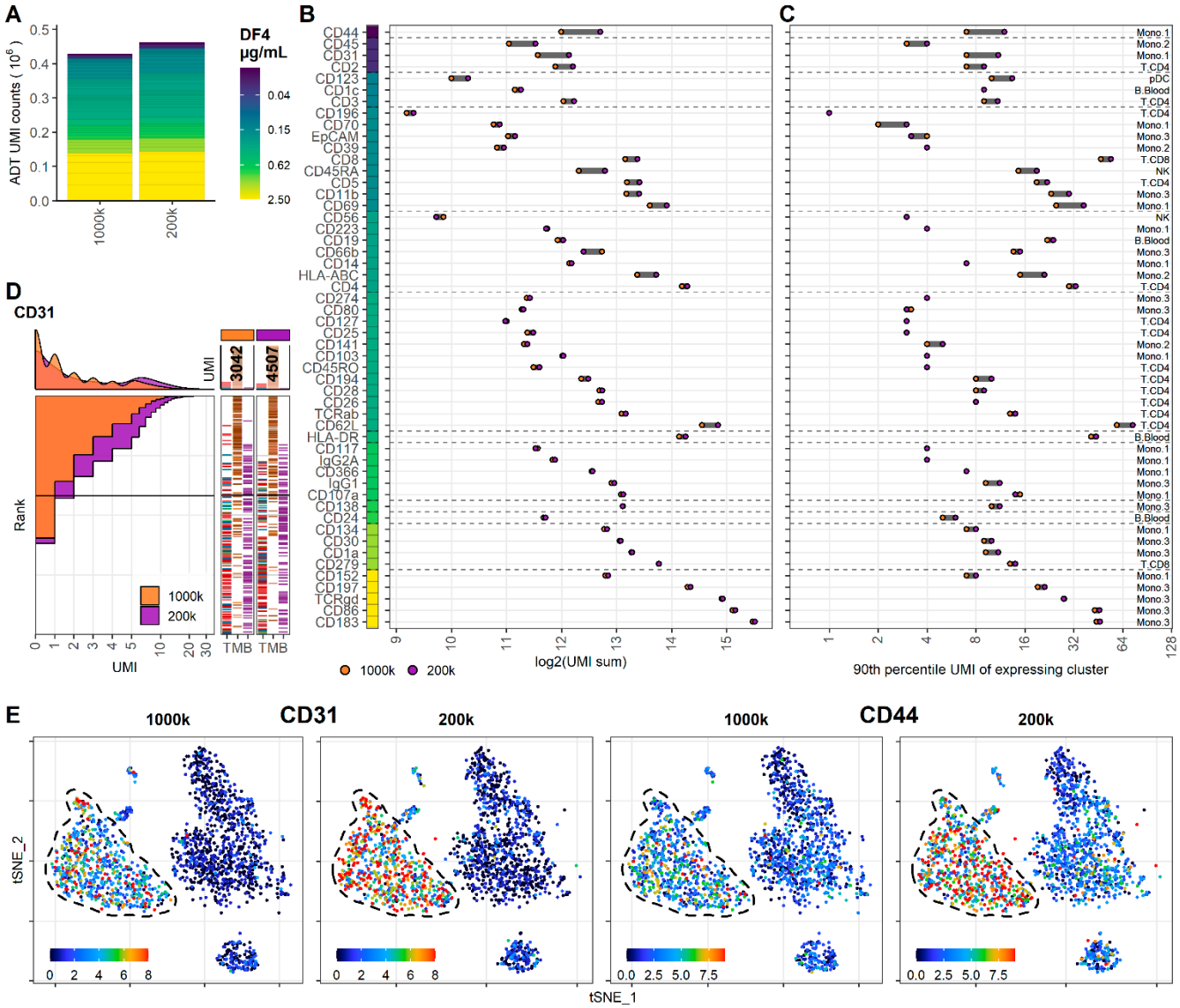


Figure 4: Reducing cell number during staining increases signal for antibodies at low concentration.

Comparison of PBMC samples stained in 25µl antibody staining solution at dilution factor 4 (DF4) at two cell densities: $1x10^6$ (1000k; same sample as 25µl in Figure 3) or $0.2x10^6$ (200k) cells. **A**. Summarized UMI counts within cell-containing droplets segmented by the individual antibodies colored by their concentration. **B, C**. Change in ADT signal for each antibody by reducing cell numbers at staining from $1x10^6$ to $0.2x10^6$ cells. Individual antibodies are colored by their concentration. Quantified by (**B**) sum of UMIs within cell-containing droplets assigned to each antibody and (**C**) 90th percentile UMI count within cell type with most abundant expression (the assayed cell type is annotated to the right). **D**. Titration plot (Marker UMI count vs. Normalized cell rank) for CD31 signal response when reducing cell numbers at staining from $1x10^6$ to $0.2x10^6$ cells. Histogram depicts distribution of UMIs at each condition. Barcode to the right depict cell type occurrence at the corresponding rank to visualize cell specificity of the antibody. Numbers on top of the small bar plot denote total UMI count assigned to CD31 within cell-containing droplets from each condition. **E**. Non-normalized UMI counts visualized on tSNE plot of CD31 and CD44 which affected by the reduction in staining volume, mitigated by a concomitant reduction in cell density. Dashed line indicate region where expression levels vary between cell densities. Titration plots for all markers can be found in Supplementary Figure S5.

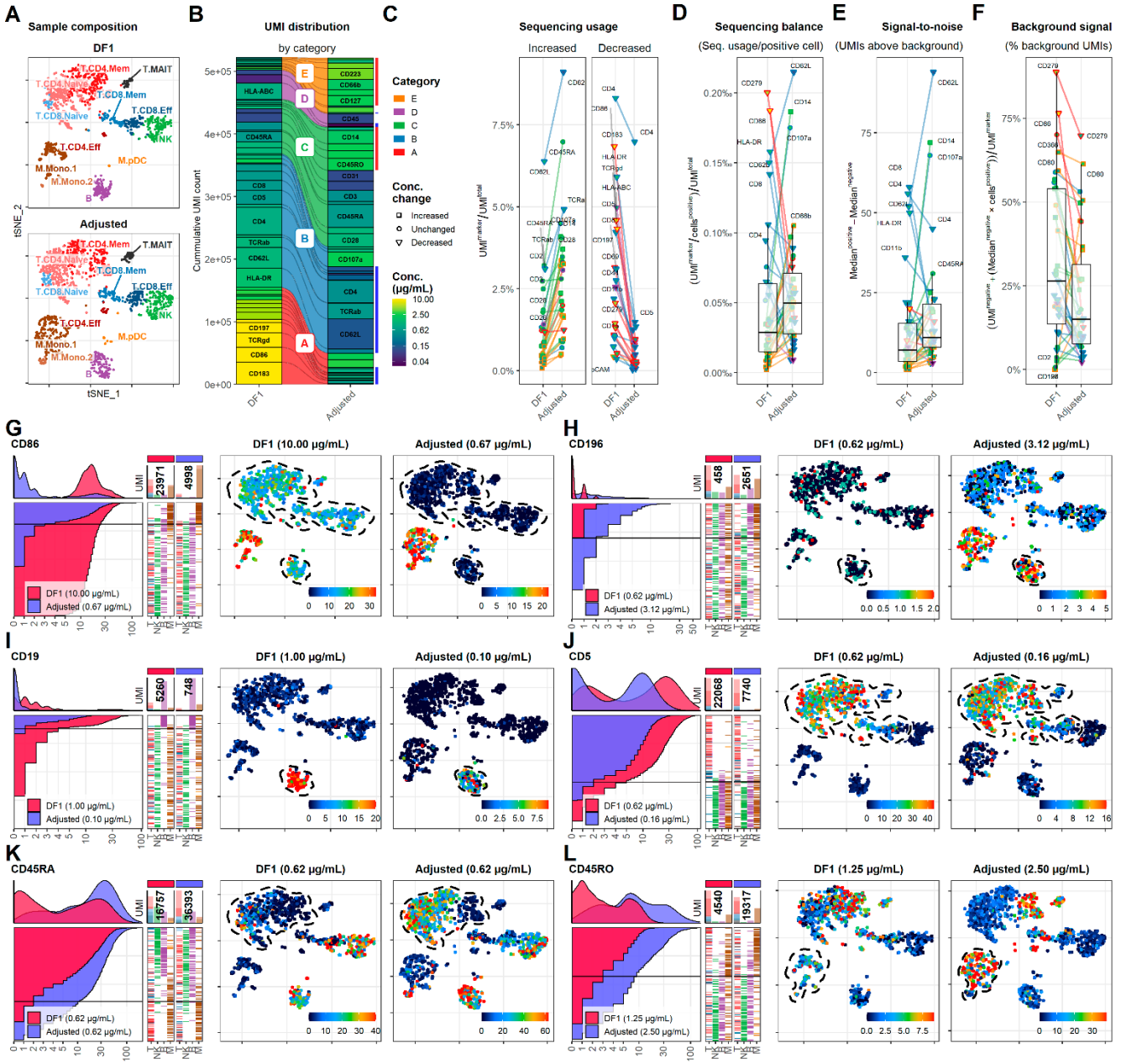


Figure 5: Adjusting antibody concentrations increases signal, lowers background and reduces costs and sequencing requirements.

A. Single cells from the DF1 and adjusted sample were integrated and selected to yield similar number of cells within each annotated cell type, visualized by tSNE. **B**. ADT reads from DF1 and adjusted samples were subsampled to yield similar number of UMIs within the selected cells. Size of each segment show the distribution of UMIs among the antibodies in the panel divided into categories that determined how they were adjusted (Table 1). **C-F**. Response of adjustment of individual antibodies assayed by (**C**) their overall sequencing usage (fraction of UMIs assigned to each marker), (**D**) their balancing (percent of UMIs used per positive cell), (**E**) signal-to-noise (difference in median UMI count within positive and negative cells) and (**F**) background signal (percentage of UMIs used for background signal). Shapes of marker denote whether the antibody concentration was changed between the DF1 and adjusted sample. Color of "shapes" denote antibody concentration. Color of connecting lines denote antibody category. Center line in box plot denote the median. **G-L**. Titration plot (left) and tSNE plots showing raw UMI counts (right) for antibodies in different categories. Dashed red lines indicate regions of interest highlighting the differences (or lack thereof) between the DF1 and adjusted samples. Titration plots for all markers by category can be found in Supplementary Figure S7.

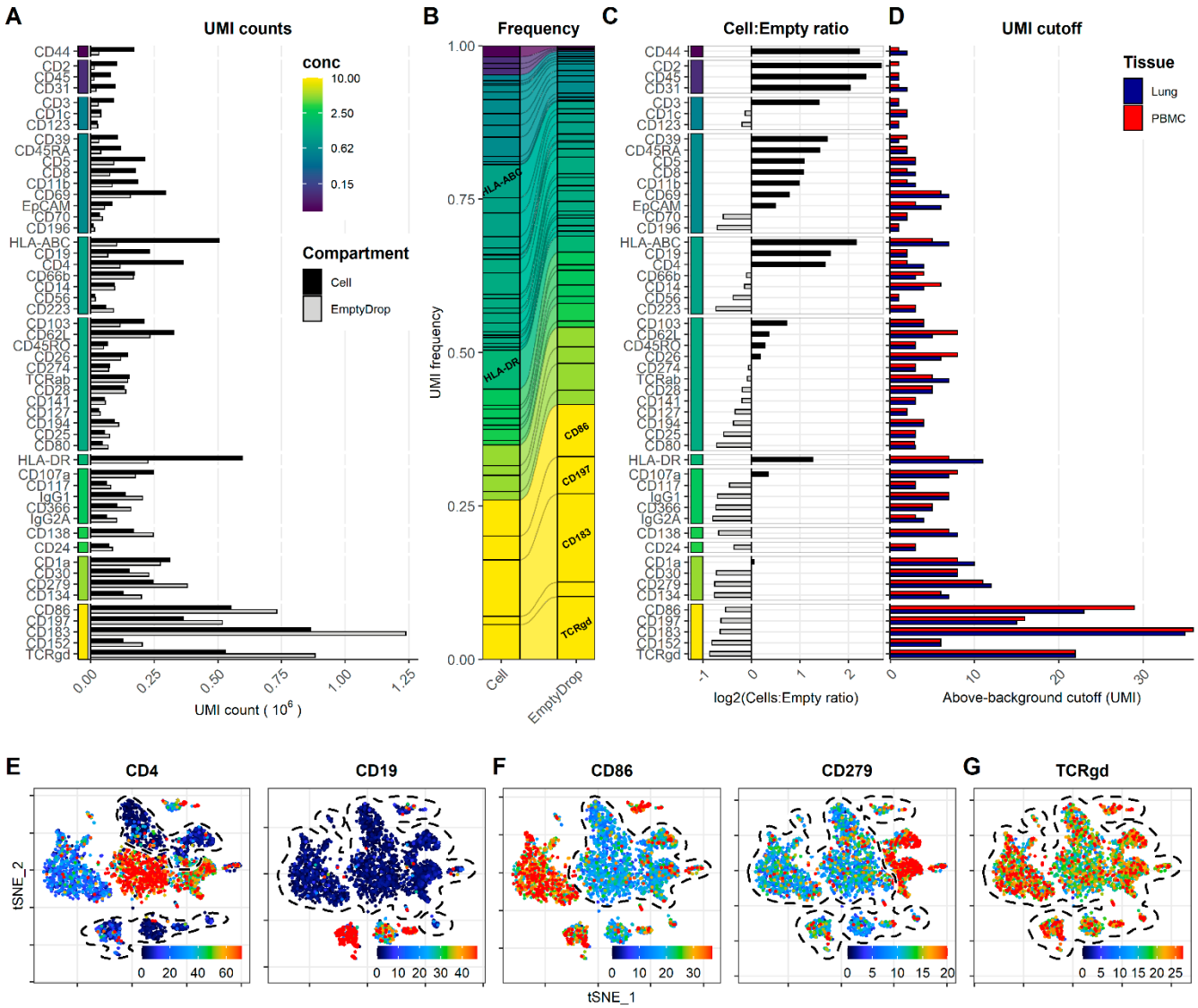


Figure 6: Background signal from oligo-conjugated antibodies are dependent on concentration and presence of epitopes.

