APPLICATION NOTE

Titration of TotalSeq-A Universal Cocktail antibodies in bulk RNA-Seq reveals key expression differences

- Optimized use of BioLegend TotalSeq antibodies for simultaneous detection of cell surface proteins and RNA transcripts in cell populations
- Robust TotalSeq antibody staining with analysis of biological and technical variables on protein expression profiles
- Efficient, cost-effective, multiomic workflow for bulk-cell samples using diluted TotalSeq antibodies while maintaining protein expression profiles

In collaboration with

BioLegend®

illumina®

Introduction

TotalSeq oligo-conjugated antibodies (BioLegend) use next-generation sequencing (NGS) to enable single-cell proteomics in parallel with RNA sequencing (RNA-Seq) for improved multiomic studies. Oligonucleotide tags conjugated to the TotalSeq monoclonal antibodies, also referred to as antibody-derived tags (ADTs), provide unique identifiers for NGS quantification of proteins. Because ADTs can be sequenced and quantified in the same way as RNA in RNA-Seq applications, TotalSeq oligo-conjugated antibodies have been repeatedly used in single-cell proteomics studies.[1-5]

However, even in bulk-cell and tissue sample analysis, RNA-Seq and proteomics data from the cell surface can elucidate important differences between gene and protein expression patterns. We have previously described the application of TotalSeq antibody staining in a bulk RNA sequencing context (BEN-Seq) for identifying changes in the cell microenvironment.[6] Here, we demonstrate that protein expression signatures are largely maintained in different bulk cell samples when using TotalSeq antibody concentrations at 1/10th of the stock undiluted antibodies. While single-cell experiments require TotalSeq antibodies at higher concentrations, this study demonstrates a more cost-effective approach for adding proteomics analysis to any bulk RNA-Seq experiment.

Methods

Sample preparation

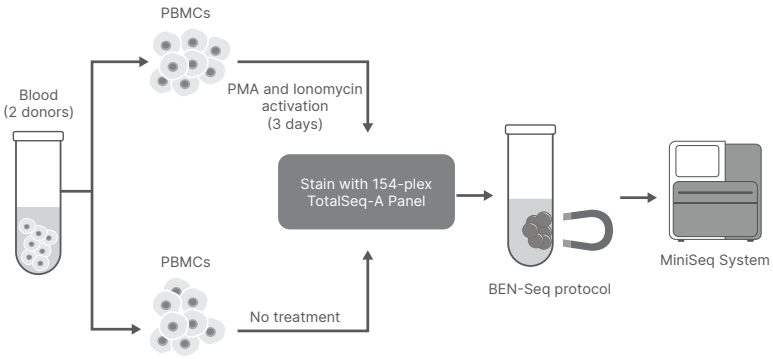
Human peripheral blood mononuclear cells (PBMCs) were isolated from whole blood samples from two volunteer donors and treated with 25 ng/ml phytohemagglutinin (PMA) and 1 µg/ml ionomycin for six hours to stimulate secretion of cytokine proteins for analysis (Figure 1).[7] For each donor, a total of three samples were prepared: resting cells, activated cells, and a mixture containing 10% activated and 90% resting cells to test sensitivity for trace levels of activated cell signal.

To determine the optimal concentration of the TotalSeq panel reagents required for bulk experiments, the six samples were further divided into four subsets and stained using the undiluted TotalSeq-A Human Universal Cocktail, v1.0 (BioLegend, Catalog no. 399907), or with a 1/10th dilution of the same panel. In total, 24 samples were used (two donors × three conditions × two dilutions × two replicates).

