APPLICATION NOTE

Correlated expression of protein and RNA reveals a unique molecular signature in Th1-polarized cells

- Optimized BEN-Seq method for simultaneous bulk protein and gene expression profiling using Illumina NGS
- Easy quantification of multiple protein targets on any Illumina sequencing platform using oligo-conjugated TotalSeq™ antibodies
- Characterization of human Th1-polarized cells using BEN-Seq and CITE-Seq

In collaboration with

BioLegend®

illumina®

Abstract

Bulk RNA sequencing (RNA-Seq) has become a widely used tool for measuring gene expression, whereas flow cytometry is considered the gold standard for studying protein expression at a cellular level. Recent developments in single-cell sequencing technologies have combined the analysis of RNA and protein in a technique called cellular indexing of transcriptomes and epitopes by sequencing (CITE-Seq). This method uses antibodies conjugated to oligonucleotides, available as TotalSeq antibodies from BioLegend. These oligo-conjugated antibodies have been incorporated into an RNA-Seq workflow for simultaneously obtaining RNA and protein data using Illumina next-generation sequencing (NGS), described as bulk epitope and nucleic acid sequencing (BEN-Seq), rendering a complete proteogenomic profile of the sample. To demonstrate the utility of this approach, we performed CITE-Seq and bulk RNA and protein sequencing with 280 TotalSeq antibodies on T helper 1 (Th1)-polarized human peripheral blood mononuclear cells (PBMCs). We found a unique molecular profile that characterized Th1 cells with a strong correlation between both methods, demonstrating how protein analysis can be used to enrich bulk RNA-Seq data sets.

Introduction

The interrelationship of DNA coding with RNA and protein expression governs all biological processes. Characterizing their regulation is crucial to understanding the development, homeostasis, and disease states of an organism. Efficient charicterization of cells requires the use of high-throughput, multiparameter analysis tools, such as flow cytometry and next-generation sequencing (NGS). RNA-Seq is a highly sensitive and accurate method for measuring expression levels across the entire transcriptome. This technology provides researchers with a global view of gene expression changes in disease and development states in response to environmental conditions or therapeutics. However, workflows that simultaneously obtain gene and protein expression information have not been systematically optimized.

The recent development of single-cell RNA sequencing (scRNA-Seq) gives researchers a choice to profile their samples in bulk or at single-cell resolution. scRNA-Seq is well suited for cell type identification and functional studies. However, different parameters, such as sample type and the nature of the scientific hypothesis tested, will define which method to use.

Furthermore, the recently developed CITE-Seq method incorporates simultaneous detection of proteins with scRNA-Seq in the same individual cell.[1,2] The technology is supported by BioLegend antibodies conjugated to carefully designed oligonucleotides, available under the product name TotalSeq. Each antibody-conjugated oligonucleotide contains a PCR handle, an RNA capture sequence, and a 15-nucleotide barcode that identifies a unique clone and, therefore, the target protein. After binding to their targets, the conjugated oligonucleotides are captured by the same method as the cellular mRNA in scRNA-Seq, providing easy workflow integration. Using Illumina NGS as a single molecular readout, the combination of RNA and protein measurement provides the proteogenomics profile of individual cells in a high-throughput manner. The multiplexing capacity of CITE-Seq expands from traditional techniques such as flow cytometry and mass cytometry (CyTOF), allowing for higher dimensional data generation and processing when needed. It also enables complex comparisons between gene and protein expression within and between distinct cell populations.[3,4]

Currently, to obtain multiomic data when doing bulk RNA-Seq, separate techniques must be employed to analyze DNA, RNA, and protein. This is due to the lack of an integrated workflow, or lack of compatibility, for simultaneously obtaining data from these three different cellular components. A common approach has been analysis of the protein fraction by Western blotting, or an equivalent protein analysis technique. However, Western blotting offers limited ability to detect multiple proteins efficiently, and it can be challenging to detect surface proteins. Conversely, flow cytometry provides robust protein characterization but lacks the flexibility for accurately measuring RNA expression.