Signal from free-floating antibodies in the cell suspension is a major source of background in droplet-based scRNA-seq and can be assayed by their signal within non-cell-containing (empty) droplets. **A-B**. Comparison of signal from each antibody within cell-containing and empty droplets (identified in Supplementary Figure S5) by (**A**) their total UMI counts or, (**B**) their relative frequency within each compartment. Color bar denotes antibody concentration at dilution factor 1 (DF1). **C**. Ratio of UMI frequencies of each marker between cell-containing and empty droplets. Markers with black bars have greater frequency in cell-containing droplets whereas grey bars have greater frequency in empty droplets. **D**. UMI Thresholds for detection above-background for each marker within PBMC and Lung tumor samples (based on gating in Supplementary Figure S1). **E-G**. Examples of tSNE plots showing non-normalized (raw) UMI counts from cells stained at dilution factor 1 (DF1) for (**E**) markers with low background, (**F**) markers with high background that still exhibit cell type-specific signal (CD86 and CD279) and (**G**) marker where positive signal is absent or obscured by the background. Regions of background signal are encircled by dashed lines. To make the color scale in the tSNE plots less sensitive to extreme values, we set the upper threshold to the 90% percentile. tSNE plots for all markers can be found in Supplementary Figure S10.

Supplementary Tables

Supplementary Table 1: Oligo-conjugated antibody panel for 5'-CITE-seq

Marker	Alias	Clone	Isotype (Mouse)	Vendor	TotalSeq C Tag	Cat#	DF1 conc. (µg/mL)	DF4 conc. (µg/mL)	Adj. conc. (µg/mL)	Adj. conc. (DF1 FC)
CD44		BJ18	IgG1	BioLegend	0125	338827	0.050	0.013	0.025	2.00
CD2		TS1/8	IgG1	BioLegend	0367	309231	0.100	0.025	0.100	1.00
CD31	PECAM-1	WM59	IgG1	BioLegend	0124	303139	0.100	0.025	0.100	1.00
CD45		HI30	IgG1	BioLegend	0391	304068	0.100	0.025	0.100	1.00
CD123		6H6	IgG1	BioLegend	0064	306045	0.500	0.125	0.500	1.00
CD3		UCHT1	IgG1	BioLegend	0034	300479	0.500	0.125	0.500	1.00
CD1c		L161	IgG1	BioLegend	0160	331547	0.500	0.125	1.000	0.50
EpCAM		9C4	IgG2b	BioLegend	0123	324247	0.625	0.156	0.063	10.00
CD8		RPA-T8	IgG1	BioLegend	0080	301071	0.625	0.156	0.078	8.00
CD11b		ICRF44	IgG1	BioLegend	0161	301359	0.625	0.156	0.125	5.00
CD5		UCHT2	IgG1	BioLegend	0138	300637	0.625	0.156	0.156	4.00
CD69		FN50	IgG1	BioLegend	0146	310951	0.625	0.156	0.156	4.00
CD39		A1	IgG1	BioLegend	0176	328237	0.625	0.156	0.625	1.00
CD45RA		HI100	IgG2b	BioLegend	0063	304163	0.625	0.156	0.625	1.00
CD70		113-16	IgG1	BioLegend	0027	355119	0.625	0.156	1.250	0.50
CD196	CCR6	G034E3	IgG2b	BioLegend	0143	353440	0.625	0.156	3.125	0.20
CD19		HIB19	IgG1	BioLegend	0050	302265	1.000	0.250	0.100	10.00
HLA-ABC		W6/32	IgG2a	BioLegend	0058	311449	1.000	0.250	0.125	8.00
CD4		RPA-T4	IgG1	BioLegend	0072	300567	1.000	0.250	0.500	2.00
CD14		M5E2	IgG2a	BioLegend	0081	301859	1.000	0.250	2.000	0.50
CD66b		6/40c	IgG1	BioLegend	0166	392909	1.000	0.250	2.000	0.50
CD223		11C3C65	IgG1	BioLegend	0152	369335	1.000	0.250	4.000	0.25
CD56		5.1H11	IgG1	BioLegend	0047	362559	1.000	0.250	4.000	0.25
CD103		BerACT8	IgG1	BioLegend	0145	350233	1.250	0.313	0.125	10.00
CD62L		DREG-56	IgG1	BioLegend	0147	304851	1.250	0.313	0.125	10.00
TCRab		IP26	IgG1	BioLegend	0224	306743	1.250	0.313	0.625	2.00
CD194	CCR4	L291H4	IgG1	BioLegend	0071	359425	1.250	0.313	1.250	1.00
CD26	DPP4	BA5b	IgG2a	BioLegend	0396	302722	1.250	0.313	1.250	1.00
CD274	B7-H1 (PD-L1)	29E.2A3	IgG2a	BioLegend	0007	329751	1.250	0.313	1.250	1.00
CD28		CD28.2	IgG1	BioLegend	0386	302963	1.250	0.313	1.250	1.00
CD127	IL7Ralpha	A019D5	IgG1	BioLegend	0390	351356	1.250	0.313	2.500	0.50
CD141	Thrombomodulin	M80	IgG1	BioLegend	0163	344125	1.250	0.313	2.500	0.50
CD25		BC96	IgG1	BioLegend	0085	302649	1.250	0.313	2.500	0.50
CD45RO		UCHL1	IgG2a	BioLegend	0087	304259	1.250	0.313	2.500	0.50
CD80	B7-1	2D10	IgG1	BioLegend	0005	305243	1.250	0.313	2.500	0.50
HLA-DR		L243	IgG2a	BioLegend	0159	307663	1.650	0.413	0.330	5.00
CD117	C-kit	104D2	IgG1	BioLegend	0061	313243	2.500	0.625	0.625	4.00
IsoIgG1		MOPC-21		BioLegend	0090	400187	2.500	0.625	1.000	2.50
IsoIgG2A		MOPC-173		BioLegend	0091	400293	2.500	0.625	1.000	2.50
CD107a	LAMP1	H4A3	IgG1	BioLegend	0155	328649	2.500	0.625	2.500	1.00
CD366		F38-2E2	IgG1	BioLegend	0169	345049	2.500	0.625	2.500	1.00
CD138		DL-101	IgG1	BioLegend	0831	352327	3.000	0.750	3.000	1.00
CD24		ML5	IgG2a	BioLegend	0180	311143	3.300	0.825	3.300	1.00
CD134	OX40	Ber-ACT35	IgG1	BioLegend	0158	350035	5.000	1.250	1.000	5.00
CD1a		HI149	IgG1	BioLegend	0402	300135	5.000	1.250	1.000	5.00
CD279	PD-1	EH12.2H7	IgG1	BioLegend	0088	329963	5.000	1.250	1.000	5.00
CD30		BY88	IgG1	BioLegend	0028	333919	5.000	1.250	1.000	5.00
CD183	CXCR3	G025H7	IgG1	BioLegend	0140	353747	10.000	2.500	0.500	20.00
TCRgd		B1	IgG1	BioLegend	0139	331231	10.000	2.500	0.500	20.00
CD86	B7-2	IT2.2	IgG2b	BioLegend	0006	305447	10.000	2.500	0.667	15.00
CD152	CTLA4	BNI3	IgG2a	BioLegend	0151	369621	10.000	2.500	1.000	10.00
CD197	CCR7	G043H7	IgG2a	BioLegend	0148	353251	10.000	2.500	1.000	10.00

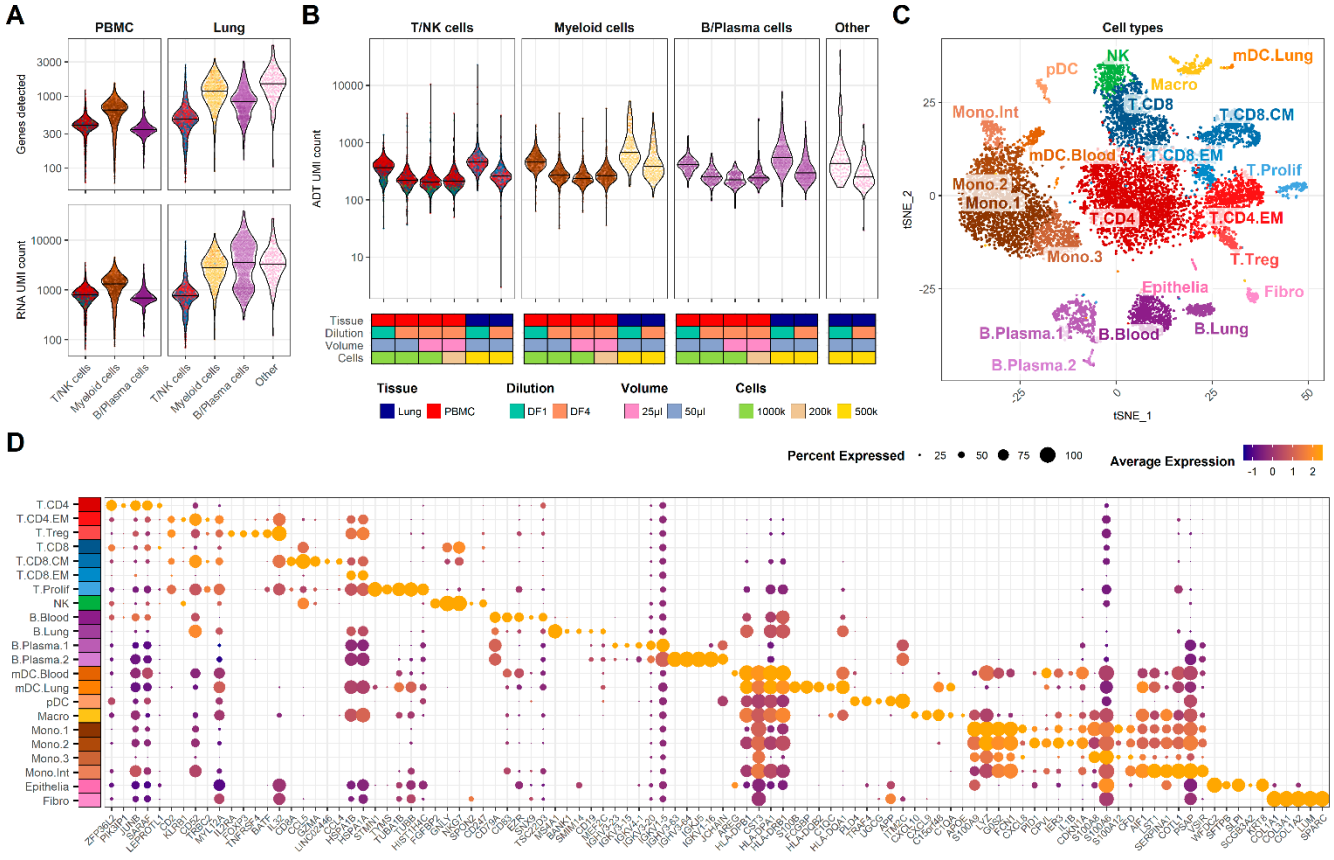
Supplementary Table 2: Antibody usage and costs

Panel data	Vendor rec.	DF1	Adjusted	Units
Average Ab conc	10.000	2.306	1.184	*µg/mL*
Staining volume	100	50	25	*µl*
Total Ab used*	52.000	5.996	1.539	*µg*
Cost of Abs per sample**	1690	194.857	50.024	*USD*
Fold Reductions in cost compared to				
Vendor rec.	1.0	8.7	33.8	*Fold*
DF1	-	1.0	3.9	*Fold*
Vendor rec. (same volume)	1.0	4.3	8.4	*Fold*
DF1 (same volume)	-	1.0	1.9	*Fold*

*The table is based on the 52 marker TotalSeq-C antibody panel shown in Figure 1-Figure Supplement 1
**Price calculations based on USD 325 per 10µg of TotalSeq-C antibody (list price as of 2021-02-23 from BioLegend).