Library preparation

ADT fractions of PBMCs were isolated following the pelleted cells protocol for the Quick-RNA Whole Blood kit (Zymo, Catalog no. R1201) and RNA Clean & Concentrator

A. Experimental workflow



B. Bulk-Seq input specifications

	Bulk-RNA	Bulk-ADT
Input (cells)	0.5–1M	0.5–1M
Input (ng)	100 ng	—
Features	Transcriptome (~30K genes)	154 antibodies
Reads per feature	—	~3.3K
Reads	25M	1M

Figure 1: Bulk-cell proteogenomic sequencing parameters—(A) Experimental design for comparing unstimulated and Th1-polarized human PBMCs using BEN-Seq proteogenomics. (B) Input requirements for bulk-cell proteogenomics samples.

kit (Zymo, Catalog no. R1017), using proteinase K and without the DNase step. ADTs were separated using streptavidin-magnetic beads loaded with complementary oligonucleotides.[8] Isolated ADT fractions were subjected to first-strand synthesis using qScript Flex cDNA Synthesis kit (Quantabio, Catalog no. 95049-025), and then purified using SPRI beads (Beckman Coulter, Catalog no. B23317).

Resulting libraries were amplified with sparQ HiFi PCR Master Mix (Quantabio, Catalog no. 95192-250) for 16 cycles with the appropriate index primers (RPI1-24), followed by SPRI bead purification. The fragment size and concentration of DNA libraries were verified on the Agilent TapeStation, using High Sensitivity D1000 ScreenTape (Agilent, Catalog no. 5067-5584) and the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Catalog no. Q32851), respectively. The amplified libraries were diluted to 1 nM with nuclease-free water. Sequencing was performed using 30 µl of each diluted library and assayed using paired-end reads (2 × 30 bp) and read depth of 1M reads/ADT, using two MiniSeq™ High Output Reagent Kits (75-cycles, Illumina, Catalog no. FC-420-1001).

Data analysis

FASTQ sequencing files were quantified with kallisto software[9] using an alignment index that included the 15-base protein barcode sequences in the ADT panel and all possible sequences within two Hamming distances from the barcode sequences. DESeq2 software[10] was used to identify differentially expressed (DE) ADTs, and to perform variant-stabilized transformation for downstream analyses and visualization (Figure 2).

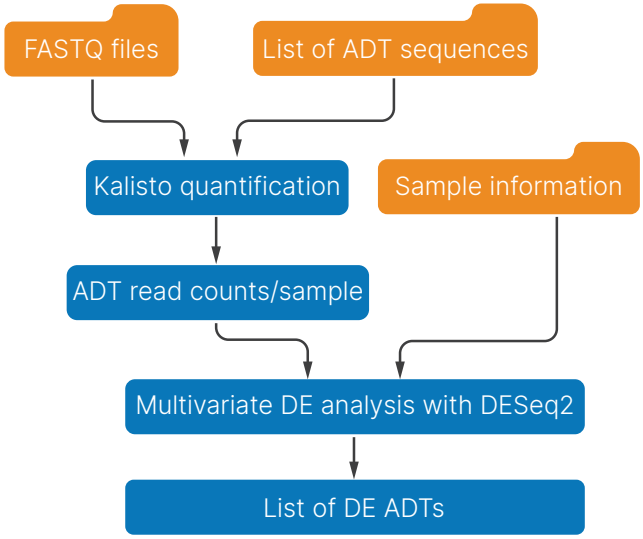


Figure 2: Overview of the workflow for expression analysis of ADTs.

Results

Basic alignment metrics

Before using diluted TotalSeq antibodies to draw experimental conclusions for bulk-cell studies, it is important to make sure that the reduced concentration does not negatively impact basic sequencing performance compared to undiluted antibodies. Alignment and sequencing metrics for each of the experimental conditions, using the 1/10th dilution of antibodies or undiluted antibodies, are shown in Figure 3. All samples had read counts between 1.1M to 3.2M reads and alignment rates above 90%. The dilution of the TotalSeq antibodies did not have a significant effect on read counts for the bulk RNA-Seq (unpaired t-test: p=0.37).

Differential protein expression analyses

DE analyses were performed on all three pairwise comparisons for the different cell states: 100% resting, 100% activated, and 10% activated + 90% resting. Regardless of antibody concentration, comparisons between 100% activated cells with either 100% resting or 10% activated yielded similar lists of DE ADTs (Figure 4, Figure 5).