In this paper, we demonstrate how to incorporate protein detection into bulk RNA-Seq and develop a workflow for BEN-Seq. We polarized Th1 cells from human PBMCs and characterized their mRNA and protein profile using unpolarized cells as controls. The protocol was optimized for compatibility with any Illumina sequencing instrument. We found a strong correlation between the single-cell and bulk proteogenomic data and we observed unique patterns of molecular changes in both T-cells and antigen-presenting

cells (APCs). Furthermore, the method can be applied to current bulk RNA-Seq workflows to generate complementary protein data and more informative data sets overall.

Materials and methods

Sample processing and cytokine detection

Human PBMCs were isolated from healthy donors and stimulated with phytohemagglutinin (PMA) and Ionomycin for 4 hours. An aliquot was also treated with monensin to measure intracellular cytokines using flow cytometry and secreted cytokines using the BioLegend LEGENDplex™ assay kit following the manufacturer protocols.

Flow cytometry was performed using a BD LSRFortessa instrument and the data analyzed using FlowJo software.

Bulk sequencing

To incorporate protein analysis into bulk RNA-Seq, PBMCs for BEN-Seq were stained with 280 TotalSeq antibodies using the BioLegend BEN-Seq protocol.

RNA and antibody-derived tag (ADT) fractions of PBMCs were isolated using a column-based whole blood RNA isolation kit (pelleted cells protocol) with Proteinase K but without the DNase step. RNA and ADT fractions were separated using streptavidin-conjugated magnetic beads loaded with complementary oligonucleotides following the BioLegend BEN-Seq protocol (Figure 3). The remaining RNA fraction was processed using the Illumina TruSeq™ Stranded mRNA Library Prep Kit with Illumina TruSeq RNA UD Indexes. Samples were sequenced on Illumina MiniSeq™, NextSeq™ 550, and NextSeq 2000 Sequencing Systems.

Single-cell sequencing

PBMCs for scRNA-Seq were stained with 280 TotalSeq antibodies using BioLegend TotalSeq-A Antibodies and Cell Hashing and the 10x Single Cell 3′ Reagent Kit v3/3.1 protocol. Cells were counted with a Countess II FL Automated Cell Counter and loaded into a 10x Genomics Chromium Controller. The Chromium Single Cell 3′ Reagent Kit v3.1 was used according to the manufacturer protocol for single-cell RNA (scRNA) isolation. ADT libraries were generated in parallel using BioLegend TotalSeq-A antibodies. Samples were multiplexed using BioLegend Cell Hashing reagents, and sequenced on Illumina NextSeq 550, NextSeq 2000, and NovaSeq™ 6000 Sequencing Systems.

Data analysis

Bulk RNA-Seq raw reads were normalized using regularized log as implemented by DESeq2. For scRNA-Seq analysis, the number of unique molecular identifiers (UMI) mapped to each gene/barcode was normalized using an arsinh (inverse hyperbolic sine) transformation. Heat maps were plotted using the R ComplexHeatmap module.