Supplementary Figures

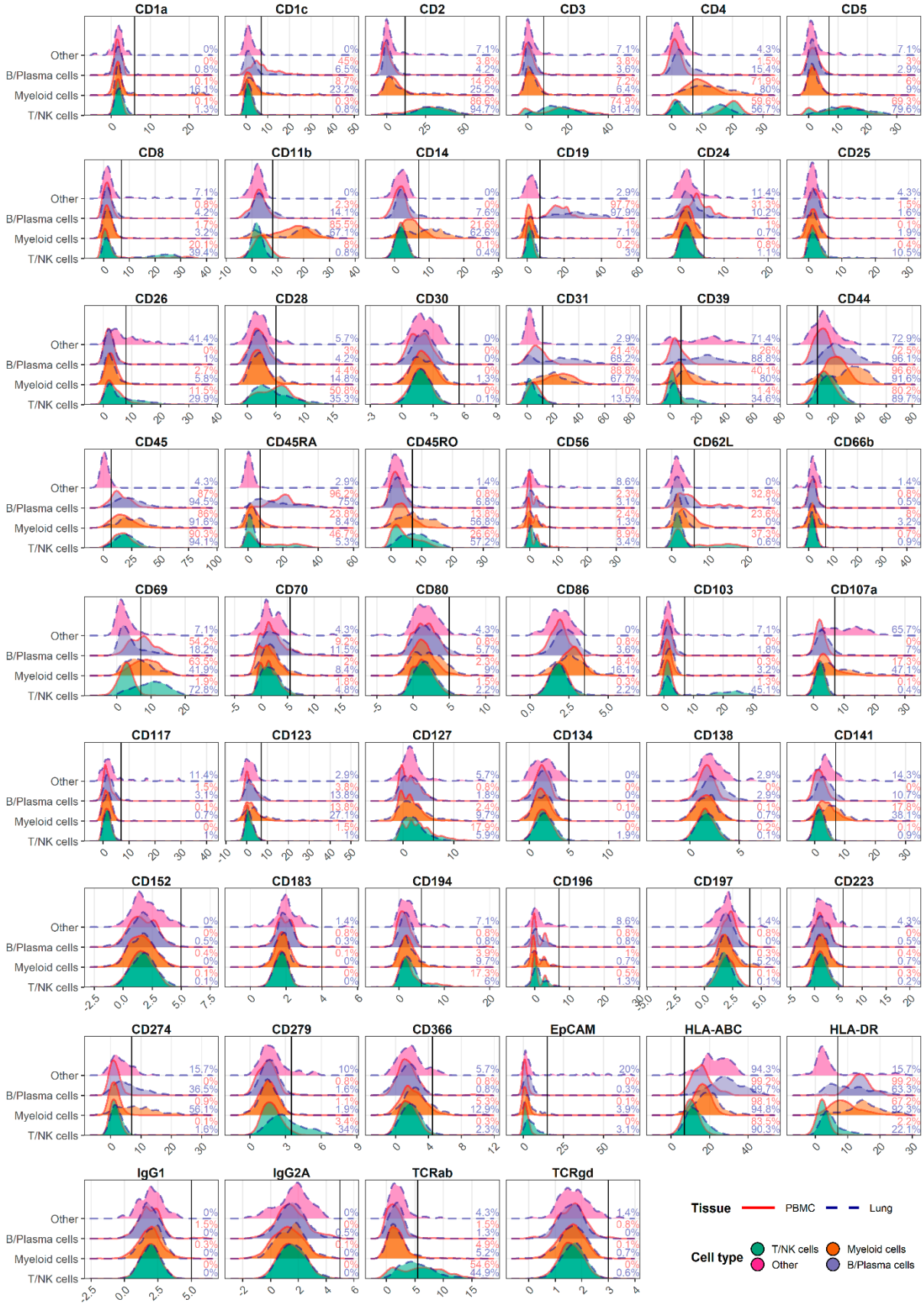
Supplementary Figure S1



Supplementary Figure S1: Quality control metrics and cell type annotation

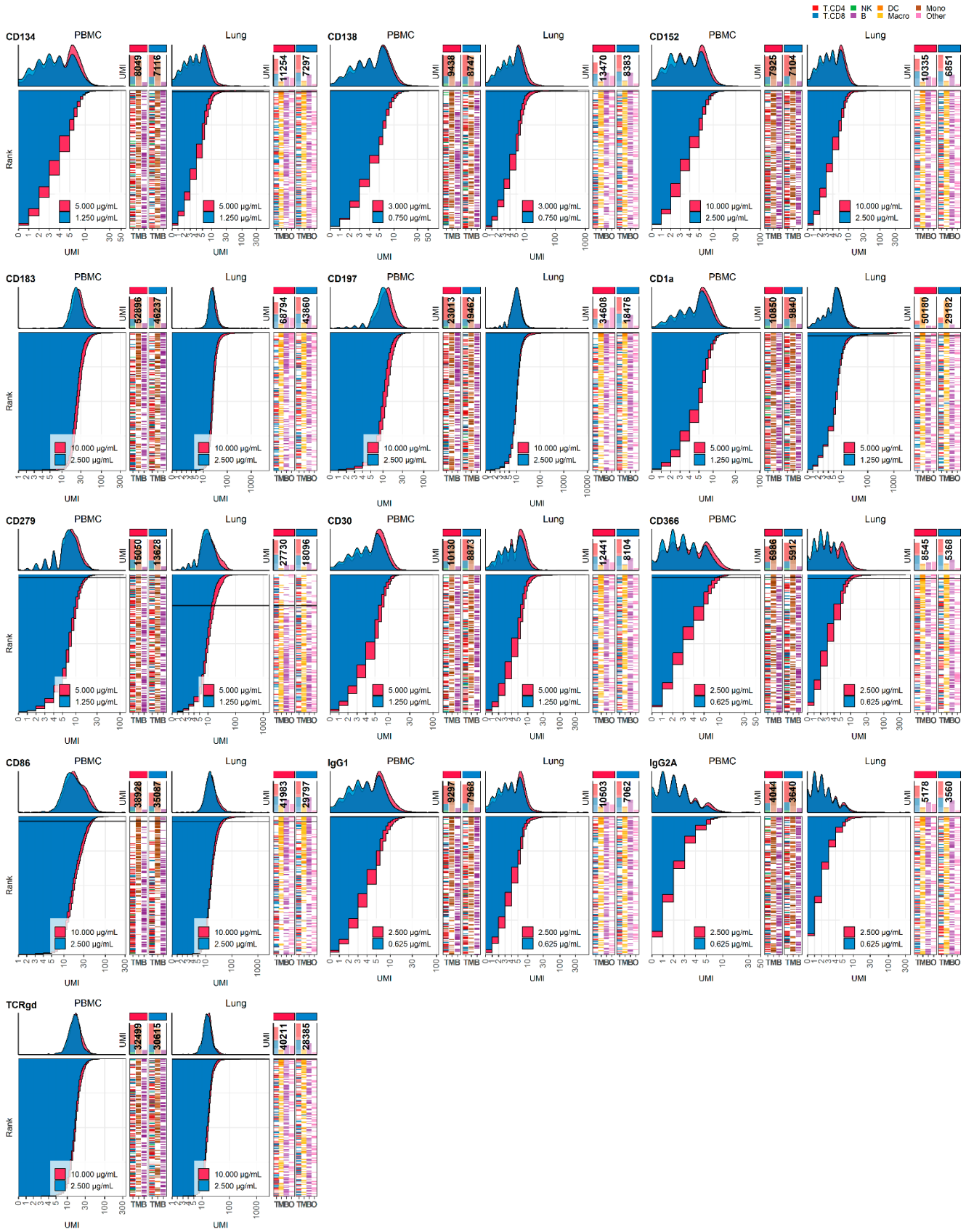
A. Number of genes detected (top) and UMI count (bottom) for each cell across major cell lineages within the mRNA modality. **B**. UMI count within ADT modality across different experimental conditions and cell lineages. **C**. Detailed cell type annotation visualized by tSNE. **D**. Top 5 marker genes for each cell type (expressed by > 30% of cells and having being at least 20% more abundant compared to other cell types). Only showing dots for genes expressed in > 10% of the cluster.

Supplementary Figure S2

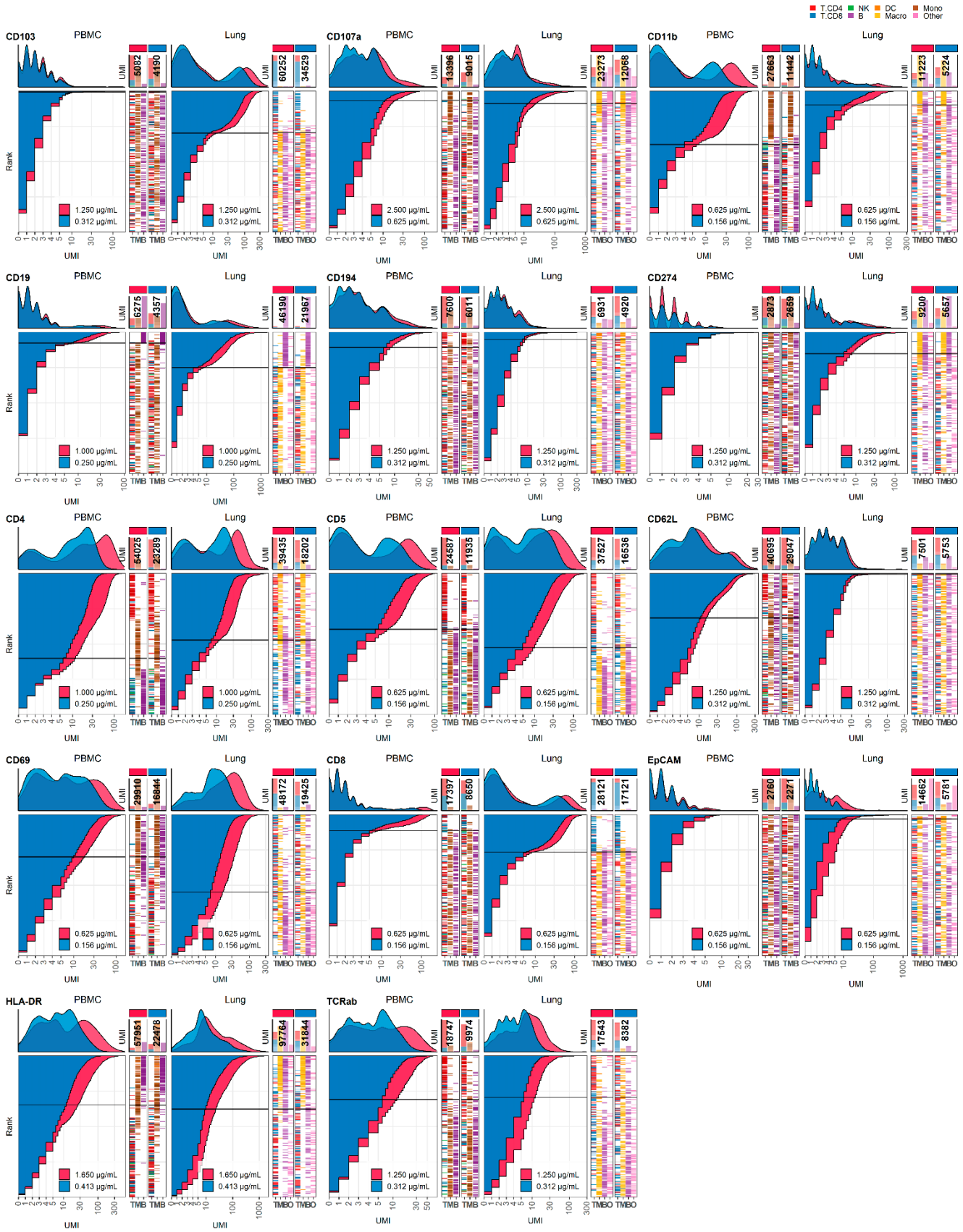


Supplementary Figure S2: Gating positive cells based on ADT signal at dilution factor 1 (DF1). Histograms for normalized ADT expression of each marker within major cell types in PBMC (red) and Lung (blue). Gating threshold indicated by vertical line and numbers denote percent positive within each cell type of each tissue.