One of the key advantages of simultaneously detecting RNA and protein is that it removes the need to make inferences about one molecule from the other. Biologically, it is not uncommon for some gene–protein pairs to be negatively correlated and it is important to note that in this study the dilution of the TotalSeq antibodies had no effect on the directionality of DE targets. Not only did antibody concentration maintain the directionality of differential expression results, we also observed that the most differentially expressed ADTs in these comparisons were largely consistent (Figure 5B and Figure 5E). *CD62L*, *CD99*, and *CD4* are highly downregulated genes whereas *CD69*, *CD63*, and *CD154* are frequently upregulated in fully activated PBMCs stained with the 1/10th diluted TotalSeq panel (Figure 5B). These results are consistent with single-

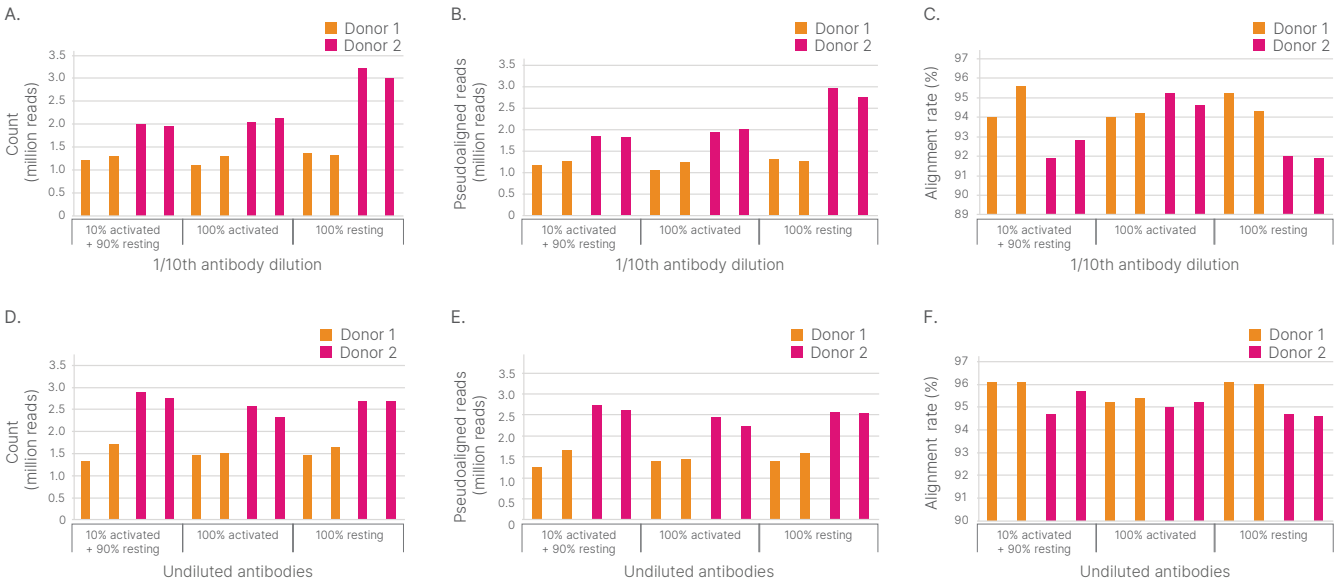


Figure 3: Sequencing and alignment comparisons for data generated from 1/10th diluted TotalSeq antibodies or undiluted antibodies—Two replicates were created for donor 1 and donor 2 samples in each condition. (A) Read counts for DE experiments with 1/10 antibody dilution, (B) pseudoaligned reads for DE experiments with 1/10 antibody dilution, (C) percent alignment for DE experiments with 1/10 antibody dilution, (D) Read counts for DE experiments with undiluted antibodies, (E) pseudoaligned reads for DE experiments with undiluted antibodies, (F) percent aligned for DE experiments with undiluted antibodies.