Results

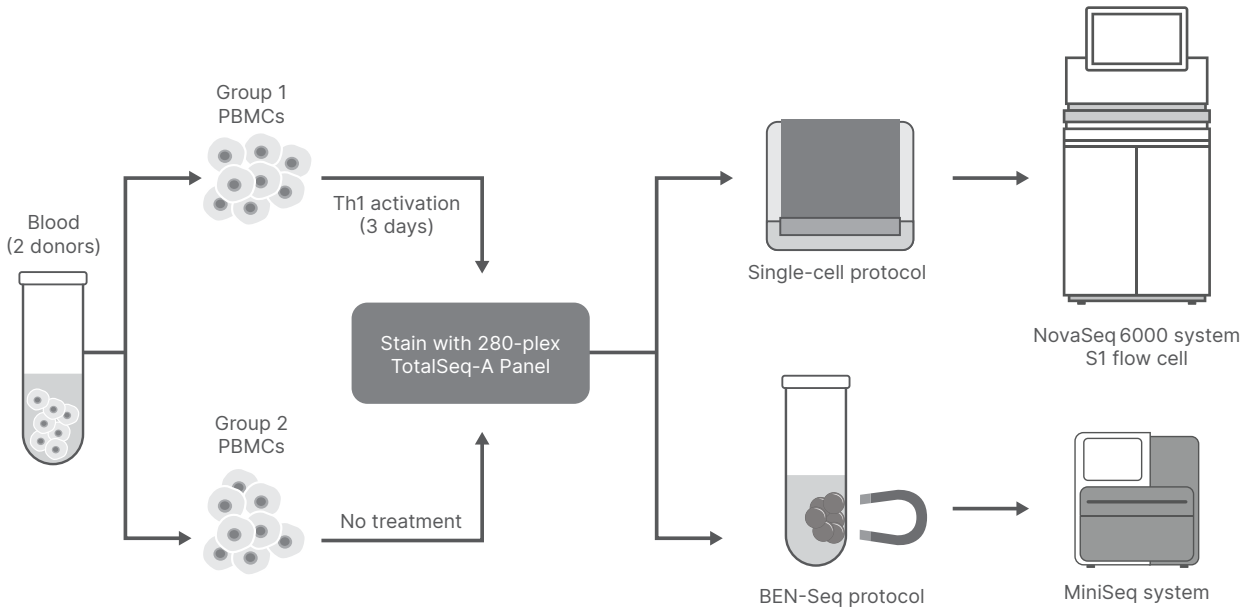
PBMCs from two donors were isolated and separated into two groups. Cells from Group 1 underwent a five-day Th1 polarization and Group 2 served as a negative control. After the five-day polarization period, cells were stained with a 280-antibody TotalSeq-A panel. Both groups were further divided into subgroups for single-cell or bulk RNA-Seq experiments (Figure 1).

To verify the outcome of the Th1-polarization, intracellular expression of the Th1 associated cytokines IL-2, IFN-γ, and TNF-α was measured by flow cytometry on polarized cell suspensions (Figure 2, dot plots) and by LEGENDplex on frozen supernatant (Figure 2, bar graphs).

For BEN-Seq, TotalSeq ADT, RNA, and genomic DNA were isolated using commercially available purification-spin columns and streptavidin-conjugated magnetic beads loaded with complementary oligonucleotides (Figure 3).

Analysis of BEN-Seq data revealed that genes for canonical Th1 markers are upregulated in response to Th1 polarization (Figure 4). Several Th1-associated transcription factor genes were upregulated, including *STAT4* and *TBX21*. Transcription of *IFN-γ*, *TNF-α*, and *TNF-β* genes was also upregulated as expected in response to Th1 polarization (Figure 4A). Gene and protein expression levels of Th1-associated cell surface markers were observed. Figure 4B shows a summary of 53 different molecules that were either upregulated or downregulated as compared to the control group. Following Th1 polarization, CD40, a

A.



B.

	Bulk-RNA	Bulk-ADT
Input (cells)	0.5–1M	0.5–1M
Input (ng)	100 ng	—
Features	Transcriptome (~30K genes)	282 antibodies
Reads per feature	~1.6K	~3.3K
Reads	50M	1M

	sc-RNA	sc-ADT
Input experiment (cells)	1M	1M
Input (cells loaded)	8K	8K
Features	Whole transcriptome (~30K genes)	282 antibodies
Read pairs per cell	20K	10K
UMI per cell	4K	1K
No. cells detected	5K	5K
No. genes detected	> 1K	—

Figure 1: Single-cell and bulk sequencing parameters—(A) Experimental design to compare single-cell and bulk sequencing. The diagram shows a workflow for comparing unstimulated and Th1-polarized human PBMCs using BEN-Seq or single-cell proteogenomics. (B) Input requirements for bulk and single-cell proteogenomics samples.