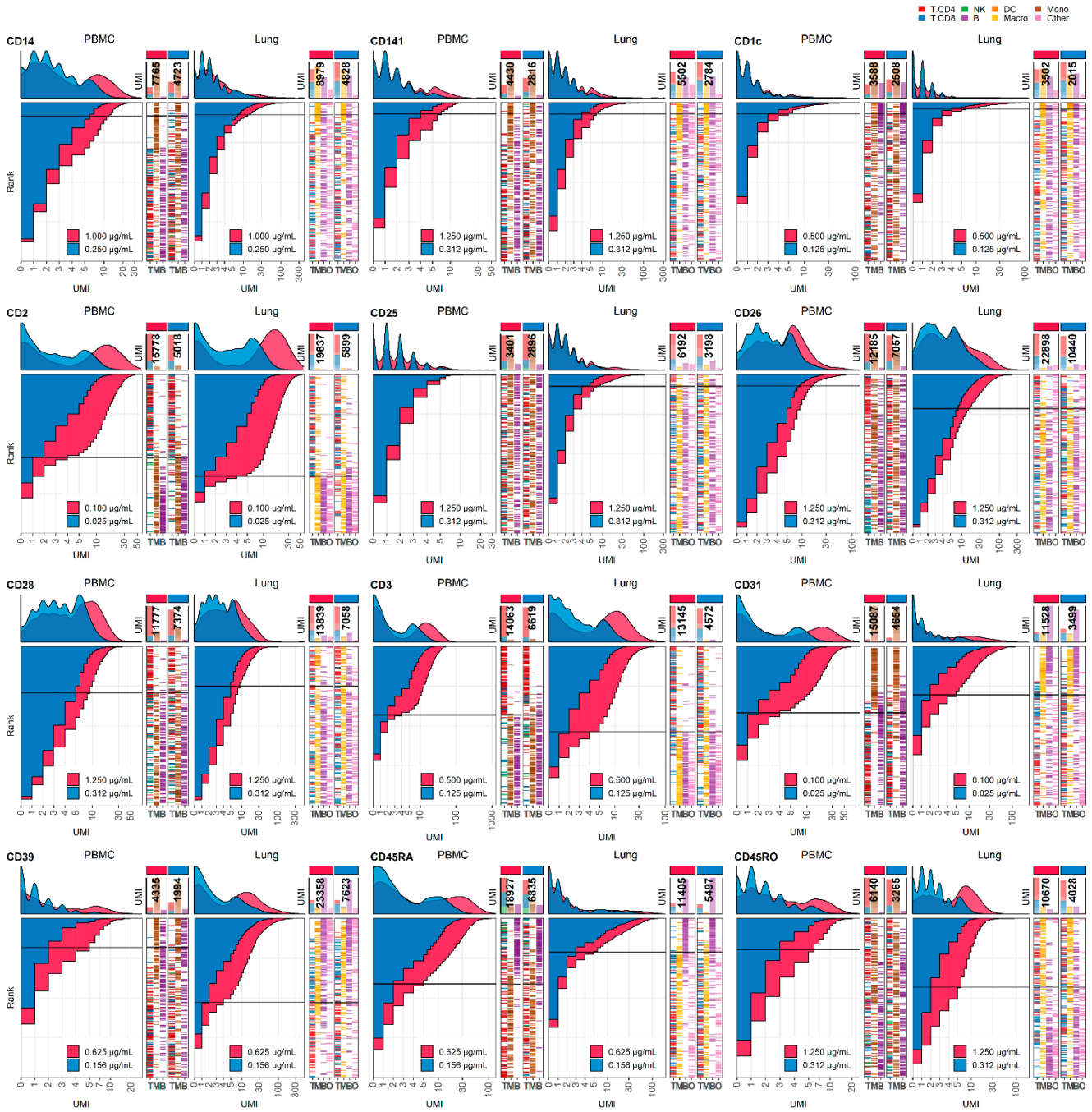
Supplementary Figure S3A



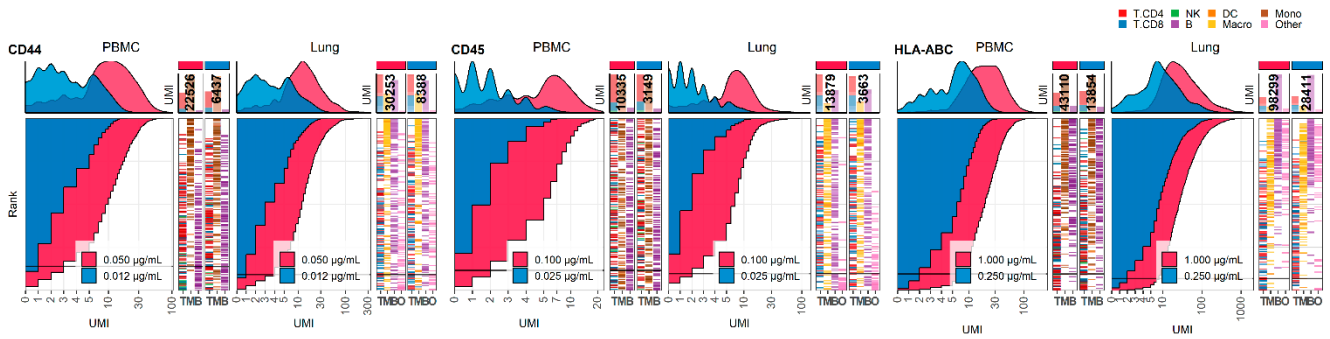
Supplementary Figure S3B



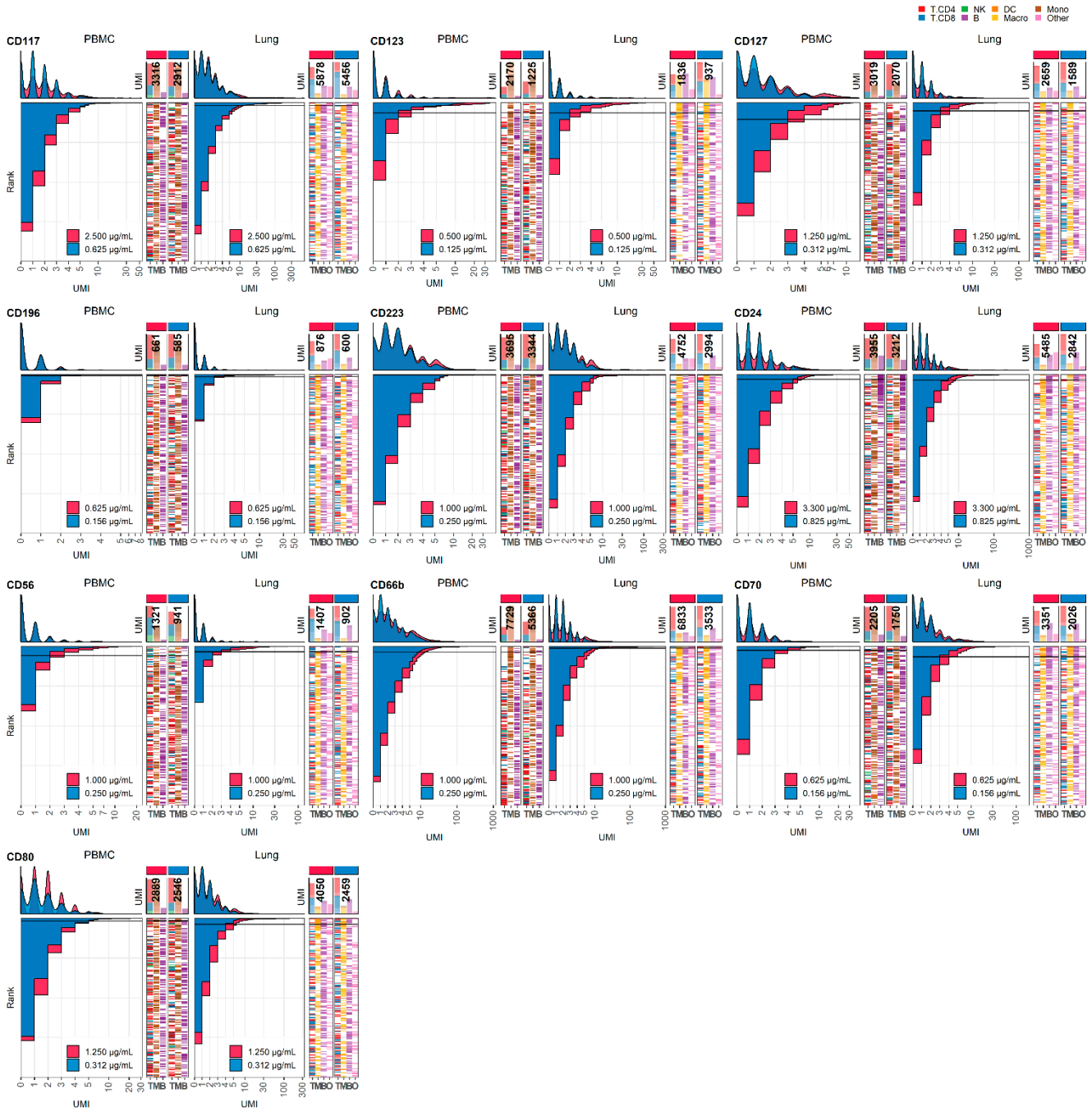
Supplementary Figure S3C



Supplementary Figure S3D

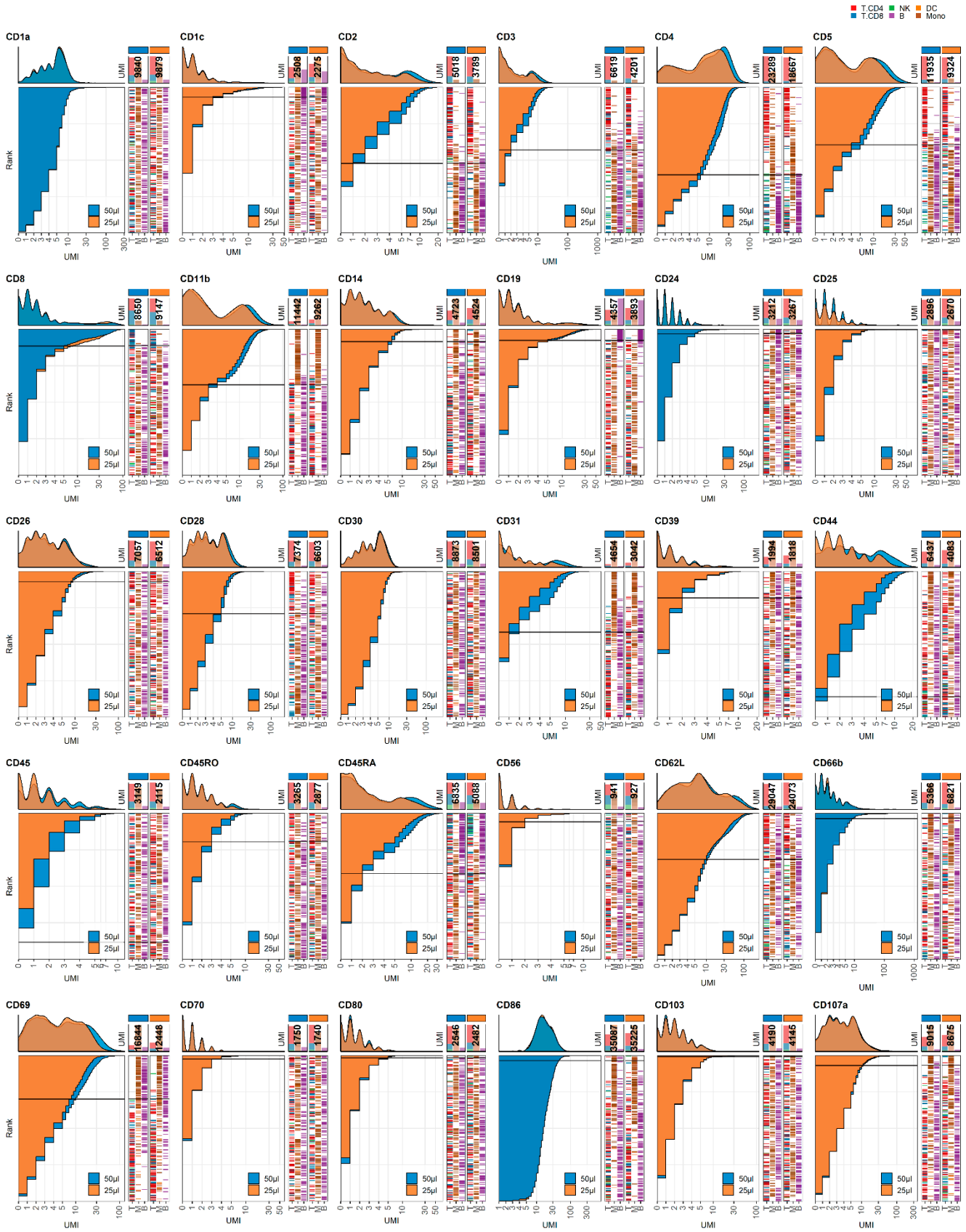


Supplementary Figure S3E

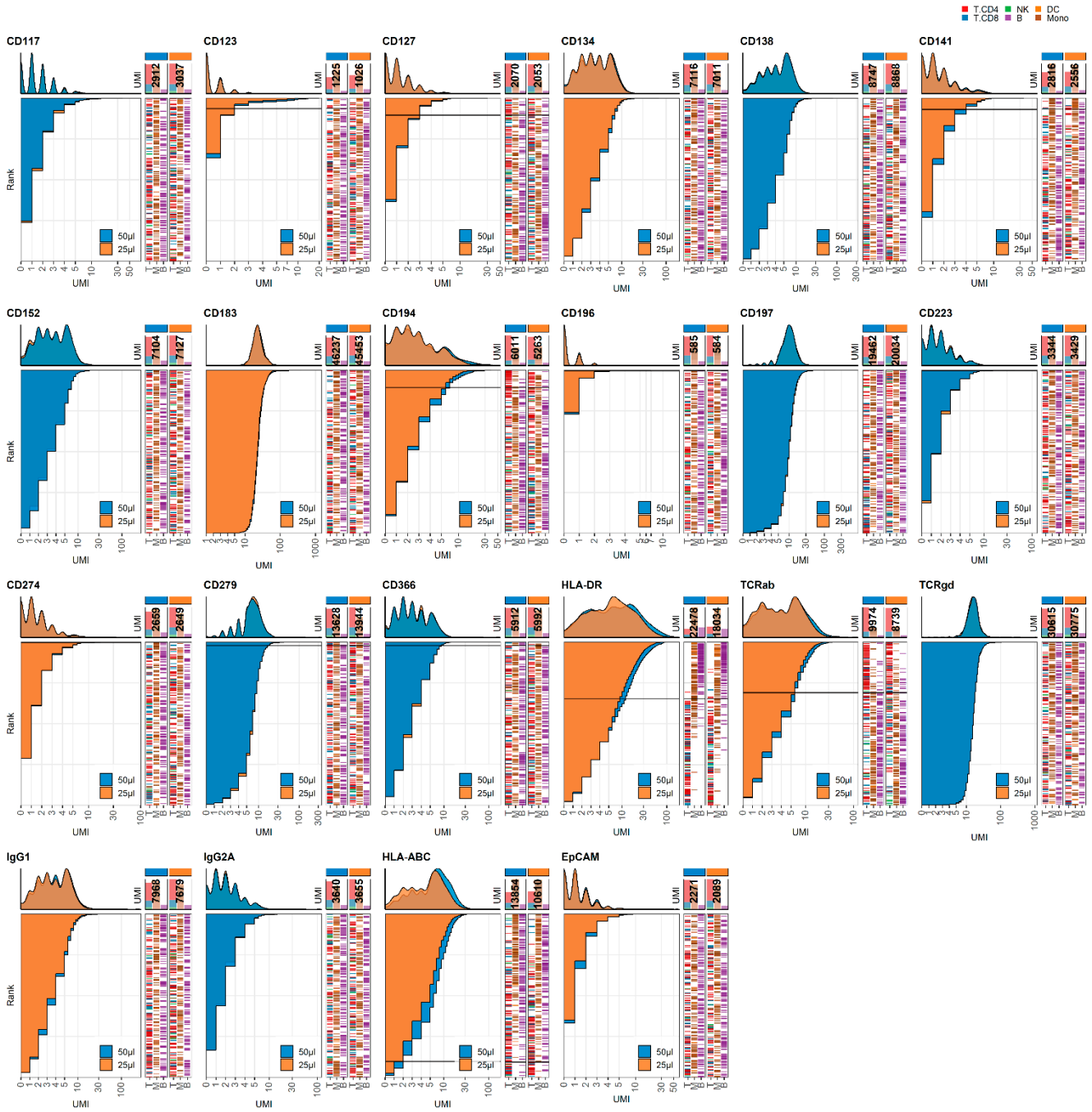


Supplementary Figure S3: Response of individual antibodies to four-fold reduction in concentration in PBMCs and Lung tumor immune cells.
Titration plots (marker UMI count vs. cell rank) showing response to reduction in antibody concentration from dilution factor 1 (DF1) to DF4 within PBMC (left) and Lung (right). Histogram depicts distribution of UMIs at each condition colored by dilution factor (and annotated with concentration). Numbers within bar plot denote total UMI count within cell-containing droplets at each antibody concentration. Barcodes to the right depicts