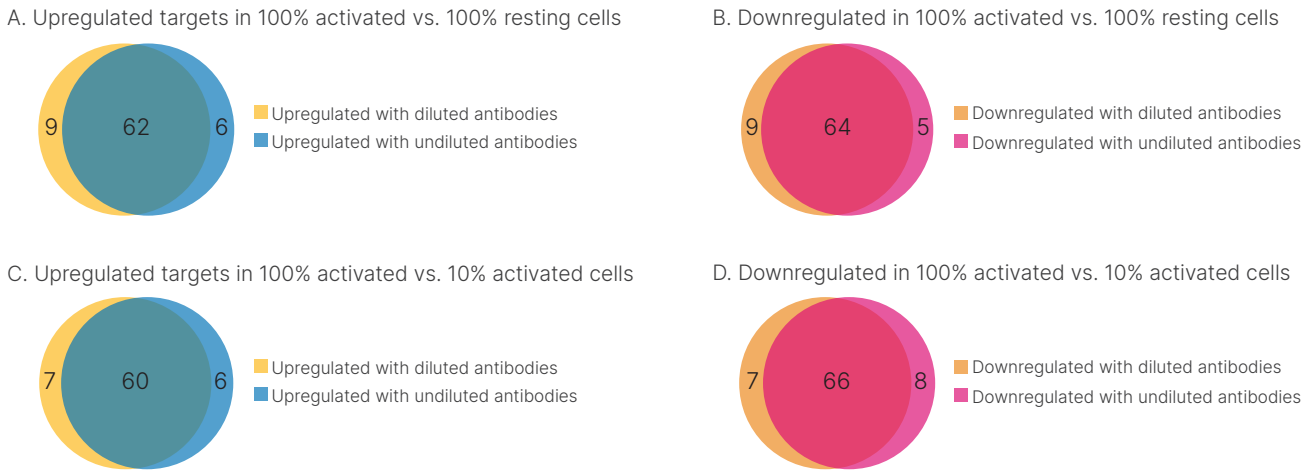


Figure 4: Summary of differentially expressed genes detected in bulk cell samples using 1/10th diluted TotalSeq antibodies or undiluted antibodies—Each diagram represents a DE ADT comparison of TotalSeq antibodies concentrations: undiluted, or 1/10 dilution.