co-stimulatory molecule expressed on APCs,[5] was down-regulated only at the protein level with mRNA expression being increased. This is in contrast to CD154 (CD40L), the ligand for CD40, which is expressed on several T-cell populations[6] and shows upregulation in both transcript and protein levels. This dynamic likely serves as a control for immune response, as CD40 overexpression is implicated in the pathology of many autoimmune diseases.[7]

After performing a high-level BEN-Seq analysis, cell characterization may be further dissected with an identified pattern of RNA and protein expression that can be assigned to individual cells. Single-cell multiomic analysis resolves cellular heterogeneity of each sample, using cell surface protein expression (Figure 5A). Following this approach, we were able to identify four distinct

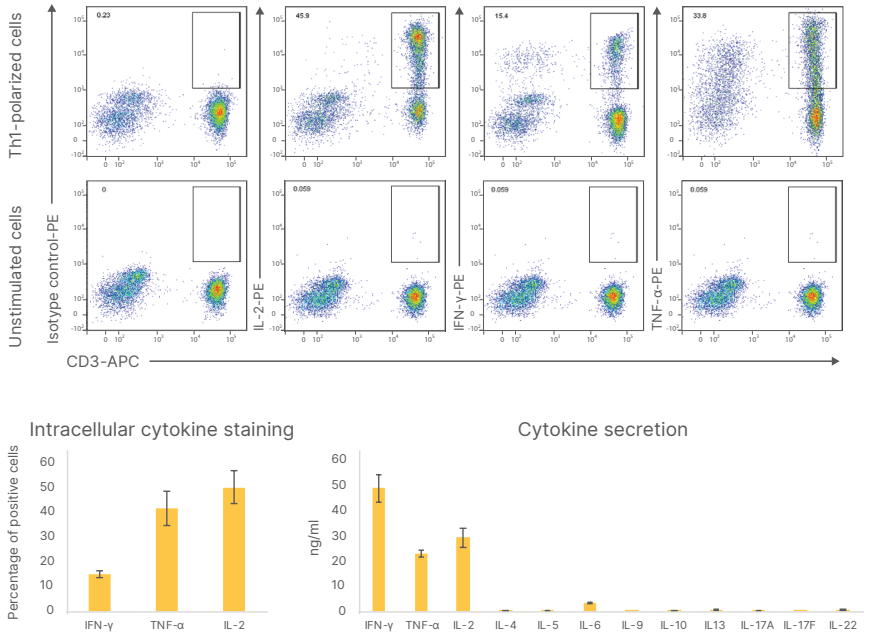


Figure 2: Th1-polarized cells show the expected cytokine profile—Flow cytometry analysis of Th1-polarized PBMCs shows production of classical Th1-associated cytokines (dot plots, top panels). Secreted levels of cytokines measured using the BioLegend Th1 LEGENDplex panel showed strong correlation of high IL-2, IFN-γ, and TNF-α levels, as well as low levels of Th2 phenotype cytokines such as IL-4, IL-5, and IL-10.