cell type by color at the corresponding rank to visualize specificity of the antibody. Horizontal line depicts gating threshold for cells considered positive for the marker.

Supplementary Figure S4A

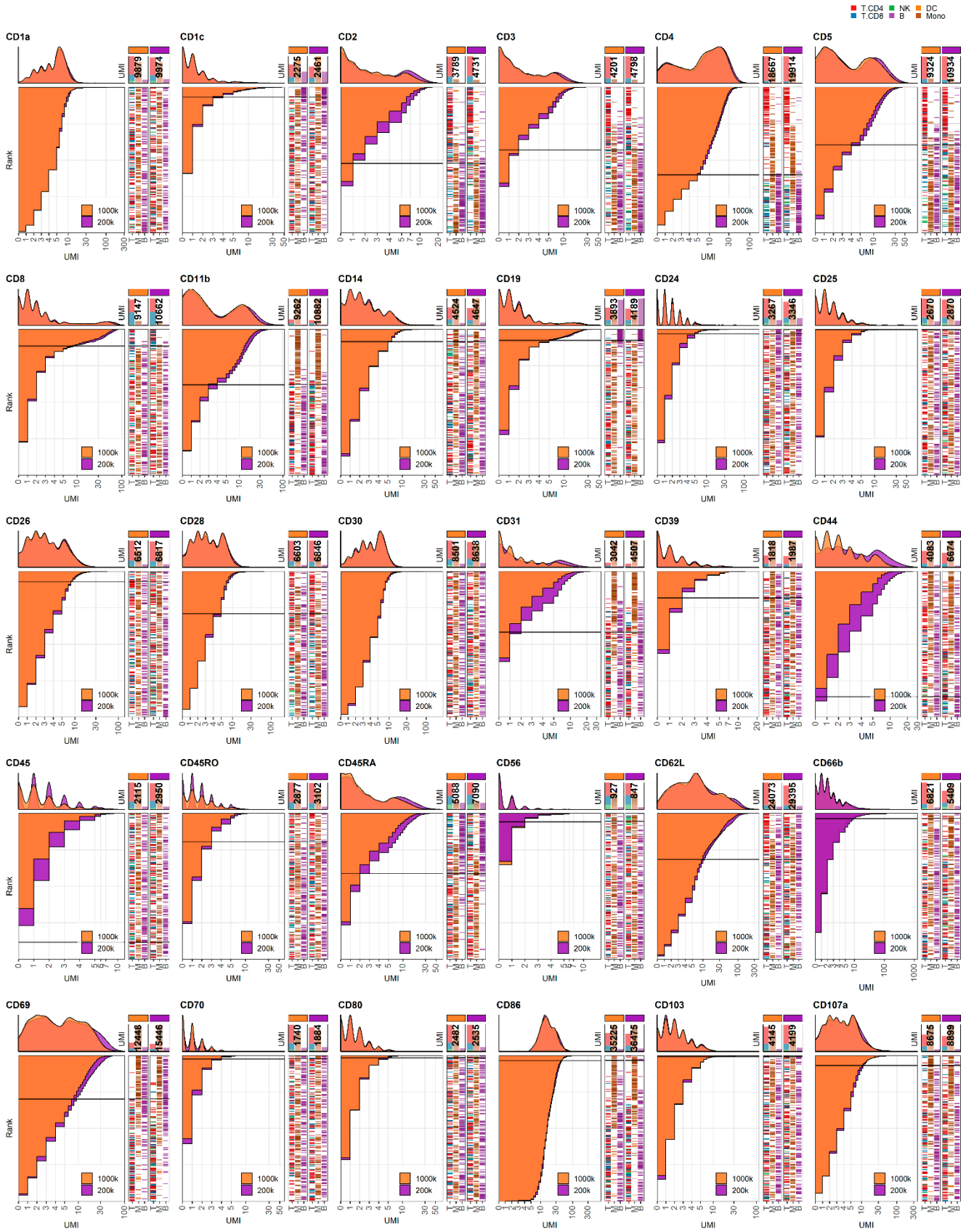


Supplementary Figure S4B

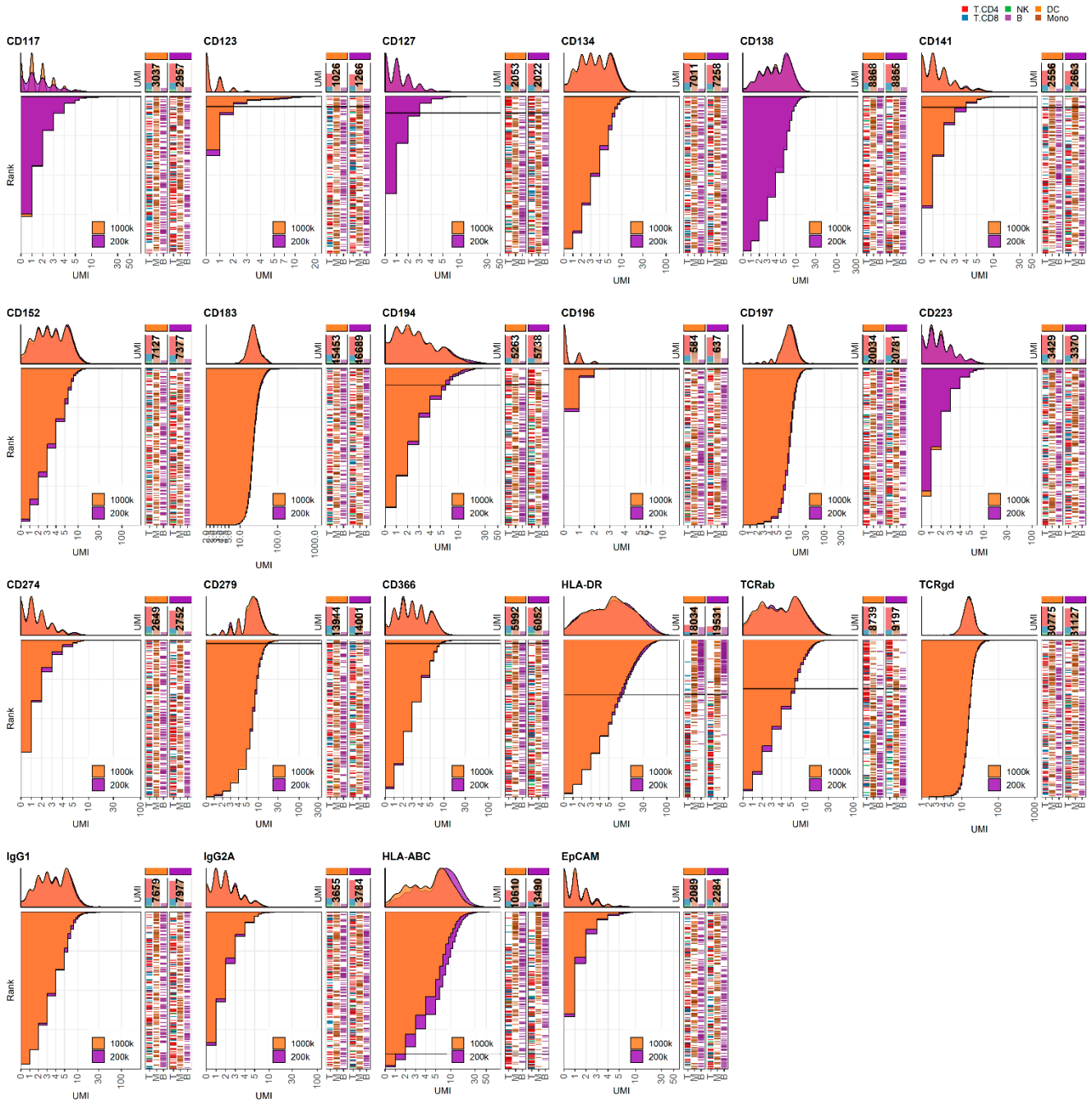


Supplementary Figure S4: Response of individual antibodies to reduction in staining volume. Titration plots (marker UMI count vs. cell rank) showing response to reducing staining volume from 50 µl to 25 µl. Histogram depicts distribution of UMIs at each condition colored by condition. Numbers within bar plot denote total UMI count within cell-containing droplets at each antibody concentration. Barcodes to the right depicts cell type by color at the corresponding rank to visualize specificity of the antibody. Horizontal line depicts gating threshold for cells considered positive for the marker.