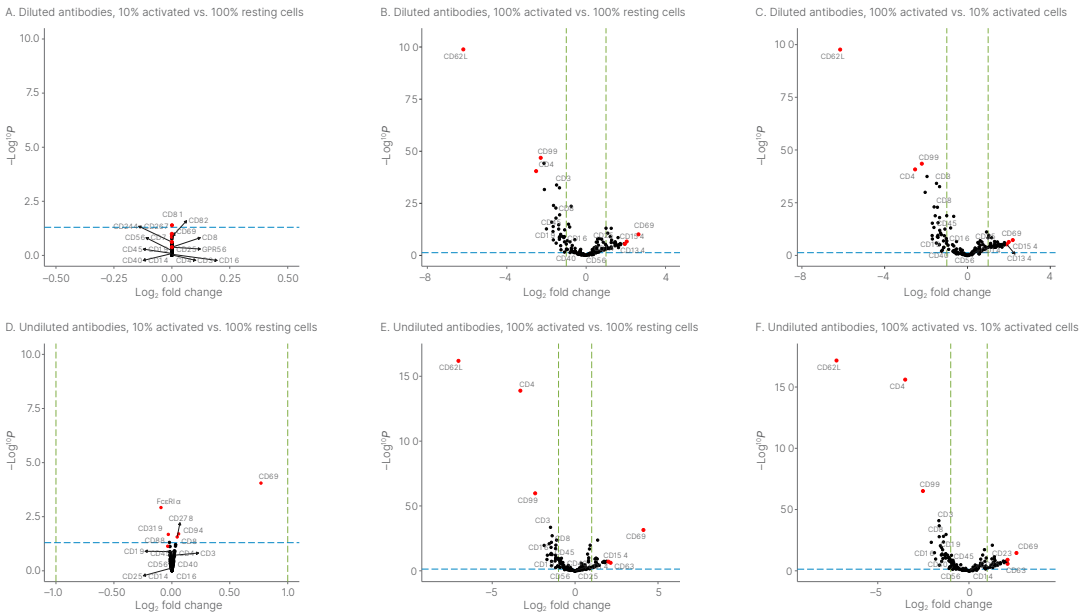


Figure 5: Differential expression of ADTs detected using either 1/10th diluted TotalSeq antibodies or undiluted antibodies—Dashed blue horizontal lines indicate p-value = 0.05 and dashed green vertical lines represent a $\log_2$ fold change +/– 1. The top three ADTs in terms of fold-change difference in either direction are marked red. The sample with the lowest activated cell content is used as the denominator in fold-change calculations.

cell protein data using TotalSeq antibodies and RNA-Seq data for PBMC activation using PMA/Ionomycin, which, based on internal research data from BioLegend, mainly involves T-cell differentiation and maturation (Figure 6).

In this study, five DE ADTs were identified between 100% resting and 10% activated samples using undiluted antibodies, while only one DE ADT was identified using diluted antibodies (Figure 5A and Figure 5D). The DE ADTs identified in these comparisons tend to have small fold changes, but they are still biologically relevant to the expected effect of PMA activation. For reference, a heat map of variance stabilizing transformation (VST)-normalized expression data for ADTs showing the largest expression differences between the 100% activated and 100% resting states are shown in Figure 7.

Principal component analysis

To make sure that the differences we observed were associated with the experimental conditions and not the TotalSeq antibody dilution or technical replication, we performed a principal component analysis (PCA) on the VST-normalized ADT expression matrix. Scatter plots of the first two principal components demonstrate that the biological factors, donor and proportion of activated cells, are determining factors of ADT expression differences (Figure 8A). Factors of a technical nature, antibody concentration and replication, do not have a major effect on ADT counts (Figure 8B). This further demonstrates the robustness of the TotalSeq antibody staining.

Discussion

We have previously demonstrated that TotalSeq-A Universal Cocktail antibodies can be applied to a bulk RNA-Seq analysis to identify changes in sample-level cell-surface protein composition.[6] Here, we show how different variables, such as biological and technical replicates and concentration of the initial antibodies, affect protein expression profiles. In particular, we show that at a 1/10th dilution of the original TotalSeq-A Universal Cocktail antibodies, expression profiles are similar to those obtained with undiluted antibodies, indicating that the TotalSeq-A Universal Cocktail can be used at lower concentrations to add cost-effective proteomics analysis to bulk RNA-Seq experiments. We believe that our results can be extrapolated to single TotalSeq antibodies, as the single conjugates should be able to produce similar patterns in bulk sequencing when titrated at lower levels.