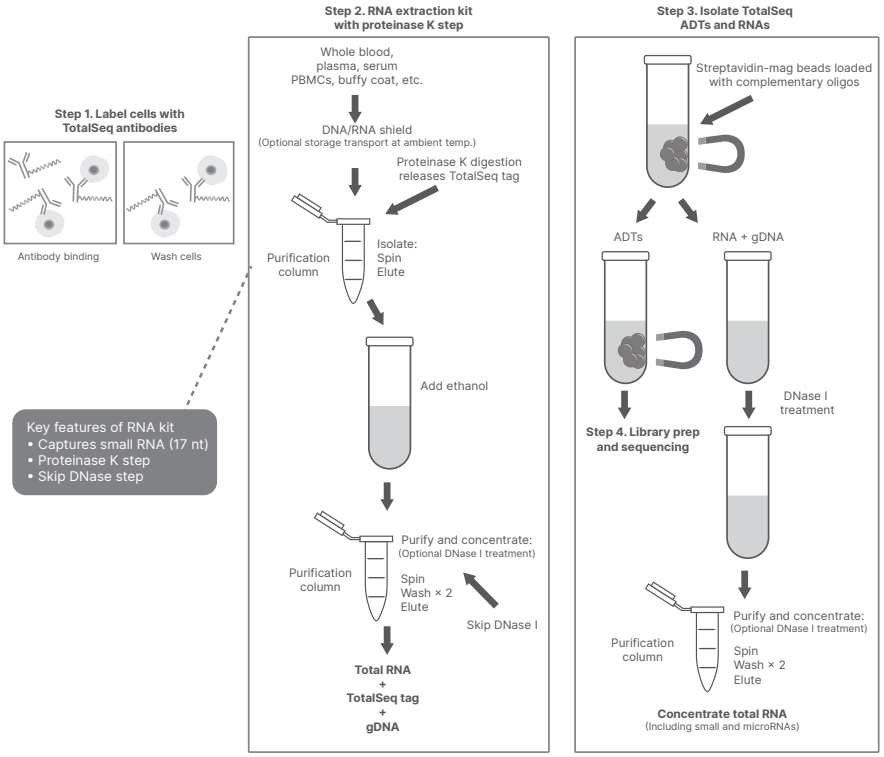


Figure 3: Optimized workflow for BEN-Seq—Cells are stained with TotalSeq antibodies then lysed and total nucleic acid is collected. ADTs are separated from cellular nucleic acid for library prep and sequencing. Cellular nucleic acid is treated with DNase I to remove genomic DNA. The remaining RNA is ready for downstream library preparation and sequencing workflows.

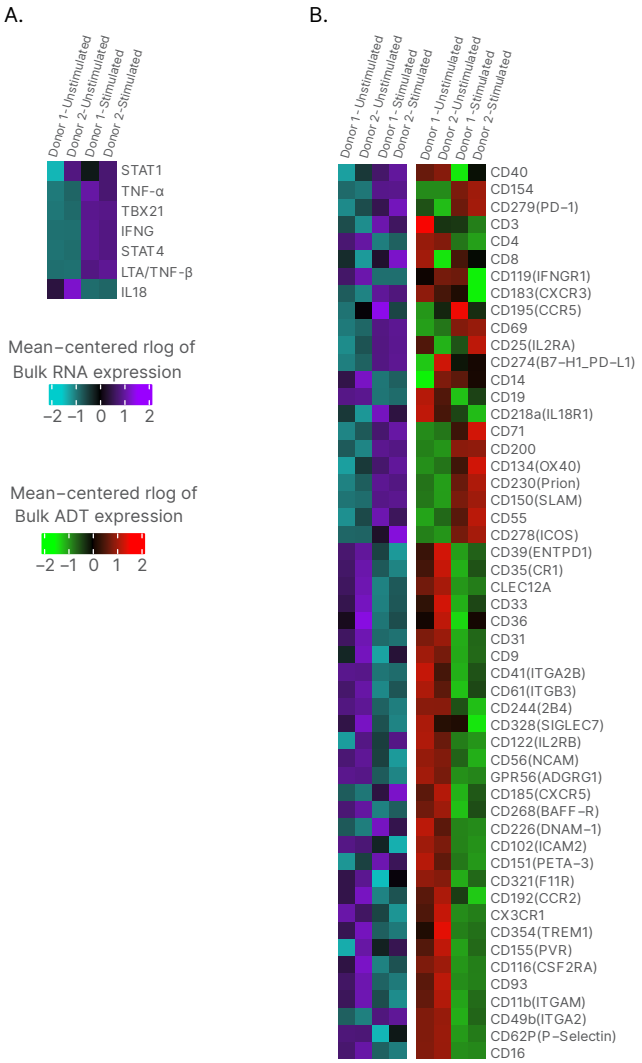


Figure 4: Bulk-Seq gene and protein expression reveals Th1 profile in polarized cells— (A) Gene expression in unstimulated and Th1-polarized human PBMCs reveals upregulation of Th1-associated molecules in polarized Th1 cells. (B) Gene and protein expression profile detected by BEN-Seq shows defined molecular patterns in Th1-polarized cells as compared to unstimulated cells.