Supplementary Figure S5A



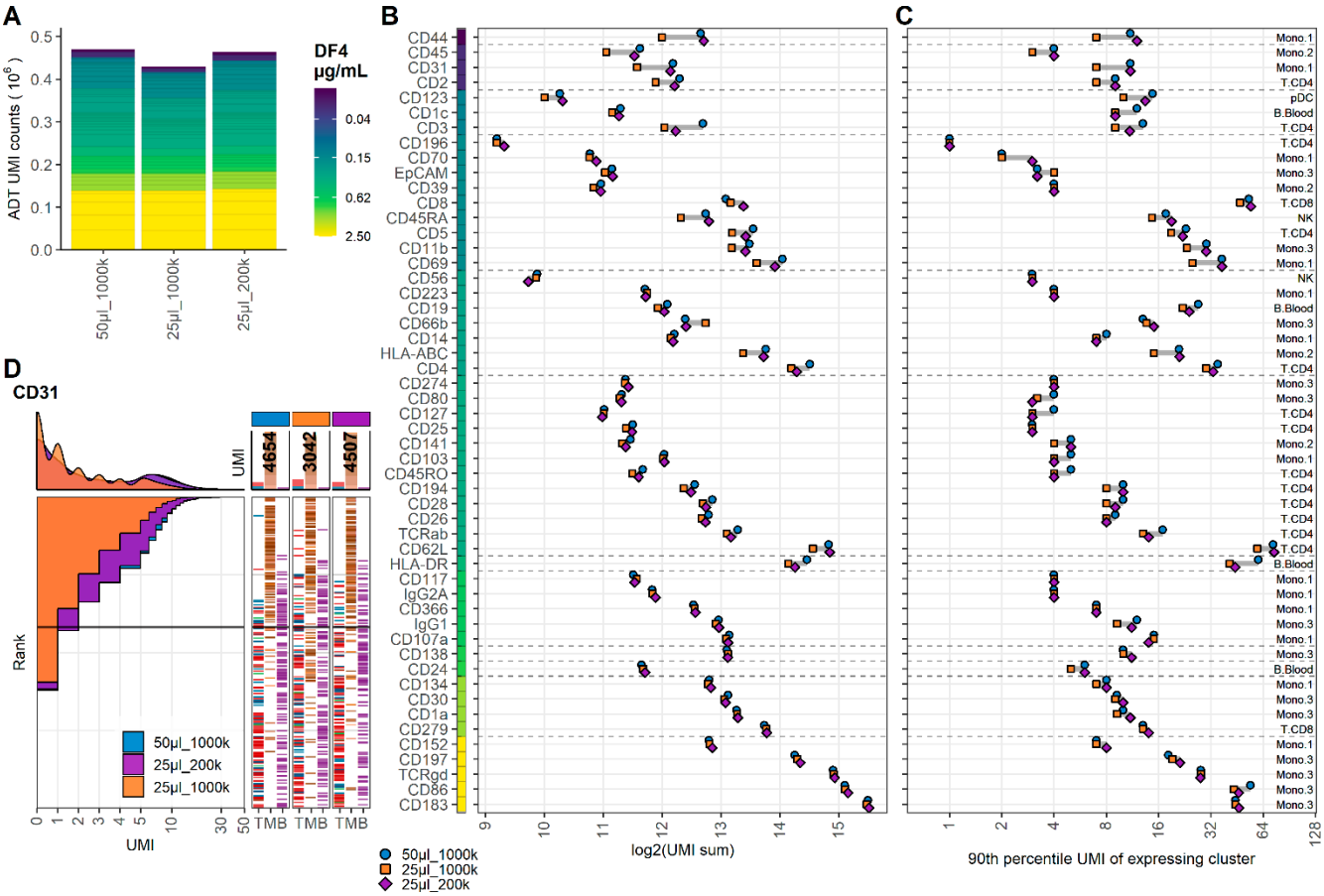
Supplementary Figure S5B



Supplementary Figure S5: Response of individual antibodies to reduction in cell numbers at staining.

Titration plots (marker UMI count vs. cell rank) showing response to reducing cell numbers at staining from $1x10^6$ to $0.2x10^6$ cells. Histogram depicts distribution of UMIs at each condition colored by condition. Numbers within bar plot denote total UMI count within cell-containing droplets at each antibody concentration. Barcodes to the right depicts cell type by color at the corresponding rank to visualize specificity of the antibody. Horizontal line depicts gating threshold for cells considered positive for the marker.

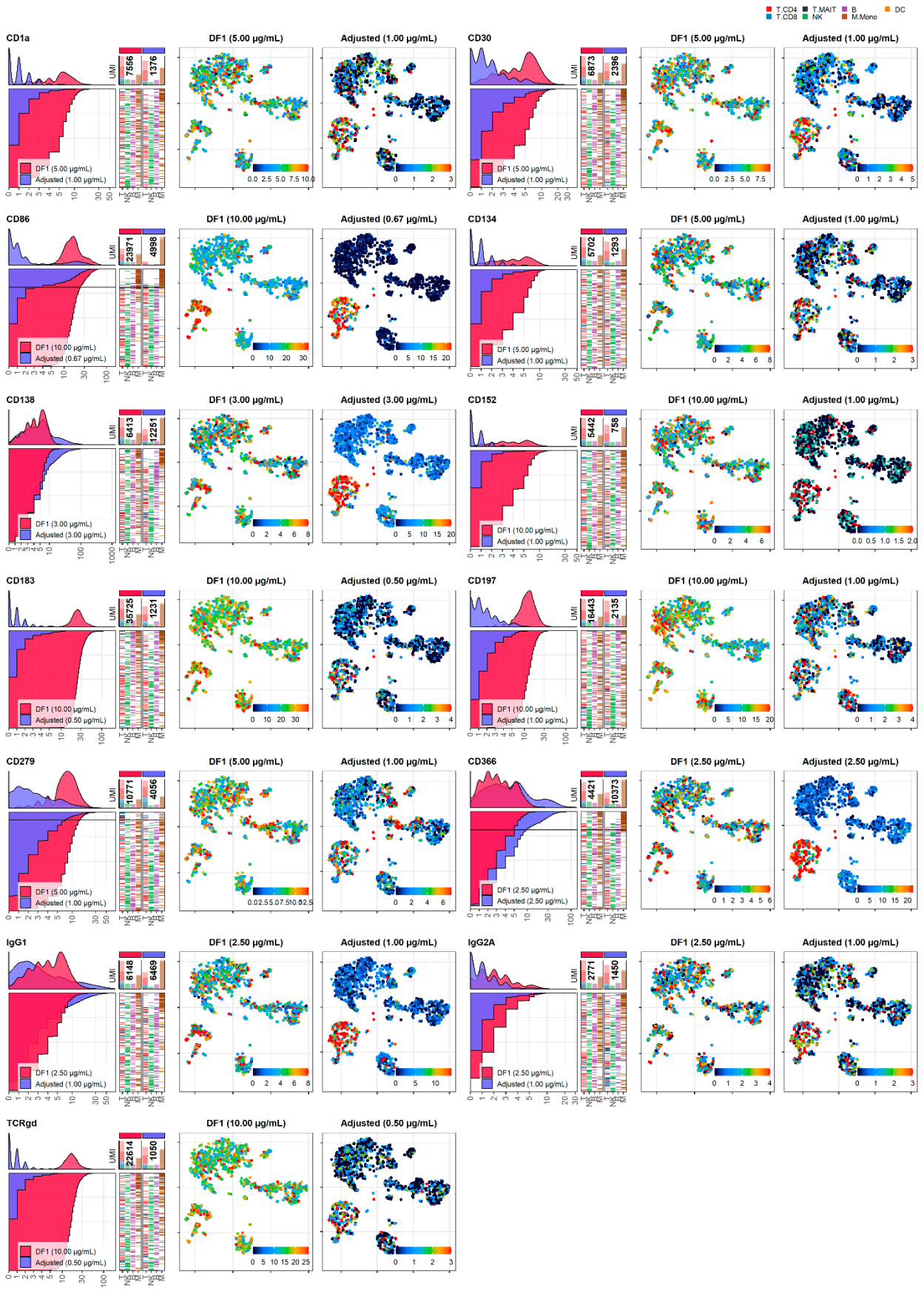
Supplementary Figure S6



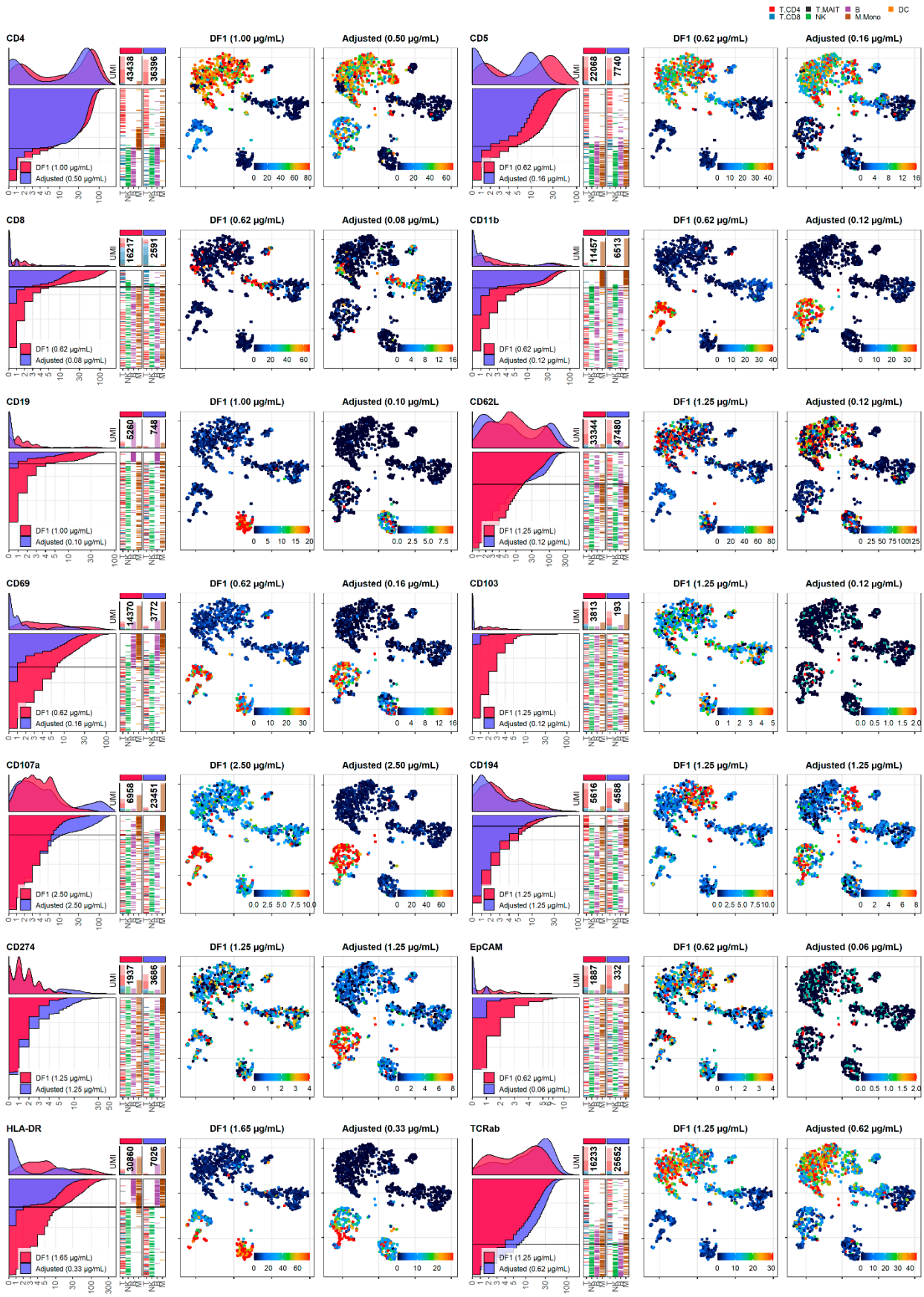
Supplementary Figure S6: 5-fold reduction in cell density mitigates but does not supersede 2-fold reduction in staining volume.