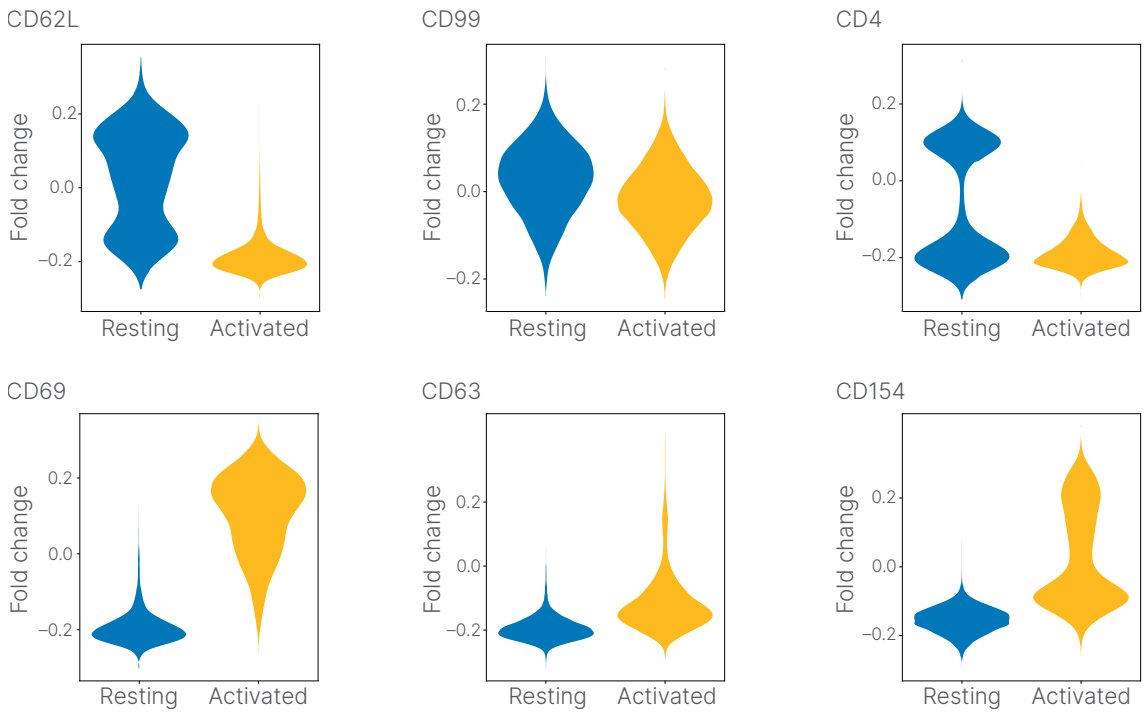


Figure 6: Differential expression levels of select ADTs from a single-cell TotalSeq and RNA-Seq study of PBMC activation—ADT expression data from single-cell experiments shown for ADTs that were identified in the bulk-cell study. Expression patterns are largely consistent for the effects of PBMC activation by PMA/Ionomycin in single-cell and bulk-cell studies.

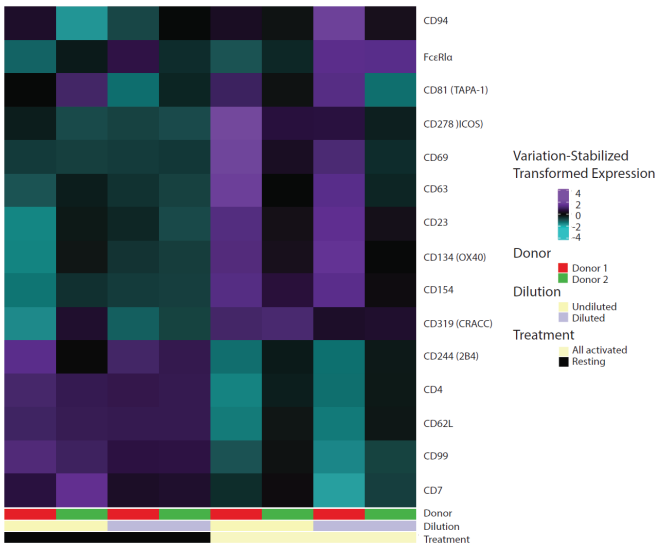


Figure 7: Expression heat map of ADTs with the largest expression differences between activated and resting states.

Overall, our results show an optimization of the previously published BEN-Seq protocol[8] and provide further evidence of the utility of BEN-Seq, paving the way for expanding to multidimensional data in bulk RNA sequencing applications. The inclusion of protein data in this type of analysis can help characterize complex samples at a deeper level to enable better understanding of biological processes. The BEN-Seq method described is also highly flexible and can be easily adapted for RNA-Seq applications on any Illumina sequencing system.

It is important to point out that the results reported in this study are specific to bulk-seq sequencing applications. Diluting optimized antibody mixtures to the same degree in single-cell sequencing applications could result in loss of necessary sensitivity and resolution in the data generated.