CD4+ T-cell subsets with increased expression of CD154 in response to Th1 polarization (Figure 5B, left red box). The data also reveal that several cells with antigen presentation capacity downregulate expression of CD40 in response to Th1 polarization, including B-cells, monocytes, and dendritic cells (Figure 5B, right red boxes).

To illustrate the visualization of this result, we generated uniform manifold approximation and projection (UMAP) graphs with CD4+ Memory T-cells and B-cells from both donors. By projecting the protein expression level of CD154 and CD40 along with the corresponding gene expression data, we can see how the modulation of these two important markers changes in response to Th1 polarization at a given point in time (Figure 6).

Discussion

Bulk RNA-Seq has been an instrumental tool in understanding global gene expression changes in a myriad of recent biological studies. However, measuring gene expression changes only tells part of the biological story. Adding Western blotting data can be crucial to providing orthogonal validation of gene expression measurements and characterizing cellular processes. However, this approach is limited by technical difficulties in measuring expression levels of large sets of proteins. The introduction of oligo-conjugated TotalSeq antibodies, integrated to single-cell sequencing methods, provides an excellent avenue for the simultaneous analysis of multiple protein targets and RNA expression. Thanks to the extensive use of flow cytometry in cell biology studies, protein profiling is recognized as the gold standard of cell analysis and classification. The ability to measure protein expression via sequencing at a single-cell level unifies flow cytometry and RNA-Seq, providing a holistic approach to cell analysis. The utility of TotalSeq antibodies extends to bulk RNA-Seq as it easily allows for the measurement of protein expression, independent of single-cell analysis techniques.

In this application note, we have demonstrated the utility of adding TotalSeq antibodies to bulk RNA-Seq, termed BEN-Seq, to obtain protein expression data in addition to RNA expression data. We were able to evaluate expression levels of 280 proteins and relate changes in protein expression to their corresponding genes. Furthermore, we observed well-defined expression trends that correlated with single-cell protein and RNA expression data. We identified distinct T-cell populations that upregulated expression of CD154, PD-1, and other classical Th1-associated molecules in response to Th1 polarization. CD154 is an important cell surface protein that is expressed by activated T-cells and is the ligand for CD40, which is expressed by professional APCs.[6] We also observed that B-cells downregulated the expression of CD40 at the protein level, which may serve as a means of controlling cell-mediated responses. It is also possible that the two