Comparison of PBMC samples stained in 50µl, 25µl and 25µl antibody staining solution at dilution factor 4 (DF4) at cell densities: $1x10^6$ (50µl_1000k), $1x10^6$ (25µl_1000k) or $0.2x10^6$ (25µl_200k) cells, respectively. **A**. Summarized UMI counts within cell-containing droplets segmented by the individual antibodies colored by their concentration. **B, C**. ADT signal for each antibody at each condition. Individual antibodies are colored by their concentration. Quantified by (**B**) sum of UMIs within cell-containing droplets assigned to each antibody and (**C**) 90th percentile UMI count within cell cluster with most abundant expression (the assayed cluster is annotated by numbers inside the). **D**. Titration plot (marker UMI count vs. cell rank) showing response to changing staining volume and/or cell density. Histogram depicts distribution of UMIs at each condition colored by condition. Numbers on top of the small bar plot denote total UMI count assigned to CD31 within cell-containing droplets from each sample. Barcodes to the right depicts cell type by color at the corresponding rank to visualize specificity of the antibody. Horizontal line depicts gating threshold for cells considered positive for CD31.

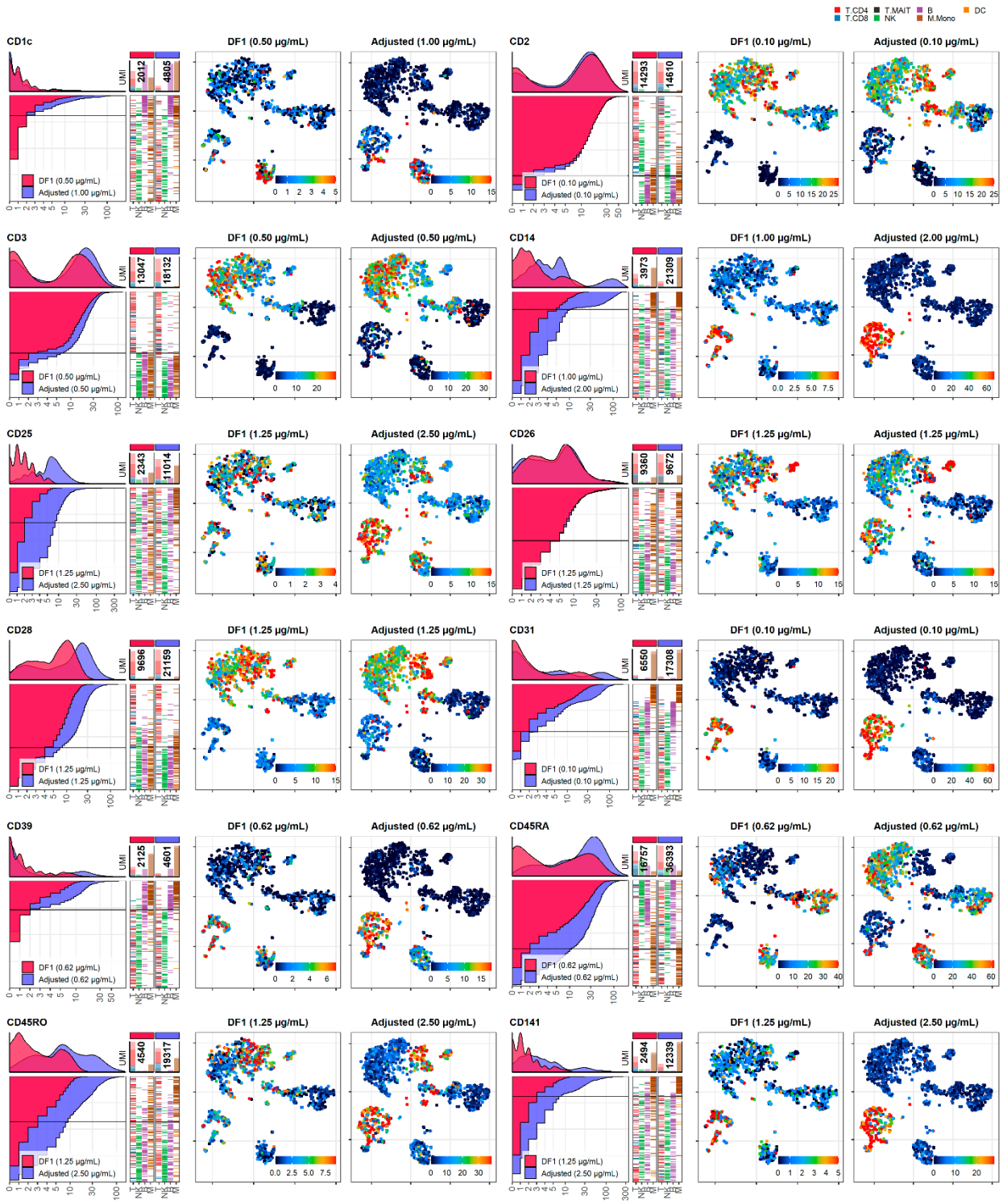
Supplementary Figure S7A



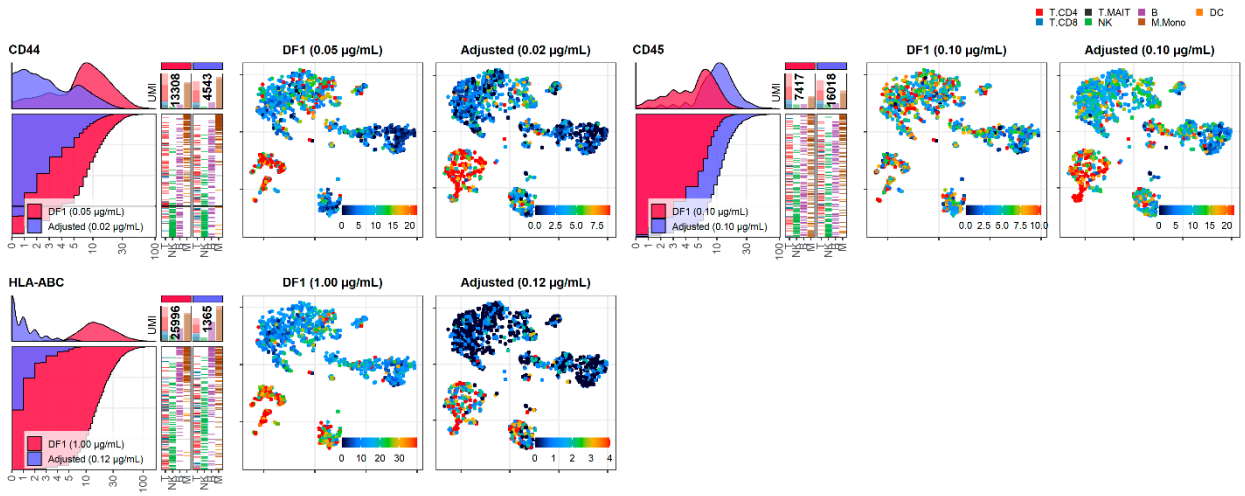
Supplementary Figure S7B



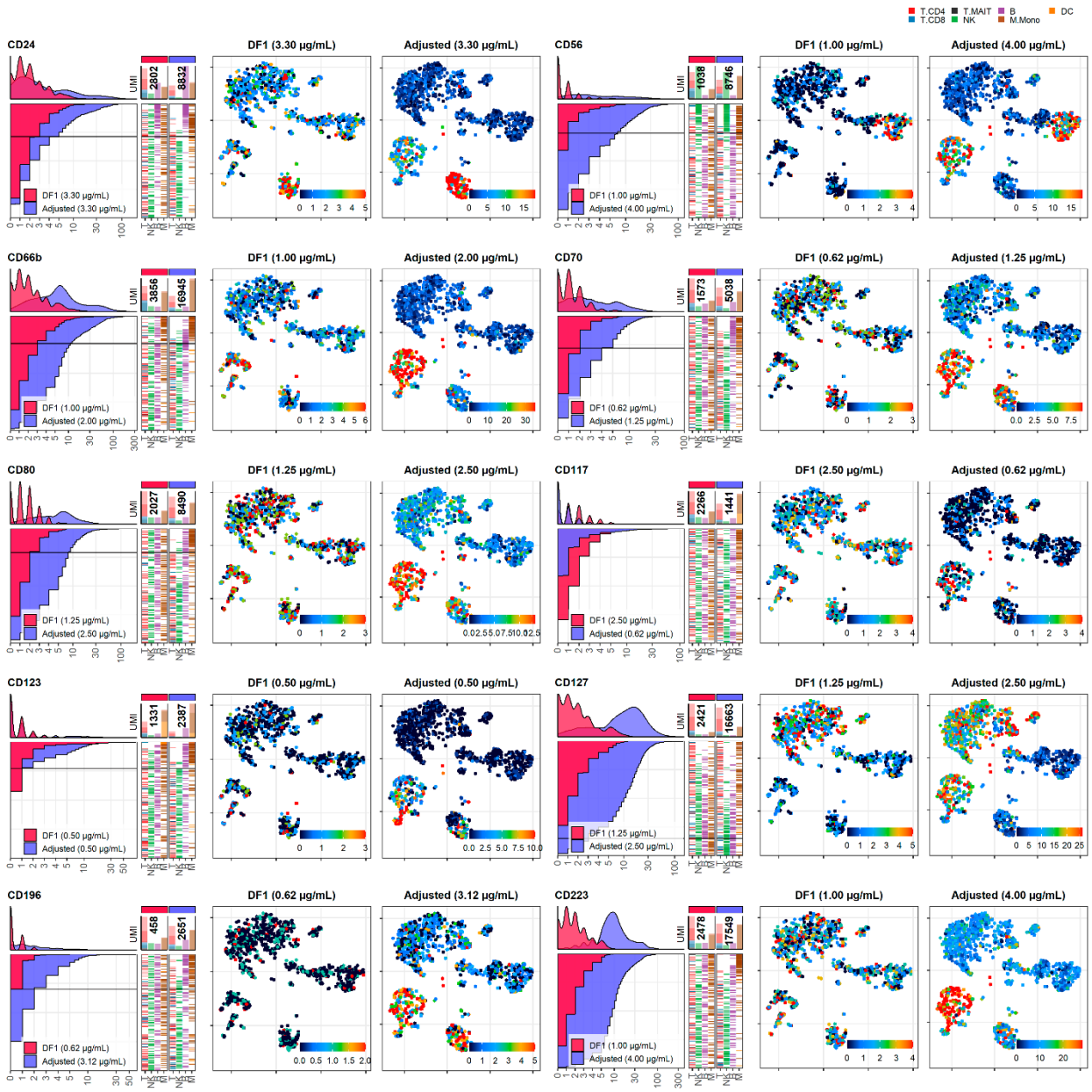
Supplementary Figure S7C



Supplementary Figure S7D

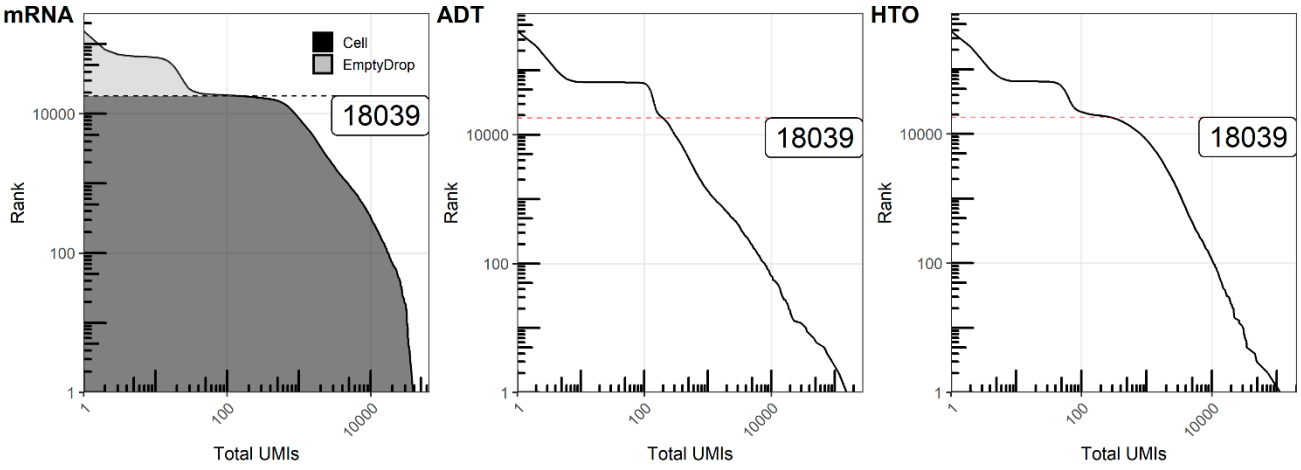


Supplementary Figure S7E



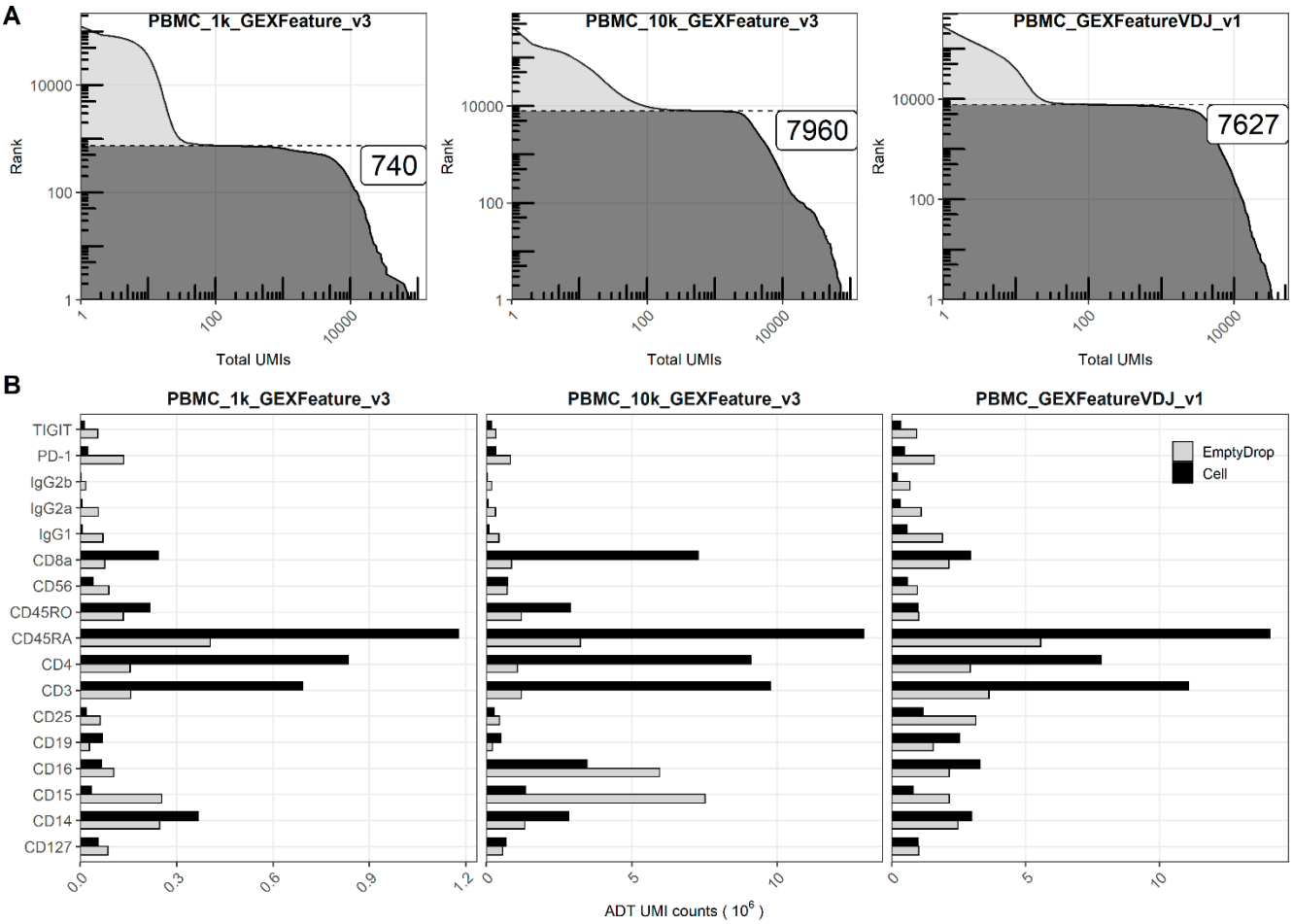
Supplementary Figure S7: DF1 versus adjusted antibody concentration comparisons