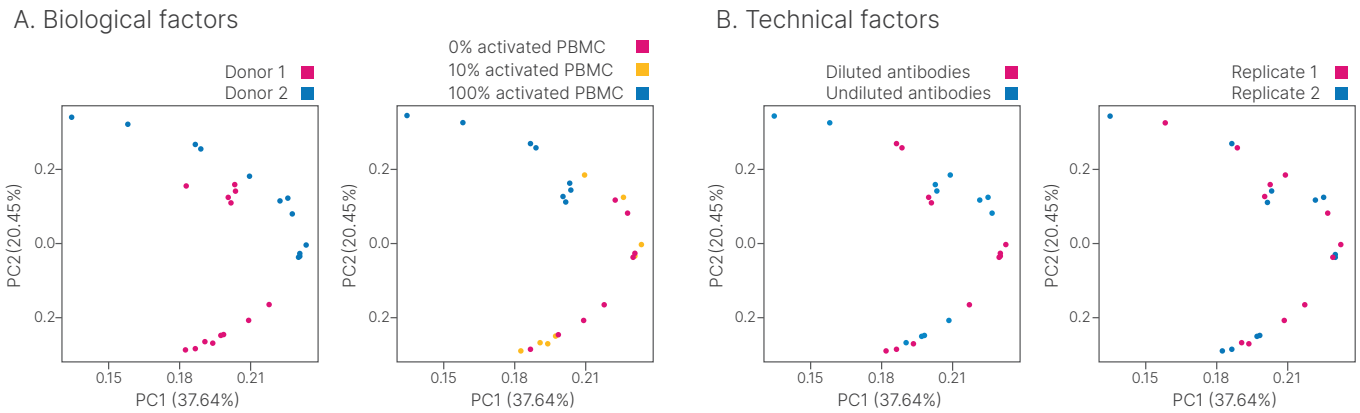


Figure 8: Principle component analysis scatter plots of the first two principal components of ADT protein expression—(A) biological components donor and proportion of activated cells, (B) technical components antibody dilution and replication.

References

1. Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*. 2017;14(9):865-868. doi:10.1038/nmeth.4380
2. Gaublomme JT, Li B, McCabe C, et al. Nuclei multiplexing with barcoded antibodies for single-nucleus genomics [published correction appears in Nat Commun. 2020 Nov 2;11(1):5619]. *Nat Commun*. 2019;10(1):2907. Published 2019 Jul 2. doi:10.1038/s41467-019-10756-2
3. Karagiannis TT, Cleary JP Jr, Gok B, et al. Single cell transcriptomics reveals opioid usage evokes widespread suppression of antiviral gene program. *Nat Commun*. 2020;11(1):2611. Published 2020 May 26. doi:10.1038/s41467-020-16159-y
4. Su Y, Chen D, Yuan D, et al. Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19. *Cell*. 2020;183(6):1479-1495.e20. doi:10.1016/j.cell.2020.10.037
5. Arunachalam PS, Charles TP, Joag V, et al. T cell-inducing vaccine durably prevents mucosal SHIV infection even with lower neutralizing antibody titers. *Nat Med*. 2020;26(6):932-940. doi:10.1038/s41591-020-0858-8
6. Illumina. Correlated expression of protein and RNA reveals a unique molecular signature in Th1-polarized cells. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/biolegend-ben-seq-app-note-m-gl-00024/biolegend-ben-seq-app-note-m-gl-00024.pdf. Accessed November 29, 2021.
7. Ai W, Li H, Song N, Li L, Chen H. Optimal method to stimulate cytokine production and its use in immunotoxicity assessment. *Int J Environ Res Public Health*. 2013;10(9):3834-3842. Published 2013 Aug 27. doi:10.3390/ijerph10093834
8. BioLegend. BEN-Seq Protocol. Updated November 2020. https://www.biolegend.com/en-us/protocols/bulk-epitope-and-nucleic-acid-sequencing-ben-seq-protocol. Accessed November 8, 2021.
9. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification [published correction appears in Nat Biotechnol. 2016 Aug 9;34(8):888]. *Nat Biotechnol*. 2016;34(5):525-527. doi:10.1038/nbt.3519
10. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-00475 v1.0