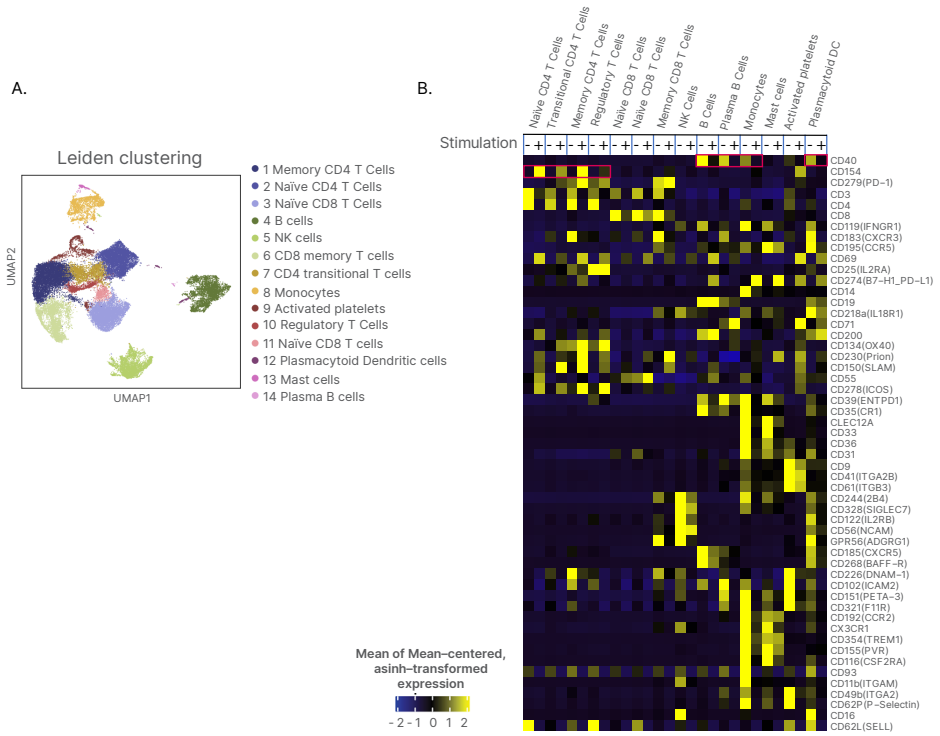


Figure 5: scRNA-Seq results correlate with bulk RNA-Seq profiles and assigns changes to individual cell types—(A) Major cell lineages classified using TotalSeq antibodies after single-cell analysis. (B) Surface protein expression determined by single-cell analysis with TotalSeq antibodies. Red boxes indicate differential responses in Th1 polarization between unstimulated and stimulated PBMCs.

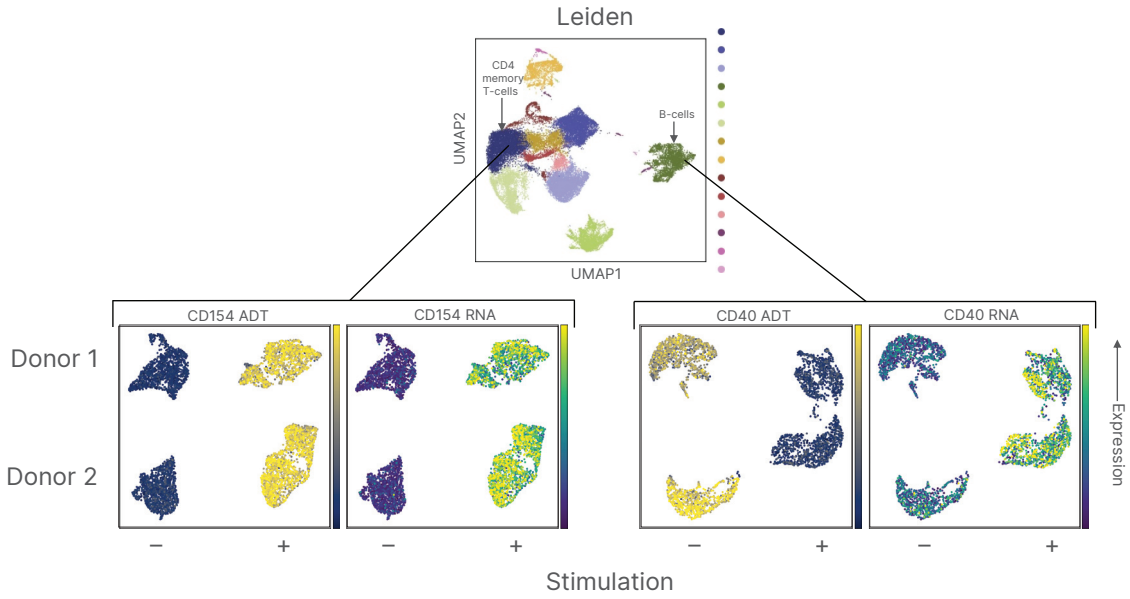


Figure 6: Analysis of mRNA and protein expression of CD154 and CD40—Expression of genes and proteins (ADT) of CD154 in CD4+ memory T-cells and CD40 in B-cells shows differential dynamics in upregulation of the two types of molecules (nucleic acid and protein) in the two target populations.

molecules are governed by different regulatory kinetics as a way to execute immunoregulatory functions.[8] These observations were highly concordant between bulk and single-cell experiments, demonstrating the reproducibility and consistency of the protein and RNA expression profiles.