Titration plots (left; marker UMI count vs. cell rank) and tSNE plots colored by raw UMI counts (right) showing response to changing the antibody concentration from dilution factor 1 (DF1) to the adjusted sample. Within Titration plots, Histogram depicts distribution of UMIs at each condition colored by sample (and annotated with antibody concentration). Numbers within bar plot denote total UMI count within cells at each antibody concentration. Barcodes to the right depicts cell type by color at the corresponding rank to visualize cell type specificity of the antibody. Horizontal line depicts Gating threshold for cells considered positive for the marker.

Supplementary Figure S8



Supplementary Figure S8: Quantifying UMIs within cells and empty droplets of ADT and HTO. Knee plots (barcode rank vs. total UMIs within barcode) for mRNA, antibody-derived tag (ADT) and hashtag-oligo (HTO) libraries. Cell-containing droplets were filtered based on the total UMI count at the inflection point in the mRNA plot. Red lines depict the position of rank 18039, largely corresponding to the cutoff for cell-containing droplets.

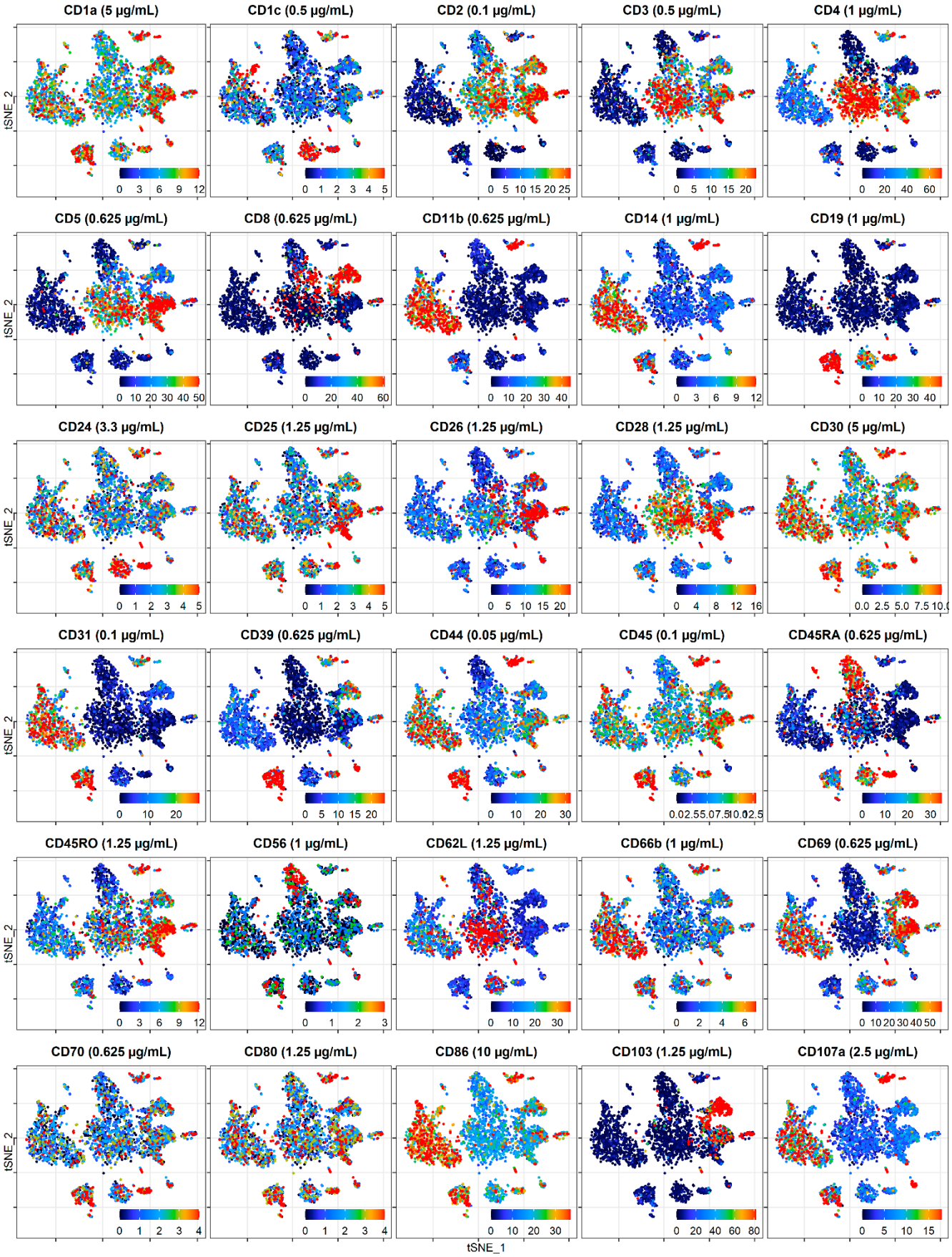
Supplementary Figure S9



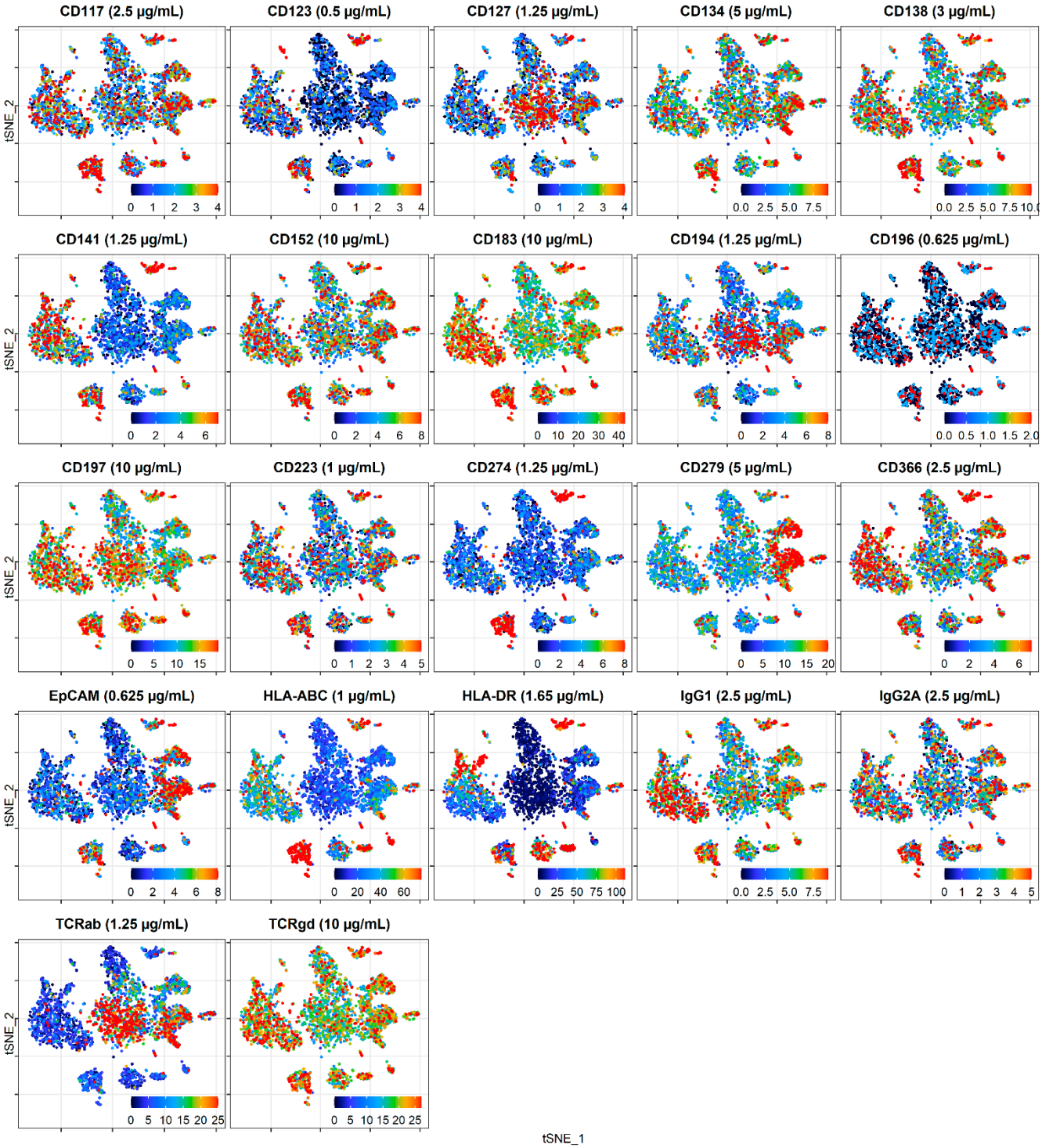
Supplementary Figure S9: Quantifying UMIs within cells-containing and empty droplets from public 10X datasets.

A. Knee plots (barcode rank vs. total UMIs within barcode) for mRNA libraries within three publicly available single-cell RNA sequencing runs (from the 10X Genomics website) showing filtering of cell-containing and empty droplets based on the total UMI count at the inflection point. **B**. Total UMI counts for the individual antibodies from each of the antibody-derived tag (ADT) libraries within cell-containing and empty droplets.

Supplementary Figure S10A



Supplementary Figure S10B



Supplementary Figure S10: Cellular distribution of raw UMI counts
tSNE plots displaying raw (unnormalized) UMI counts from the cells stained at dilution factor 1 (DF1) for each antibody. To make the color scale less sensitive to outliers, we set upper threshold at the 90% percentile.