Both bulk and single-cell sequencing are used as the foundation to analyze cell physiology. For example, sequencing techniques are used on sectioned tissue to add structural data, a field that is being now called spatial transcriptomics.[9,10] In addition, nucleic acid characterization is employed on cancer cells to explore gene mutations.[11] Assay for transposase-accessible chromatin using sequencing (ATAC-seq) has been applied to single-cell sequencing in the past two years to study single-cell epigenetics.[12] Compatibility and use of TotalSeq antibodies for these and other bulk and single-cell sequencing applications are on the horizon. TotalSeq is already being applied to single-cell ATAC with select antigen profiling by sequencing (ASAP-Seq) to study protein and chromatin accessibility simultaneously.[13] ASAP-Seq has also explored the use of TotalSeq antibodies to detect intracellular proteins. All of these sequencing techniques may be adapted in the future to simultaneously add their counterpart protein data set, helping researchers achieve their goals quicker and provide a more integrated analysis of their scientific question.

Additional resources

BioLegend protocols, including BEN-Seq, are available at biolegend.com/en-us/technical-protocols

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.
2. Mimitou EP, Cheng A, Montalbano A, et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat Methods*. 2019;16(5):409-412.
3. Kotliarov Y, Sparks R, Martins AJ, et al. Broad immune activation underlies shared set point signatures for vaccine responsiveness in healthy individuals and disease activity in patients with lupus. *Nat Med*. 2020;26(4):618-629.
4. Granja JM, Klemm S, McGinnis LM, et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol*. 2019;37(12):1458-1465.
5. Elgueta R, Benson MJ, de Vries VC, Wasiuk A, Guo Y, Noelle RJ. Molecular mechanism and function of CD40/CD40L engagement in the immune system. *Immunol Rev*. 2009;229(1):152-172.
6. Lederman S, Yellin MJ, Krichevsky A, Belko J, Lee JJ, Chess L. Identification of a novel surface protein on activated CD4+ T-cells that induces contact-dependent B cell differentiation (help). *J Exp Med*. 1992;175(4):1091-1101.
7. Peters AL, Stunz LL, Bishop GA. CD40 and autoimmunity: the dark side of a great activator. *Semin Immunol*. 2009;21(5):293-300.
8. Ford GS, Barnhart B, Shone S, Covey LR. Regulation of CD154 (CD40 ligand) mRNA stability during T cell activation. *J Immunol*. 1999;162(7):4037-4044.
9. Ståhl PL, Salmén F, Vickovic S, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*. 2016;353(6294):78-82.
10. Rodriques SG, Stickels RR, Goeva A, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*. 2019;363(6434):1463-1467.
11. Albertí-Servera L, Demeyer S, Govaerts I, et al. Single-cell DNA amplicon sequencing reveals clonal heterogeneity and evolution in T-cell acute lymphoblastic leukemia [published online ahead of print, 2020 Aug 18]. *Blood*. 2020;blood.2020006996.
12. Mezger A, Klemm S, Mann I, et al. High-throughput chromatin accessibility profiling at single-cell resolution. *Nat Commun*. 2018;9(1):3647.
13. Mimitou EP, Lareau CA, Chen KY, et al. Scalable, multimodal profiling of chromatin accessibility and protein levels in single cells. *BioRxiv*. 2020.09.08.286914.

illumina®

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